

# The interaction of notional number and morphophonology in subject–verb agreement: A role for working memory

Quarterly Journal of Experimental Psychology  
2019, Vol. 72(4) 890–900  
© Experimental Psychology Society 2018  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1747021818771887  
qjep.sagepub.com



Heidi Lorimor<sup>1</sup> , Carrie N Jackson<sup>2</sup> and Janet G van Hell<sup>3</sup>

## Abstract

Research shows that cross-linguistically, subject–verb agreement with complex noun phrases (e.g., *The label on the bottles*) is influenced by notional number and the presence of homophony in case, gender, or number morphology. Less well-understood is whether notional number and morphophonology interact during speech production, and whether the relative impact of these two factors is influenced by working memory capacity. Using an auditory sentence completion task, we investigated the impact of notional number and morphophonology on agreement with complex subject noun phrases in Dutch. Results revealed main effects of notional number and morphophonology. Critically, there was also an interaction between morphophonology and notional number because participants showed greater notional effects when the determiners were homophonous and morphophonologically ambiguous. Furthermore, participants with higher working memory scores made fewer agreement errors when the subject noun phrase contained homophonous determiners, and this effect was greater when the subject noun phrase was notionally singular. These findings support the hypothesis that cue-based retrieval plays a role in agreement production, and suggests that the ability to correctly assign subject–verb agreement—especially in the presence of homophonous determiners—is modulated by working memory capacity.

## Keywords

Agreement; morphophonology; notional number; cue-based retrieval; working memory; Dutch

Received: 2 September 2017; revised: 22 March 2018; accepted: 23 March 2018

## Introduction

Fluent speakers of a language often feel like language proceeds effortlessly. In fact, the very term “fluent” suggests that speakers are able to put together sentences without intrusive disfluencies or repairs. However, in spoken language, words are produced incrementally—with speakers articulating only one word at a time—even though the conceptual planning of speech is much less linear. The question of how speakers manage to build structures and create grammatical strings has been a long-standing one in psycholinguistic research (Bock & Cutting, 1992). For example, research has focused on which aspects of sentence production are largely automatic, and which depend, at least in part, on working memory (Kemper, Herman, & Lian, 2003; MacDonald, 2016; Slevc, 2011).

The computation of subject–verb agreement is especially well-suited to address the role of working memory in sentence production, since there is a reliable dependency between the number marking on the subject head

noun and the verb, which can be used to measure deviations in the syntactic production process. In sentences that contain multiple nouns, each noun is a possible agreement controller. This creates opportunities for speakers to “lose track” of the subject head noun, leading to agreement errors such as agreement attraction, which occurs when the

<sup>1</sup>Department of Languages, Cultures & Linguistics, Bucknell University, Lewisburg, PA, USA

<sup>2</sup>Department of Germanic and Slavic Languages and Center for Language Science, Pennsylvania State University, University Park, PA, USA

<sup>3</sup>Department of Psychology and Center for Language Science, Pennsylvania State University, University Park, PA, USA

### Corresponding author:

Heidi Lorimor, Department of Languages, Cultures & Linguistics, Bucknell University, One Dent Drive, Lewisburg, PA 17837, USA.  
Email: hml003@bucknell.edu

verb of a sentence agrees with the grammatical features of a “local,” or non-subject, noun (Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001). Under working memory models, constraints on working memory should have the greatest impact when a noun’s morphophonology, or the phonological realisation of morphological features, is ambiguous because such ambiguity should increase the chances of retrieving the incorrect subject head noun, resulting in more agreement attraction errors (Badecker & Kuminiak, 2007). Indeed, morphophonology has been shown to affect agreement in many languages, including Dutch (Hartsuiker, Schriefers, Bock, & Kikstra, 2003), French, Italian, and Spanish (Franck, Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008), Russian (Lorimor, Bock, Zalkind, Sheyman, & Beard, 2008), and Slovak (Badecker & Kuminiak, 2007).

Conflict between notional number representations and grammatical number valuations is also a source of agreement variability, as noun phrases (NPs) that are grammatically singular—but that refer to more than one thing (e.g., “the label on the bottles”)—are more likely to take plural agreement than notionally singular NPs (Vigliocco, Hartsuiker, Jarema, & Kolk, 1996). While notional agreement is sometimes a grammatical option (Haskell & MacDonald, 2003), it still reflects agreement with something other than the grammatical number of the head noun. However, no models of agreement currently exist that simultaneously account for working memory and notional number effects on agreement, although, a priori, participants with greater working memory skills should have better success at keeping track of subject head nouns overall (Hartsuiker & Barkhuysen, 2006). Furthermore, no studies to date have investigated morphophonology, notional number, and working memory at the same time, even though understanding how working memory interacts with morphophonology and notional number in tandem provides more insight than studying each of these components separately. As long as we continue to study these factors in isolation, we run the risk of advancing models that only capture pieces of the larger mechanisms, rather than gaining insight into all of the mechanisms involved in agreement production. In this study, we use an auditory agreement production task in Dutch that manipulates notional number and morphophonology while measuring participants’ working memory abilities, to determine whether working memory affects agreement with all types of sentences equally, or whether there are differential effects of working memory that are modulated by morphophonological ambiguity or variations in notional number.

### *Morphophonological and notional number effects on agreement*

Previous work has investigated morphophonology and notional number, and whether or not they interact with

each other, but without including tests of working memory. Antón-Méndez and Hartsuiker (2010) investigated rates of agreement attraction in Dutch, depending on the morphophonology and whether the subject NP referred to one thing (single-token) or many things (multiple-token). Their morphophonological manipulation was based on the Dutch determiners *de* and *het*. The determiner *de* is ambiguous for number because it is used for both common-gender singular nouns and all plural nouns. In contrast, the other determiner, *het*, which is used with singular neuter nouns, is not ambiguous for number. Therefore, in the items with number ambiguity, the singular determiner on the head noun and the plural determiner on the local noun would have the same form (*de*). Sample items from Antón-Méndez and Hartsuiker are shown in 1a-d:

(1a) De buurt met de fietsroutes (*de-de*; ambiguous; single token)

“The neighborhood with the bike routes”

(1b) Het dorp met de fietsroutes (*het-de*; unambiguous; single token)

“The village with the bike routes”

(1c) De dop op de flessen (*de-de*; ambiguous; multiple token)

“The cap on the bottles”

(1d) Het etiket op de flessen (*het-de*; unambiguous; multiple token)

“The label on the bottles”

Antón-Méndez and Hartsuiker’s goal was to determine whether there was an interaction between morphophonology and notional number during agreement production. They argued that a significant interaction between these two variables would provide evidence for interactive, rather than serial, models of language production. However, they did not find an interaction between morphophonology and notional number, even though they found main effects of each because there was more plural agreement with *de-de* items and with multiple-token items. Furthermore, in discussing their findings, they outlined the possibility that some types of interactions between morphophonology and notional number could also be explained not only by interactivity but also by monitoring effects, such as those predicted by working memory models (Badecker & Kuminiak, 2007). Given the renewed interest in how working memory should affect agreement, especially with regard to morphophonological effects on cue-based retrieval (Lago, Shalom, Sigman, Lau, & Phillips, 2015; Lorimor, Jackson, & Foote, 2015), we wanted to test for effects of morphophonology and notional number on agreement, and to add an independent measure

of working memory, to directly test working memory models of agreement production.

### Working memory

If working memory does affect agreement, an important question is whether it affects all instances of agreement equally, or whether it exerts greater effects in the presence of notional number conflicts or morphophonological ambiguity. One early study on agreement attraction that examined the role of working memory was Bock and Cutting (1992), which measured how often participants agreed with the plural features on the non-subject nouns, depending on whether the noun was embedded within a phrasal modifier (2a) or a clausal modifier (2b):

(2a) The demo tape from the popular rock singers . . . (phrasal)

(2b) The demo tape that promoted the rock singers . . . (clausal)

They found that agreement attraction was more common with phrasal modifiers than with clausal modifiers, suggesting that syntactic structure played an important role in whether the features on a local noun would interfere in the normal subject–verb agreement processes and lead to attraction. To investigate the role of working memory, Bock and Cutting conducted speaking-span tasks and found a significant correlation between speaking-span scores and agreement attraction in only one of their three experiments. Therefore, while they could not conclude that working memory played no role in agreement attraction, Bock and Cutting argued that working memory does not explain the majority of agreement attraction errors in English and that agreement is more of an automatic process that takes place during syntactic structure building.

Further work has shown that working memory effects on agreement may appear when secondary tasks are added (Fayol, Largy, & Lemaire, 1994) and in special populations, like persons with aphasia (Slevc & Martin, 2016). For instance, Slevc and Martin (2016) found that persons with aphasia who also exhibited working memory deficits exhibited higher rates of agreement attraction than participants in a control group. In bilingual and monolingual children, Veenstra, Antoniou, Katsos, and Kissine (2017) similarly found significant relationships between participants' rates of agreement attraction and scores on some working memory tasks.

Hartsuiker and Barkhuysen (2006) added a secondary task to increase memory load when investigating agreement attraction among nonimpaired college-aged adults, while also examining whether working memory effects were more evident when there was a conflict between notional and grammatical number. Using an auditory sentence completion task in Dutch, they manipulated working

memory by giving half of the participants three words that they needed to recall after each trial, thereby increasing the overall memory load of the task. In a separate measure, they also measured participants' speaking spans, using a task modelled on Daneman and Green (1986). Hartsuiker and Barkhuysen found an overall effect of notional number, in that participants produced more plural verb agreement with the multiple-token (notionally plural) items than with single-token items. They also found higher rates of agreement errors among the participants with low speaking spans, but only when those participants were under higher memory load. However, they found no interaction between memory load and notional number, suggesting that notional number agreement is not caused by a lack of working memory resources and that notional number agreement is, to some degree, separable from agreement attraction. Hartsuiker and Barkhuysen, however, did not manipulate morphophonology, and instead only used *de-de* items (determiners that are ambiguous for number information). Therefore, while their study provided important insights into the separability of notional effects from the role of working memory, questions remain about whether morphophonology interacts with working memory, or whether there might be a three-way interaction between working memory, morphophonology, and notional number.

### Separable components of the agreement process

The separation of notional number agreement from agreement attraction, which is supported by the lack of an interaction between notional number and working memory load in Hartsuiker and Barkhuysen (2006), is codified in the Marking and Morphing model (Eberhard, Cutting, & Bock, 2005). Within Marking and Morphing, two sources of information (notional number information from the message and grammatical number information from the individual nouns and determiners) jointly set the probability of either singular or plural agreement during subject encoding. The grammatical number computation differentially weights elements in the subject NP, so that the number on the head noun has the greatest weight, but number information on determiners or on embedded nouns can also affect the probability of plural agreement, although to a lesser extent than the number information on the head noun. When agreement errors are primarily caused by grammatical number, rather than notional number, this is called agreement "attraction."

Marking and Morphing includes in its grammatical number computation a mechanism to account for morphophonological effects that relies on the ambiguity of number cues. In Marking and Morphing, items like the Dutch determiner *het*, which are unambiguously singular, obtain a number value of  $-1$ . Items that are unambiguously

plural obtain a number value of +1. Items that are ambiguous for number (like the Dutch determiner *de*) have a number value of zero. Therefore, any unambiguously singular items will lead to a higher probability of singular agreement and therefore less agreement attraction, while number-ambiguous items will not (see Antón-Méndez & Hartsuiker, 2010, for an implementation).

A third component, beyond notional number agreement and agreement attraction, has also been proposed to account for certain types of agreement errors stemming from mis-retrieval of the subject head noun (Lorimor et al., 2015). The basic idea is similar to that of a working memory model of agreement (Badecker & Kuminiak, 2007), in which there is an additional process of subject retrieval at the point of verb production (e.g., cue-based retrieval) that might incorrectly locate the subject head noun when the form of the verb is being planned. This mechanism is similar to proposals for cue-based retrieval in agreement comprehension (Schlueter, Williams, & Lau, 2018), in which participants launch a search for the subject head noun upon encountering a verb, the success of which can be lessened by similarity-based interference (Villata, Tabor, & Franck, 2018). This retrieval component of agreement may be relevant in accounting for morphophonological effects in agreement, above and beyond those predicted by the Marking and Morphing model (Eberhard et al., 2005), and may also be a natural place to look for effects of working memory in agreement production.

### Present study

Previous studies on morphophonology in Dutch have explained the effect of grammatical gender based on number ambiguity (Hartsuiker et al., 2003), which is also the explanation given by the Marking and Morphing model (Eberhard et al., 2005). An alternate consideration is that, when the determiners on the head and local noun are ambiguous for number, they also have the same form. When the determiners have the same form, retrieval of the subject head noun should be more difficult, due to increased similarity-based interference, and effects of working memory should be evident. Therefore, we examined agreement attraction in Dutch, using the stimuli from Antón-Méndez and Hartsuiker (2010). Our study differed from Antón-Méndez and Hartsuiker in four main ways. First, we included a larger number of participants to increase statistical power. Second, we used an auditory oral sentence completion task (like Hartsuiker & Barkhuysen, 2006) instead of a visual completion task. Third, because we were looking specifically at the role of working memory, we included items with scorable disfluencies in our analyses, whereas Antón-Méndez and Hartsuiker treated all disfluent responses as miscellaneous items. This is because we anticipated that working memory retrieval effects would be more common among participants with lower working memory spans, and that

failures in working memory might lead to disfluencies, so if we excluded disfluent items, we would be eliminating an important set of data. Finally—and most importantly—we added a Dutch version of an operation span (OSpan) task (Turner & Engle, 1989) to obtain an objective measure of participants' working memory capacity, which we included as a continuous variable in our statistical analysis. Our main goal was to determine whether working memory affects agreement in Dutch, and if so, whether there were interactions between working memory, notional number, and morphophonology.

In terms of morphophonological effects, Antón-Méndez and Hartsuiker (2010) and Hartsuiker et al. (2003) have already shown that agreement errors are more likely on *de-de* items, compared to *het-de* items. This effect can be explained in two ways. First, the number ambiguity on the *de* determiner could lead to more agreement attraction, as described above. Second, the fact that the determiners have the same form could make retrieval of the subject head noun more difficult, leading to less reliable agreement with the subject head noun. If we replicate the higher rate of agreement errors in *de-de* items, and also show that working memory differentially affects agreement with *de-de* items, this will provide crucial evidence that the reduction in agreement errors on *het-de* items is due to the presence of additional retrieval cues on the determiners, and not solely by the mechanisms for morphophonology described by Marking and Morphing. Furthermore, by investigating notional effects at the same time as morphophonology and working memory, we can gain a more complete picture of all of the processes involved in shaping agreement.

## Method

### Participants

In total, 55 Dutch native speakers in the Netherlands participated in the experiment. Due to technical difficulties, responses from one participant were not recorded. An additional three participants were excluded because they answered fewer than 65% of the math problems on the OSpan task correctly, suggesting they may have privileged remembering the target words over solving the math equations, rather than attending to both components of the task. All results are based on the remaining 51 participants (41 female, 10 male). The mean age of participants was 19.3 years (standard deviation [*SD*] = 1.7, range = 18–24 years).

### Materials

The 80 experimental items were based on the sentence preambles from Antón-Méndez and Hartsuiker (2010). Minor changes were made to 16 items to avoid vocabulary overlap with other items in the task (see Supplementary

Material B for a list of all experimental items). All items contained a singular head noun followed by a prepositional phrase containing a plural local noun. They varied according to whether the head noun was common gender, marked with the ambiguous determiner *de*, as in (1a) and (1c), or neuter gender, marked with the unambiguously singular determiner *het*, as in (1b) and (1d). In total, 40 items were classified as having a single referent (i.e., notionally and grammatically singular), as in (1a) and (1b), and 40 items were classified as having a distributive referent (i.e., notionally plural but grammatically singular), as in (1c) and (1d).

The 80 complex NPs were split into two experimental lists, such that participants saw 10 items in each of the four conditions. The 40 experimental items in each list were presented along with 112 filler items in a randomised order. Filler NPs included conjoined NPs (e.g., *De krant en het tijdschrift* “The newspaper and the magazine”), and simple singular and simple plural NPs (e.g., *Het groene gordijn* “the green curtain”). Across all items, an equal number of NPs were—prescriptively speaking—grammatically plural and grammatically singular. All experimental and filler NPs were recorded by a female Dutch native speaker.

### Procedure

Participants were tested individually in a quiet room on a computer using E-Prime v2.0 (Psychology Software Tools, 2012), and their responses were digitally recorded. Participants received both aural and written instructions. They listened to each preamble and were instructed to repeat the preamble exactly as they heard it, and to then complete the sentence by describing where the things are (Lorimor, 2007). Participants were instructed to use the copula verb in either the present or past tense (i.e., *is/zijn* “is/are”; *was/waren* “was/were”) and to respond as quickly and fluently as possible. Prior to the experiment, participants were given several example completions (e.g., *in de stad* “in the city,” *op de maan* “on the moon”) and one complete example. Participants received feedback if they repeatedly used a verb besides the copula or if they did not specify a location in their response, but they received no feedback on their repetition of the preamble or use of singular or plural verbs in their completions. Participants completed 10 practice items at the beginning of the task.

For each item, participants saw a fixation point for 500ms and then heard the recorded preamble. Then an exclamation point appeared on the screen, prompting participants to repeat the preamble and complete the sentence. The experimenter advanced between trials manually with a mouse click. Sessions lasted 15–20 min.

After finishing the sentence completion task, participants filled out a language background questionnaire and completed an OSpan task (Engle, 2002; Turner & Engle, 1989). In this task, participants saw a simple math

equation, along with an answer to the equation, on the computer screen, and they had to decide whether the answer provided was correct or not. For half of the items, the provided answer was correct, and for half of the items the provided answer was incorrect. This equation appeared for 3,750ms or until the participant responded. Then a Dutch word appeared on the screen for 1,250ms and participants were instructed to remember this target word. After a set number of equation–word pairs, the word RECALL appeared on the screen, and participants were prompted to type in as many of the target words as they could remember. The number of equation–word pairs in each set increased from two to six pairs as the experiment progressed, with three sets per level.

### Scoring

Responses on the production task were transcribed and coded as singular, plural, or miscellaneous. Miscellaneous responses included instances where the participants did not correctly repeat the preamble or failed to complete the sentence. All singular and plural responses were additionally coded as fluent or disfluent. Fluent utterances consisted of responses in which participants produced the preamble and the verb without any hesitations, filled pauses or repetitions, and used a form of the copula (*is, zijn, was, waren*). Disfluent utterances consisted of responses in which participants hesitated or repeated a portion of the preamble or the verb, or used a verb other than the copula. In instances where participants changed the number marking on the verb upon repetition, we coded the verb as singular or plural based on the first complete verb that was produced. A second coder scored 5% of the data, with an interrater reliability of 97.6%. There were 326 (16.0%) plural responses (226 fluent, 100 disfluent), 1,602 (78.5%) singular responses (1,263 fluent, 339 disfluent), and 112 (5.5%) miscellaneous errors. All responses are presented in Table 1.

On the OSpan task, participants received one point for each word they correctly recalled on the task. Their score then represented the total number of correctly recalled words, as this is a more accurate means of reporting OSpan scores than the largest set-size for which participants recalled all of the target words (Conway et al., 2005).

### Results

All 51 participants included in the results answered at least 65% of the math problems on the OSpan task correctly ( $M = 84.7\%$ ,  $SD = 8.4$ , range = 65.0%–98.3%). These participants recalled, on average, 49.8 words out of a maximum of 60 words on the OSpan task ( $SD = 5.8$ , range = 34–59).

Analyses were conducted using mixed-effect logistic regression models (Jaeger, 2008) with the lme4 package version 1.1–12 (Bates, Mächler, Bolker, & Walker, 2015) in

**Table 1.** Distribution of response type (singular vs. plural verbs) by scoring category and condition.

Response type	Singular (fluent, disfluent)	Plural (fluent, disfluent)	Miscellaneous errors
Single token			
<i>De-de</i>	406 (321, 85)	71 (43, 28)	33
<i>Het-de</i>	449 (367, 82)	41 (28, 13)	20
Multiple token			
<i>De-de</i>	315 (237, 78)	166 (122, 44)	29
<i>Het-de</i>	432 (338, 94)	48 (33, 15)	30

R version 3.2.5 (R Development Core Team, 2016). Number marking on the verb (singular vs. plural) was the dependent variable. Notionality (single-token vs. multiple-token) and morphophonology (*het-de* vs. *de-de*) were entered as fixed effects into the model, contrast coded as  $-.5$  and  $.5$  (Davis, 2010). Given that previous research has shown that agreement errors can vary as a function of working memory (e.g., Hartsuiker & Barkhuysen, 2006; Veenstra et al., 2017), we entered participants' OSpan scores as a fixed effect, centred at the sample mean. To control for any differences in agreement as a function of fluency, we also included whether a given utterance was fluent or disfluent (see Supplemental Materials for an analysis of fluent data only).<sup>1</sup> As there were more fluent than disfluent items, this factor was effect coded and centred (fluent =  $-.38$ , disfluent =  $.62$ ), such that any main effect reflects the average of the two factor levels (Davis, 2010). The final random effect structure was determined by starting with the maximum structure justified by the experimental design, which included random intercepts for subjects and items, correlated by-item random slopes for OSpan, and correlated by-participant random slopes for the main effects of notionality and morphophonology, and their interaction. The random slope for the interaction between notionality and morphophonology was removed due to non-convergence, and random slopes correlated above 0.95 were removed to avoid over-fitting. Finally, we used model comparison to determine the most parsimonious random effect structure (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017), removing individual variables one at a time, based on which variables contributed the least amount of unique variance to the random effect structure. The final random effect structure included random intercepts for participants and items and decorrelated by-item slopes for OSpan. The inclusion of additional random slopes did not significantly improve the model fit (all  $ps > .24$ ). However, the pattern of significant fixed-effects was identical when the maximal random effect structure was used.

As seen in Table 2, there was a significant effect of fluency because participants produced more plural responses in disfluent than fluent utterances. There was a significant effect of notionality because participants produced more

plural responses for multiple-token than single-token items. There was a significant effect of morphophonology because participants produced more plural responses with *de-de* items than *het-de* items. There was also a significant interaction between notionality and morphophonology because, as seen in Figure 1, there were significantly more plural responses for *de-de* items that were notionally plural, compared to the other three conditions.

While there was no significant effect of OSpan, there was a significant two-way interaction between OSpan and morphophonology and a significant three-way interaction between notionality, morphophonology, and OSpan. As seen in Figure 2, participants produced fewer plural responses with *het-de* NPs, regardless of notionality (see also Figure 1). For multiple-token preambles containing *de-de* NPs, participants produced more plural responses regardless of OSpan score, but for single-token preambles containing *de-de* NPs, the proportion of plural responses decreased as OSpan scores increased.

## Discussion

As in Antón-Méndez and Hartsuiker (2010), speakers were sensitive to both notional number and morphophonology, as speakers used more plural agreement with notionally plural items and with *de-de* items. However, unlike Antón-Méndez and Hartsuiker, we found a significant interaction between notional number and morphophonology, as speakers showed greater notional effects with *de-de* NPs than with *het-de* NPs. In addition, we found an effect of fluency, with more agreement errors on disfluent items than on fluent items. Furthermore, while there was no main effect of working memory, as measured by OSpan, there was a significant interaction between OSpan and morphophonology, as participants with lower OSpan scores produced more agreement errors on the *de-de* items. This two-way interaction was qualified by a small, but statistically significant three-way interaction between notionality, morphophonology, and working memory. As seen in Figure 2, while agreement errors were lower on *het-de* items overall (Panel A), differences based on OSpan were most visible on the notionally singular *de-de* items (Panel B).

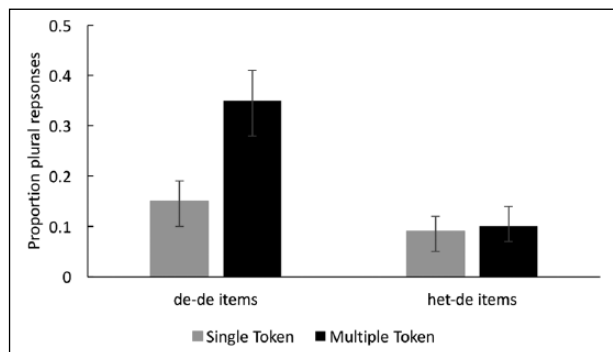
**Table 2.** Summary of the mixed logit model for complex NPs.

Predictor	Parameter estimates		Wald's test of individual coefficients		$\Delta(-2\Delta)$ -test	
	Estimate	Standard error	z-value	Pr (> z )	$\chi^2$ (df)	p
<b>Fixed effects</b>						
(Intercept)	-2.18	0.21	-10.42	<.001		
Fluency	0.46	0.17	2.69	.007	6.90 (1)	.009
Notionality	0.89	0.29	3.10	.002	30.89 <sup>a</sup> (4)	<.001
Morphophonology	1.34	0.15	8.75	<.001	119.77 (4)	<.001
Operation span (OSpan)	0.02	0.03	0.78	.435	8.64 (4)	.071
Notionality × morphophonology	1.06	0.30	3.53	<.001	18.93 <sup>b</sup> (2)	<.001
Notionality × OSpan	-0.02	0.03	-0.54	.589	5.22 (2)	.074
Morphophonology × OSpan	-0.06	0.03	-2.16	.031	8.39 (2)	.015
Notionality × Morphophonology × OSpan	0.13	0.06	2.33	.020	5.19 (1)	.023
<b>Random effects</b>						
	Groups	Name	Variance	SD		
	Participant	Intercept	0.998	0.999		
	Item	Intercept	0.573	0.757		
		OSpan	0.002	0.048		

NP: noun phrase; df: degree of freedom; SD: standard deviation.

<sup>a</sup>Likelihood ratio tests for main effects are based on omitting the main effect and any interaction terms involving that main effect.

<sup>b</sup>Likelihood ratio tests for two-way interactions are based on omitting the relevant two-way interaction term and the three-way interaction term.



**Figure 1.** Proportion of plural responses (error bars represent the bootstrapped by-participant 95% confidence intervals).

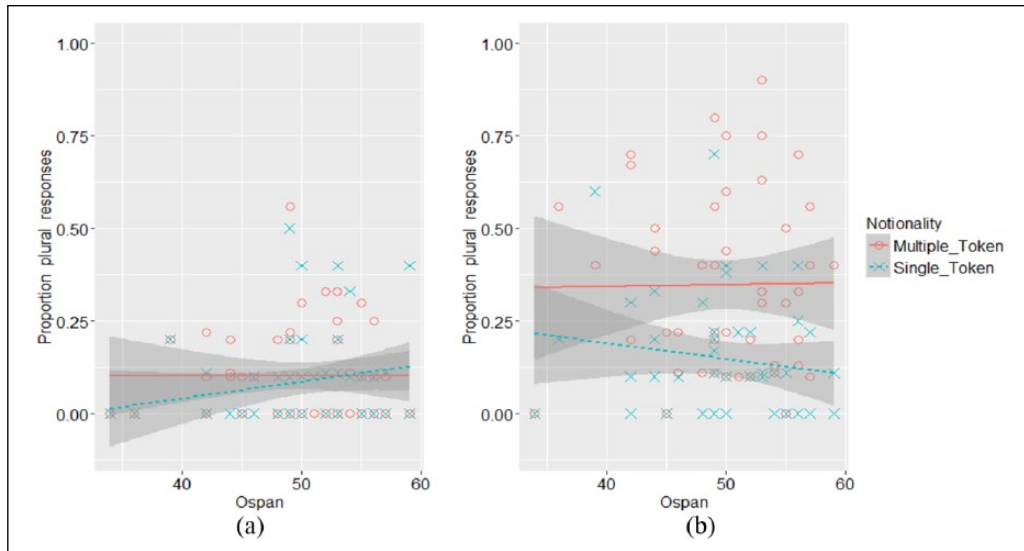
### The role of working memory on agreement production

These results support a model of agreement production that depends, at least to some extent, on working memory. The effects of working memory in this study were strongest in the *de-de* condition, when the determiners had the same form, and were ambiguous for number. The lower rate of plural agreement in the *het-de* condition suggests that the morphophonological cues provided by those distinct determiners reinforced accurate grammatical agreement by helping speakers correctly locate the subject head noun. This result is consistent with the Working Memory retrieval model (Badecker & Kuminiak, 2007), which predicts that morphophonological cues should reduce the load of working memory on agreement production. The fact

that participants made more agreement errors on items with disfluencies provides additional evidence for the role of working memory in agreement production, as disfluencies may be more likely when participants are experiencing difficulties retrieving the subject head noun.

At the same time, working memory is a multifaceted, rather than a unitary construct, and consists of distinct, yet related processes (Logie, 2011; Unsworth, Fukuda, Awh, & Vogel, 2014). Each of these processes—capacity, attentional control, and the ability to retrieve information from secondary memory—can contribute to differences in fluid intelligence and to individual differences in higher order cognitive tasks, like agreement production. Vandierendonck, Loncke, Hartsuiker, and Desmet (2018) provides insight into which aspects of working memory affect agreement production by showing a link between executive control abilities and agreement errors, using a gaze-contingent tone discrimination task. It is important to extend this line of research to consider, more precisely, how the mechanisms underlying working memory correlate with language production if we want to better understand the potential causal relationship between working memory and language production.

Furthermore, it is critical to replicate the current three-way interaction via future research, both in Dutch and in other languages, as interactions may be under-powered and thus susceptible to both Type I and Type 2 errors. The three-way interaction in this study—while statistically reliable according to conventional hypothesis testing criteria—was arguably small. To provide additional evidence about the reliability of the reported three-way interaction between notional number, morphophonology, and OSpan,



**Figure 2.** Proportion of plural responses as a function of operation span score for (a) *het-de* items and (b) *de-de* items. The proportion of plural responses for multiple token items is represented by circles and the solid regression line. The proportion of plural responses for single-token items is represented by crosses and the dotted regression line.

we conducted the same analyses reported above with Bayesian mixed effects models using *rstanarm* (Stan Development Team, 2017). Following the procedures and benchmarks by Nicenboim and Vasishth (2016), we considered there to be strong evidence of an effect for any parameter where the 95% credible interval did not include zero. According to the criteria in Nicenboim and Vasishth (2016), if the edge of the 95% credible interval overlaps with zero, there can be weak evidence for an effect if the probability of the effect is still relatively high. This Bayesian model showed the same general pattern of results as the ones that used *lme4*, and the 95% credible interval for the parameter for the three-way interaction was (0.00–0.02). Therefore, as the lower bound of this three-way interaction overlapped with zero, the Bayesian analysis suggests there is weak evidence for this interaction. When combined with the primary analysis, we consistently find evidence for this three-way interaction, suggesting that it is a small but reliable effect. Further research, especially if conducted with larger sample sizes, is especially important because understanding the role of individual differences can lend important insights into the mechanisms behind language processing (Kidd, Donnelly, & Christiansen, 2018).

### *The interaction between notional number and morphophonology*

The interaction between notional number and morphophonology also lends important insight into the role of morphophonology on number agreement. The reduction in plural agreement on *het-de* items, compared to *de-de* items, is consistent with findings from previous work

(Antón-Méndez & Hartsuiker, 2010; Hartsuiker et al., 2003). However, the impact of morphophonology on agreement in our study was modulated by notional number because notional effects were greater on the *de-de* items, compared to the *het-de* items. The fact that notional effects were greater among *de-de* items, compared to *het-de* items, provides additional evidence that distinct determiners facilitate correct retrieval of the subject head noun, which can lead to smaller notional effects among *het-de* items.

There are several reasons that we may have found an interaction between notional number and morphophonology, while an interaction between notional number and morphophonology was not observed in Antón-Méndez and Hartsuiker (2010). First, we included disfluent utterances in our statistical analysis, while Antón-Méndez and Hartsuiker treated those as miscellaneous items, because we were concerned that discarding disfluent utterances would eliminate many of the sentences in which speakers were having trouble retrieving the subject head noun. Including these items increased our statistical power. However, the inclusion of these disfluent items in our analyses did not change the pattern of effects, and a significant interaction between notional number and morphophonology emerged even in the analysis of fluent-only data (see Supplementary Material A). Second, we had more statistical power to detect an interaction because we tested more participants (51 participants vs. 36 participants in Antón-Méndez and Hartsuiker). Third, participants used more plural agreement in our study overall than in Antón-Méndez and Hartsuiker, which increased our ability to detect an interaction. This could have been due to differences in population, or to the differences in task, as we used an auditory sentence completion task, while Antón-Méndez and



Hartsuiker used a visual completion paradigm. Regardless of the cause, the rate of plural agreement in Antón-Méndez and Hartsuiker was lower than in most studies on Dutch agreement (see Table 5 in Antón-Méndez and Hartsuiker), which would limit their ability to detect an interaction.

### Cue-based retrieval in agreement production

The role of morphophonology in agreement and how it interacts with OSpan and notional number provides important insights into the mechanisms that drive the production of subject–verb agreement. We, like other studies on Dutch (Antón-Méndez & Hartsuiker, 2010; Hartsuiker et al., 2003), found lower rates of agreement errors on *het-de* items, compared to *de-de* items. Within the Marking and Morphing model (Eberhard et al., 2005), this is explained by the unambiguously singular morphology on the determiner *het*, which lowers the grammatical number value for the whole NP. What Marking and Morphing cannot explain, however, is individual differences in working memory or why working memory effects should be most evident on the *de-de* items, compared to the *het-de* items. In contrast, a cue-based retrieval account provides a straightforward explanation, both for why there are fewer agreement errors among *het-de* items overall and also for why effects of OSpan should be most evident among *de-de* items. This is because the distinct retrieval cues provided by the number-marked determiners would help speakers retrieve the correct subject noun at the point of producing a verb.

We propose, similar to Lorimor et al. (2015), that agreement can be affected by both number encoding processes and subject retrieval processes. During number encoding, agreement attraction errors and notional number agreement may be generated through a process like the one described in Marking and Morphing (Eberhard et al., 2005), in which grammatical number values (that may lead to attraction) are combined with notional number valuations to obtain a number specification that is encoded on the subject NP. In addition to these processes involved in number encoding, we propose that there is a later retrieval process during which speakers check the number of the verb with the number on the subject head noun that they are holding in content-addressable memory, and that this retrieval process can be facilitated by the presence of distinct morphophonological cues. Although there are reasons to treat production as separate from comprehension (Tanner, Nicol, & Brehm, 2014), this proposal is similar to recent proposals for how cue-based retrieval would work in agreement comprehension (Nicenboim, Engelmann, Suckow, & Vasishth, 2017; Schlueter et al., 2018). Furthermore, it is consistent with evidence that, even in comprehension, number information is processed both during NP encoding and at the point of encountering the verb (Vandierendonck et al., 2018).

A three-component agreement model that includes notional number, agreement attraction, and cue-based retrieval (as outlined in Lorimor et al., 2015) accounts for the present data as follows. During encoding, some agreement attraction errors were created by spreading activation from the grammatical features on the local noun and from other number features within the subject NP (as outlined in Eberhard et al., 2005), which explains why there are agreement errors in all four experimental conditions. In addition, notional number information on the subject NP explains the main effect of notional number and why participants were more likely to use plural verbs with multiple-token, compared to single-token items. At the point of planning the verb, an additional step of cue-based retrieval reduced the rate of plural agreement with the *het-de* items because there was less similarity-based interference on the nouns in the subject NP. In the *de-de* condition, while participants attempted to retrieve the subject head noun, the fact that the determiners on the head and local nouns had the same form made the process of cue-based retrieval more difficult. Therefore, while participants with higher OSpan were better able to maintain the subject head noun in content-addressable working memory and establish accurate subject–verb agreement, participants with lower OSpan scores were less accurate in their retrieval processes on the *de-de* items. It is possible that interactivity was an additional factor influencing our results (as discussed by Antón-Méndez and Hartsuiker, 2010), as feedback from the phonology on the *de-de* items increased the notional effects on the multiple-token items. However, our results can also be explained through cue-based retrieval, and the existence of interactivity in the language system does not preclude the possibility of an additional step of retrieval. Furthermore, interactivity cannot account for the fact that working memory affected agreement errors among the *de-de* items, but not the *het-de* items.

### Conclusion

By manipulating morphophonology and notional number in Dutch, while at the same time collecting data on individual participants' working memory, we provide evidence about how working memory affects agreement attraction in Dutch, even in the absence of a secondary task. In all, these findings support a model of agreement production that incorporates notional number, agreement attraction, and cue-based retrieval as three distinct—yet related—mechanisms. We also provide evidence for a limited role of working memory in the agreement production process. In doing so, we show that agreement is not fully automatic, and that working memory does play a role in ensuring correct agreement; however, at the same time, we show that features of the language itself, such as distinct morphophonological cues, play an important role in facilitating correct subject–verb agreement.

## Acknowledgements

Portions of this work were presented at the 31st Annual CUNY Sentence Processing Conference. The authors would like to thank Maartje Dona and Anouk Raeven for their assistance with data collection. They also thank the reviewers for their helpful comments and suggestions.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

This project was particularly supported by NSF grants BCS-1349110 and DUE 1561660.


## Supplementary material

Supplementary material is available at journals.sagepub.com/doi/suppl/10.1177/1747021818771887.

## Note

1. In an initial analysis, we also included fluency as a full factor (i.e., including fluency not just as a main effect but also with the complete set of interaction terms). However, the inclusion of these additional interaction terms did not improve the model fit,  $\chi^2(6)=6.15$ ,  $p=.522$ , so they are not included in the final model reported here.

## ORCID iD

Heidi Lorimor  <https://orcid.org/0000-0002-2241-2060>

## References

- Antón-Méndez, I., & Hartsuiker, R. J. (2010). Morphophonological and conceptual effects on Dutch subject-verb agreement. *Language and Cognitive Processes*, 25, 728–748.
- Badecker, W., & Kuminiaik, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, 56, 65–85.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, 31, 99–127.
- Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology*, 43, 83–128.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*, 12, 769–786.
- Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, 25, 1–18.
- Davis, M. J. (2010). Contrast coding in multiple regression analysis: Strengths, weaknesses, and utility of popular coding structures. *Journal of Data Science*, 8, 61–73.
- Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, 112, 531–559.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19–23.
- 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: Human Experimental Psychology*, 47, 437–464.
- Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes*, 23, 329–374.
- Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes*, 21, 181–204.
- Hartsuiker, R. J., Schriefers, H. J., Bock, K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory and Cognition*, 31, 1316–1326.
- Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject-verb agreement. *Journal of Memory and Language*, 48, 760–778.
- Jaeger, T. F. (2008). Categorical data analyses: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Kemper, S., Herman, R. E., & Lian, C. H. T. (2003). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech or noise. *Psychology and Aging*, 18, 181–192.
- Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, 22, 154–169. doi:10.1016/j.tics.2017.11.006
- 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.
- Logie, R. H. (2011). The functional organization and capacity limits of working memory. *Current Directions in Psychological Science*, 20, 240–245. doi:10.1177/0963721411415340
- Lorimor, H. (2007). *Conjunctions and grammatical agreement* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.
- Lorimor, H., Bock, K., Zalkind, E., Sheyman, A., & Beard, R. (2008). Agreement and attraction in Russian. *Language and Cognitive Processes*, 23, 769–799.
- Lorimor, H., Jackson, C. N., & Foote, R. (2015). How gender affects number: Cue-based retrieval in agreement production. *Language, Cognition and Neuroscience*, 30, 947–954.
- MacDonald, M. C. (2016). Speak, act, remember: The language-production basis of serial order and maintenance in verbal memory. *Current Directions in Psychological Science*, 25, 47–53.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315.
- Nicenboim, B., Engelmann, F., Suckow, K., & Vasishth, S. (2017). *Number interference in German: Evidence for cue-based retrieval*. Retrieved from psyarxiv.com/u2kqg

- Nicenboim, B., & Vasisht, S. (2016). Statistical methods for linguistics research: Foundational ideas: Part II. *Language and Linguistics Compass*, 10, 591–613.
- Psychology Software Tools. (2012). E-Prime 2.0. Available from <http://www.pstnet.com>
- R Development Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from <http://www.R-project.org>
- Schlueter, Z., Williams, A., & Lau, E. (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:10.1016/j.jml.2017.10.002
- 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, 1503–1514.
- Slevc, L. R., & Martin, R. C. (2016). Syntactic agreement attraction reflects working memory processes. *Journal of Cognitive Psychology*, 28, 773–790.
- Stan Development Team. (2017). *rstanarm: Bayesian applied regression modeling via Stan*. R package 2.15.3. Available from <http://mc-stan.org>
- 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.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154.
- Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, 71, 1–26. doi:10.1016/j.cogpsych.2014.01.003
- Vandierendonck, A., Loncke, M., Hartsuiker, R. J., & Desmet, T. (2018). The role of executive control in resolving grammatical number conflict in sentence comprehension. *The Quarterly Journal of Experimental Psychology*, 71, 759–778. doi:10.1080/17470218.2016.1276610
- Veenstra, A., Antoniou, K., Katsos, N., & Kissine, M. (2017). The role of executive control in agreement attraction in monolingual and bilingual children. In M. LaMendola & J. Scott (Eds.), *Proceedings of the 41st Annual Boston University Conference on Language Development* (pp. 706–717). Somerville, MA: Cascadia Press.
- Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes*, 11, 407–442.
- Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology*, 9, 2. doi:10.3389/fpsyg.2018.00002