

# UC Santa Cruz

## UC Santa Cruz Previously Published Works

### Title

Memory for linguistic features and the focus of attention: evidence from the dynamics of agreement inside DP

### Permalink

<https://escholarship.org/uc/item/95z5v033>

### Journal

Language Cognition and Neuroscience, 37(9)

### ISSN

2327-3798

### Authors

Wagers, Matthew  
McElree, Brian

### Publication Date

2022-10-21

### DOI

10.1080/23273798.2022.2057559

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

**Memory for linguistic features and the focus of attention: evidence from the dynamics of agreement inside DP**

Matthew Wagers<sup>a\*</sup> and Brian McElree<sup>b</sup>

*<sup>a</sup>Department of Linguistics, University of California, Santa Cruz, USA; <sup>b</sup>Department of Psychology, New York University, New York, USA*

*WAGERS\**

*Department of Linguistics,*

*University of California Santa Cruz, Santa Cruz, CA 95064, USA*

*mwagers@ucsc.edu*

*\*Corresponding Author*

*MCELREE*

*Department of Psychology,*

*New York University, New York, NY 10003, USA*

*brian.mcelree@nyu.edu*

## **Memory for linguistic features and the focus of attention: evidence from the dynamics of agreement**

The amount of information that can be concurrently maintained in the focus of attention is strongly restricted (Broadbent, 1958). The goal of this study was to test whether this restriction was functionally significant for language comprehension. We examined the time course dynamics of processing determiner-head agreement in English demonstrative phrases. We found evidence that agreement processing was slowed when determiner and head were no longer adjacent, but separated by modifiers. We argue that some information is shunted nearly immediately from the focus of attention, necessitating its later retrieval. Plural, the marked feature value for number, exhibits better preservation in the focus of attention, however, than the unmarked value, singular.

Keywords: language comprehension, grammatical agreement, number, working memory, focus of attention, speed-accuracy tradeoff (SAT)

## Introduction

Language comprehension requires the coordination of information at different levels of analysis and from different segments of an expression. Successful interpretation thus depends on working memory resources to access and manipulate recently encoded information. However, working memory resources are capacity constrained (Broadbent, 1958, Cowan, 1995, Jonides, et al. 2008). An important dimension of this capacity is the scope of information that is directly and concurrently accessible to cognitive processes. Such information is said to be in the focus of attention. Data from several paradigms suggests that the number of representations that can occupy the focal state is severely limited, restricted to perhaps only one chunk (McElree, 2006, Oberauer & Bialkova, 2009, cf. Cowan et al. 2012). Speed-of-access measures consistently indicate a dichotomy between the rates at which focal information and non-focal information can influence processing (Garavan, 1998, McElree & Doshier, 1989, McElree, 1998, 2001, 2006, Oberauer, 2002, Oberauer & Bialkova, 2009, Unsworth & Engle, 2009, Verhaeghen, Cerella, & Basak, 2004, Verhaeghen & Hoyer, 2007, Zhang & Verhaeghen, 2009; cf. Vergauwe et al. 2016). fMRI experiments reveal that distinct neural substrates mediate access to focal versus non-focal contents (Nee & Jonides, 2008, Öztekin, et al. 2008, Öztekin, Davachi, & McElree, 2010). The focus of attention thus partitions encodings into privileged and non-privileged access sets. It is therefore likely to play an important functional role in language comprehension. In particular, if information that has been displaced from the focus of attention is necessary to complete current processing, it must be reinstated by memory retrieval processes. These processes require time and are subject to error. For example, they are liable to similarity-based interference (Anderson & Neely, 1996). Such interference has been shown to affect incremental sentence comprehension, by demonstrations that grammatically-inaccessible constituents can

impact the retrieval of target constituents if the two constituents are similar along a linguistically-significant dimension (Gordon, et al. 2001, 2002, 2004, 2006, Van Dyke & Lewis, 2003, Van Dyke & McElree, 2006, Vasishth, Brüssow, Lewis, & Drenhaus, 2008, Wagers, Lau & Phillips, 2009, Dillon et al. 2013, Jäger, Engelmann & Vasishth, 2017, Villata & Franck, 2020).

Understanding the interaction between the focus of attention and retrieval is crucial for formulating accurate models of real-time language processing. The factors that determine the kind and extent of linguistic information which can be concurrently maintained are not well understood. Initial research based on speed-accuracy tradeoff analysis has demonstrated that the analysis of a single embedded clause can displace the current contents of focal attention (McElree, 2001, McElree, et al. 2003). However, because clauses constitute such large linguistic domains, those findings likely do not provide strong enough constraints on our models of incremental language processing. The present study probes information inside a relatively small linguistic domain: the number features associated with a determiner phrase (DP), which is the syntactic constituent comprised of a noun, its complements and modifiers, and associated functional categories.

There are two major results. Firstly, we find that the intervention of a single word between a determiner and a noun within DP can trigger displacement of number information from focal attention. Secondly, we find that displacement may be contingent on the value of the number feature, such that the marked plural feature value is more likely to survive than the default singular value. Language comprehension thus recruits short-term memory processes for linguistic analyses well below the clause level. Its strategies for managing maintenance and retrieval are likely to be closely influenced by properties of the linguistic feature system.

### *Identifying the contents of the focus of attention*

A key diagnostic for whether an item occupies focal attention is the speed with which it is accessed in some task. McElree and Doshier (1989) reported that item recognition in list-memory tasks occurs at a uniform rate for all but the last item in the list, which is recognised 40-50% faster. McElree and Doshier (1989) used the speed-accuracy tradeoff paradigm (SAT) to make this estimate. SAT is a response-signal method in which participants are trained to give their response to a tone cue following stimulus presentation. The lag between stimulus onset and tone cue is systematically manipulated in the experiment in order to derive the full time-course of accuracy as a function of processing time. Measuring these functions is crucial because estimates of process dynamics based on free response time alone will confound item accessibility and strength (Wickelgren, 1976). Figure 1 illustrates ideal and actual SAT data with their interpretation. Panels A & B demonstrate that individual conditions could differ independently in either accuracy (A) or speed (B). Panel C contains data from McElree and Doshier (1993) and demonstrates the extreme dynamics advantage for the item in focal attention. The most recently encountered item, serial position 6, is conspicuously distinct from the rest: not only does it attain higher accuracy at test, but the SAT function achieves asymptotic accuracy at a much faster proportional rate (on average: 306% faster). Serial positions 1-4, on the other hand, only show differences in ultimate accuracy. These data suggest that the last list item can enter the processing stream more quickly than all the rest. McElree (2006) reviews an array of findings showing that the rate advantage is determined neither by physical stimulus overlap nor by presentation recency per se. For example, the same advantage obtains in a rhyme or synonym judgment task as obtains in simple item recognition. The item in focal attention can be

systematically manipulated via either an  $n$ -back task or controlled rehearsal so that the advantage accrues to a non-final item in a list.

<< Insert Figure 1 here >>

What can count as an ‘item’ for focal attention is the crucial question (Oberauer & Bialkova, 2009). As Jonides et al. (2008) observe, most cognitive tasks require the coordination of multiple pieces of information. Therefore, it is likely misleading to refer to the contents of focal attention as an ‘item’ in the sense of a single, irreducible atom. Instead, they argue, it is more plausibly conceived of as a single, functional data structure. Some evidence bearing on this question come from McElree (1998), which demonstrated that list structure can impact the units to which a dynamics advantage accrues. Experimental participants were encouraged to chunk a 9-word list as 3 triplets according to superordinate category (e.g., animals, furniture, and vehicles). In this case, the focal advantage is observed for the most recently processed three words belonging to the same category, not simply the last word.

### ***The relation between the focus of attention and language comprehension***

How to define an ‘item’ with respect to focal attention is an especially salient question for language, given the rich, hierarchical structure of expressions and the fact that many important linguistic dependencies are non-adjacent. A long psycholinguistic tradition has aimed to relate the comprehensibility and acceptability of sentences to how they are managed in working memory (Miller & Chomsky, 1963). At issue is how compositional representations like phrase-structures are segmented and whether a systematic relationship exists between the ‘grain-size’ of segmentation and the system’s capacity limitations. Click dislocation experiments were an early influential attempt to answer this question (Fodor & Bever, 1965, among others, see Townsend

& Bever, 2001, for a review). Other researchers applied word recall techniques (Jarvella, 1971) or whole-sentence recognition (Sachs, 1967, Potter & Lombardi, 1992). These studies all identified clause membership as an important determinant of what constituents were immediately or faithfully accessible. J.A. Fodor, Bever & Garrett (1974) proposed that clause boundaries triggered perceptual segmentation while Carroll & Tanenhaus (1978) argued boundaries below the clause could be sufficient. More generally, they argued that segmentability fell along a cline that was not deterministically related to fixed boundaries. In either case, however, fairly large grammatical domains were thought to delimit the memory encodings.

It soon became clear that syntactic, semantic and pragmatic processes quickly and intimately interweave as single words are incorporated into a sentence (Marslen-Wilson, 1975). That expressions are interpreted incrementally, in step-sizes much smaller than a clause, suggested that compositional structure must initially be encoded in much smaller chunks (Frazier & J.D. Fodor, 1978). However, there have been few direct empirical attempts since the 1960s to measure the scope of structured encodings within a clause. Working memory constraints have still played an active role in theory-building (e.g., Gibson, 1998, 2000, Just & Carpenter, 1991, Lewis, 1996), but inferences about encoding accessibility have usually been indirectly based on the moment-to-moment complexity measures derived from reading-time studies. Despite its important architectural role in basic theories of working memory, the focus of attention has played a relatively minor role in comprehension data and theory. McElree (2001) used the SAT technique to measure the accessibility of subject phrase features before and after modification of the subject by a relative clause. In that study clausal modification was found to be sufficient to displace at least some features of the subject phrase, a finding which was subsequently replicated in McElree, Foraker, and Dyer (2003) and Johns, Matsuki, & Van Dyke (2015). Franck and



Wagers (2020) recently found in a probe recognition study of jabberwocky French that subjects head nouns persisted in the focus of attention if they were unmodified, or if they were only modified by a PP. Finally, Foraker and McElree (2007) tested the hypothesis that the focus of attention was closely related to devices of linguistic focus (Deane, 1991, Gundel, 1998, among others), but found no evidence from at least one English construction, the *it*-cleft (but cf. Kush, Johns, & Van Dyke, 2019). Together these studies provide rough boundaries for what linguistic information the focus of attention does and does not encompass.

Lewis and Vasishth (2005) incorporated the focus-of-attention capacity constraint [in their ACT-R model of comprehension](#) by restricting maintenance of parsing goals to a single encoding. The authors assumed a phrase-sized chunk based essentially on the concept of maximal projection from X'-theory (Jackendoff, 1977, Chomsky, 1981). This decision was linguistically well-motivated and grounded in the learning mechanisms of ACT-R (Anderson & LeBiere, 1998). It is a reasonable theoretical appeal, but there remains an extreme paucity of data bearing on the question directly. Thus, an important goal of this study is to expand that database.

### ***Within-DP agreement and the present study***

The present study tests the hypothesis that grammatical information contained within a domain much smaller than a clause can be displaced from focal attention. We tested the agreement relation between a determiner and the noun with which it combines. In English, determiners and nouns only show overt agreement in cases of the demonstratives *that/those* and *this/these*. Accordingly, the phrase 'that clever monkey' is acceptable, but 'that clever monkeys' is not. Agreement relations are good candidates to test the scope of focal attention. Theoretically, agreement is significant because it is the surface manifestation of a more abstract grammatical relation between two categories (Adger & Harbour, 2008). This relation is typically reflected in

the morphological covariation of one category with another. This morphological covariation could be used to support the recovery of grammatical dependencies by providing the comprehenders with cues based on shared feature content. This benefit to comprehension might provide functional pressure to maintain agreeing features in the focus of attention. However such cues are often also provided by other sources of information, like word order. Agreeing features could also therefore be a good candidate for information that can instead be shunted from the focus of attention and only retrieved later if necessary. It is thus hard to predict in advance whether, or the extent to which, agreeing features would be maintained in the focus of attention.

A large body of research has focused on subject-verb agreement, where significant observations have come from the study of agreement attraction (Bock & Miller, 1991). Agreement attraction is illustrated in a sentence like *The phone by the toilets were out of order*. Although the grammatical controller of agreement is the entire singular DP, ‘the phone by the toilets’, the verb shows agreement with a grammatically inaccessible plural phrase ‘the toilets.’ Agreement attraction can be robustly elicited in language production tasks (Eberhard, Cutting & Bock, 2005). In its comprehension analogue, comprehenders fail to notice misagreement in an agreement attraction sentence, reflected in smaller RT disruptions (Pearlmutter, et al. 1999) or more directly in speeded acceptability judgments (Wagers, Lau, & Phillips, 2009). The fact that comprehenders failed to accurately match the number features on the verb with the number features of the subject suggest that the number features on the subject somehow lost prominence as intervening phrases were processed. Agreement attraction can be understood as a type of retrieval interference, both in production (Badecker & Kuminiak, 2007, Badecker & Lewis, 2007) and in comprehension (Wagers, Lau, Phillips, 2009, Dillon et al., 2013). If grammatical number features are especially prone to retrieval interference, then this finding suggests that they

may be a good target for studying the focus of attention: the fact that they must be sometimes retrieved in sentences with relatively simple subjects like “the phone by the toilets” suggests that, at least for such constituents as subjects, number is not always maintained in the focus of attention (cf. King, 2021).

Within-DP agreement has not figured heavily in psycholinguistic research on English for the simple reason that agreement is only apparent in English demonstrative DPs. Many other languages show more productive gender and number agreement inside DPs and this agreement has an effect on incremental comprehension. For example, in Finnish, within-DP modifier-head case agreement has been shown to have a facilitating effect on syntactic processing (Vainio, Hyönä, & Pajunen, 2008). In a variety of Germanic and Romance languages, electrophysiological studies have concluded that within-DP agreement mismatches are detected incrementally for both gender and number (Barber & Carreiras, 1995, Gunter, et al. 2000, Hagoort & Brown, 1999), in a way that interacts with case assignment (Davidson, Hanulíková & Indefrey, 2012). A prime advantage to studying within-DP agreement is that length can be easily manipulated by iterating modifiers between the determiner and noun.

However there are important differences between subject-verb agreement, undoubtedly the most investigated kind of agreement in language processing research, and within-DP agreement. The latter, also referred to as *nominal concord*, exhibits a number of critical differences from subject-verb agreement. Norris (2014, 2017) provides a comprehensive overview of these differences, many of which are not apparent in English. For example, nominal concord can have multiple exponents throughout the DP, such as a determiner and adjectives. These exponents can occur in essentially all phrase structure positions (head, specifier, adjunct). The presence of nominal concord is not typically dependent on grammatical case in the way that

subject-verb agreement often is. Finally, the features that control nominal concord may differ from the ones that control subject-verb agreement, and languages can exhibit a morphosyntactic/semantic split (Corbett, 1979, Wechsler & Zlatić, 2000). For example, varieties of British English show this split with collective nouns like *committee* or *team*. While these nouns (can) determine plural agreement on the verb, they nonetheless require a singular demonstrative. Compare an acceptable sentence in this variety of English, “This committee are deciding on a solution”, to its unacceptable counterparty “\*These committee are deciding on a solution” (Elbourne, 1999, Smith, 2017). In other languages like Lebanese Arabic (Ouwayda, 2013) or Russian (Pesetsky, 2013), higher projections within DP, like adjectives, can effectively “take over” agreement from the head noun itself and determine how the DP behaves externally; this shows that agreement need not always be strictly between a determiner and a noun. In sum, we should not expect within-DP agreement to recruit identical processing mechanisms to subject-verb agreement, because the grammatical rules by which features are shared within a DP are not the same as those by which they are shared outside a DP. We can still make use of DPs and their properties in English as a site to test whether or not such a small syntactic domain nonetheless engages working memory. However, the findings may not necessarily apply to other species of agreement, like subject-verb agreement.

We used the SAT methodology to measure the speed with which comprehenders processed number agreement in English demonstrative DPs embedded in sentence contexts. We adopted determiner-noun adjacent phrases, like ‘that monkey’, as a baseline case in which the number information of both *that* and *monkey* is maximally likely to be co-present in focal attention. We then interrupted the adjacency of determiner and noun by inserting one or two modifiers. SAT time courses were derived by scaling the rate of endorsement for the

grammatical ‘... that monkey’ (a hit in signal detection theory terms) against ‘... that monkeys’ (a false alarm) at each of 17 lags following the sentence-final presentation of ‘monkey’. If modifiers lead to the number feature’s displacement from focal attention, then the rate at which agreement contrasts are discriminated should slow from the adjacent to the non-adjacent conditions. If, on the other hand, the number feature can be maintained across the modifiers, then the rates should remain the same across all DP sizes. If the demonstrative’s number feature is displaced from focal attention in larger DPs – and thus has to be recovered via retrieval – we can make the further prediction that asymptotic accuracy will decline for the non-adjacent conditions, akin to what McElree, Foraker & Dyer (2003) found for longer *wh*-dependencies.

## **Materials and Methods**

### ***Participants***

22 members of the NYU community were recruited to participate. All were self-identified native English speakers. Participants received \$10/hr for 11 1-hour experimental sessions. 4 participants were excluded for failure to learn the task.

### ***Materials***

We created 40 sets of DPs with the English demonstrative determiners (‘this’/‘that’). Within each set, a  $2 \times 3 \times 2$  design crossed number, determiner-noun distance, and grammaticality, as shown in Table 1. For distance, the determiner was either string-adjacent to the noun, separated by one hyphenated participle (‘risk-taking’), or by a two-modifier sequence (‘clever, risk-taking’). Use of proximal (‘this’/‘these’) versus distal (‘that’/‘those’) demonstratives was counterbalanced across items.

<< Insert Table 1 here >>

Six control conditions were added which contained the same adjective sequences, but continued with head nouns whose plausibility was varied (e.g., ‘the risk-taking burglars’ v. ‘the risk-taking jewels’). Since the modifiers contribute meanings to be composed with the head noun, it is plausible that this would incur a cost even in the target agreement contrasts described above. The plausibility controls were included to allow us to gauge the potential extra cost of additional modifiers, independent of agreement. These phrases were headed by the number-ambiguous English definite determiner ‘the’ and noun number was counterbalanced across items. The use of identical modifiers prevented participants from guessing in advance that they would have to pay special attention to number, as the control conditions contained no number violations<sup>1</sup>.

Each item set’s 18 phrases were embedded in sentence final position. To avoid the determiner/complementiser ambiguity of ‘that’, each phrase was the complement either of a preposition or of a verb that could not embed a clause headed by the complementiser ‘that’. Six possible preambles were written for each item and randomly assigned to 3 conditions within the set. Preambles varied in length from 4 to 13 words (median length: 8 words), and included common names and descriptions linked to animate referents. There was no control for lexical frequency of words in the preambles. The resulting 720 sentences were combined with items from four concurrently-run experiments. These experiments were related to other topics, and

---

<sup>1</sup> As an anonymous reviewer points out, it is conceivable that comprehenders could (learn to) preactivate different kinds of information in the number-marked demonstratives compared to the simple *the*-conditions. Agreement is never at stake between *the* and the NP it combines with, which means that other effects, like plausibility, could be more pronounced. In other words, the cost of modification might vary depending on what other relations have to be computed. An alternative design that only used demonstratives could be valuable here. But we agree generally with the reviewer that there is a complex interaction between prediction and integration at all levels, one which is likely to be affected by language-general properties and but also by experiment-specific statistics.

included acceptability contrasts based on several different dimensions including transitivity (“the hose drained/\*the driver drained”) and animacy (“the sergeant complained/\*the sword complained”). Participants were trained via examples in the first session of the several ways a sentence could be unacceptable. In total, the 2460 sentences from the five experiments were presented to participants evenly distributed across 10 lists in 10 sessions.

### ***Procedure***

Participants were recruited to visit the lab for 11 sessions. An initial practice session preceded the 10 experimental sessions.

The multiple-response SAT procedure was used to estimate accuracy as a function of time (Wickelgren, Corbett & Doshier, 1980, Martin & McElree, 2008). Participants were trained to read sentences and discriminate them on the basis of their acceptability by giving a series of 17 responses, each cued by a tone. Participants responded after each tone by pressing and releasing a response key or keys; they were not allowed to hold down one response throughout the series. They responded initially by pressing the ‘acceptable’ and ‘unacceptable’ response keys simultaneously, and then switched to either the ‘acceptable’ or ‘unacceptable’ response as soon as any confidence developed in that alternative. They were trained to modulate their response, as their opinion and degree of confidence could change over the response series. Participants received feedback (“too slow”) if they failed to respond with 200 ms of each tone.

Trials began with a 1-second fixation cross in the center of the display. Sentences were presented word-by-word in rapid serial visual presentation mode (Potter, 1988). Each of the two modifiers was presented as its own word. In the two-modifier condition, a comma was presented on the same screen as the preceding word. For example, a critical DP in the two-modifier

conditions would have been presented as follows: “this || clever, || risk-taking || monkey” (RSVP breaks indicated by ‘||’). The stimulus onset asynchrony of each chunk varied by word length according to the formula  $SOA(\text{char}) = 190\text{ms} + 25 \text{ ms/char}$ , with a maximum SOA set at 400 ms. The inter-stimulus interval was constant at 100 ms, except before the last word when it was lengthened to 300 ms. 200 ms before the onset of the final word onset, the tone series began. Tone frequency was 1000 Hz, tone duration was 50 ms, and tone SOA was 300 ms. These parameters were based on a previously published study that applied MR-SAT to sentence processing, Martin & McElree (2008).

*Analysis.* Accuracy per response tone was transformed into a discriminative  $d'$  score by scaling the hit rate in each grammatical condition against the false alarm rate in the corresponding ungrammatical condition (MacMillan & Creelman, 1991). In addition, a common  $d'$  was calculated by scaling the false alarm rates against a common hit rate, which was derived by pooling responses in the grammatical conditions. Lag-latency was calculated by adding the average response time at each response tone to tone latency. The resulting  $d'$ /lag-latency series was fit by a saturating, shifted exponential function in EQ(1):

$$\begin{aligned} d' &= \lambda(1 - e^{-\beta(t-\delta)}) & , t > \delta, \\ d' &= 0 & , \text{otherwise} \end{aligned} \quad (\text{EQ1})$$

This function is described by three parameters: an asymptote,  $\lambda$ ; a rate,  $\beta$ ; and an intercept,  $\delta$ .

The  $\lambda$  parameter describes maximum achieved accuracy. The speed of processing is jointly captured by the  $\beta$  and  $\delta$  parameters. The value of  $\beta$  is the time at which accuracy reaches a common proportion of asymptotic accuracy, namely  $(1 - e^{-1})$  or approximately 63%, a value which corresponds to the rate at which information accrues. The value of  $\delta$  is the amount the curve is



shifted from the ordinate axis, reflecting the moment when discriminative information is first available.

A hierarchical model fitting procedure was followed to explore the parameter space and the best-fitting functions for both average  $d'$  scores and individual participant data. The goal of the model-fitting analysis is to determine what the best-fitting, most parsimonious set of parameters is for describing the experimental conditions (see McElree, Foraker, & Dyer, 2003, for extensive discussion). A fully-saturated model assigns a  $\lambda$ - $\beta$ - $\delta$  triple to each condition, while a null model assigns the same  $\lambda$ - $\beta$ - $\delta$  parameters to all conditions. Following the recommendations in Liu and Smith (2009), we selected the 10 best-fitting models on the basis of both the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC), goodness-of-fit measures calculated from each model's likelihood score. For either measure, better fitting models have lower scores. Both AIC and BIC were translated into model weights, a normalised probability estimate that a particular model is best (Glover & Dixon, 2004). All models incorporated 6- $\lambda$  (asymptote) parameters and a single  $\delta$  (intercept) parameter. Exploratory data analysis revealed that for any  $\beta$ - $\delta$  parameterization, a 6- $\lambda$  model outperformed all others. Furthermore, repeated measures ANOVA over participants' empirical asymptotes (average  $d'$  for tones 15-17) supported this decision, showing a reliable distance effect ( $F_1(2,34) = 31.06$ ,  $MSE = 2.31$ ,  $p < .001$ ) and a marginal interaction with number ( $F_1(2,34) = 2.65$ ,  $MSE = 0.11$ ,  $p < .10$ ). The parameter  $\delta$  was fixed as it traded with  $\beta$  during model fitting, a pattern also confirmed during exploratory analysis.

## Results

### Agreement conditions

Figure 2 shows the average data (points) and best-fit functions (smooth curves). Table 2 lists the model parameters. The best-fitting parameter allocation was a  $6\lambda-2\beta-1\delta$  model. The fast rate,  $1.56 \text{ s}^{-1}$  ( $1/0.641 \text{ s}$ ), was assigned to all plural conditions and to the adjacent singular conditions. The slow rate,  $1.36 \text{ s}^{-1}$  ( $1/0.735 \text{ s}$ ), was assigned to non-adjacent singular conditions. This rate difference corresponds to a 94 ms slowing for non-adjacent singular DPs compared to all other conditions.

<< Insert Figure 2 here >>

<< Insert Table 2 here >>

Model parameterizations and competition results are given in Table 3. The best-fitting model corresponds to model number 6. Each IC measure is inherently biased to more (AIC) or fewer (BIC) parameters, but we observed that they both converged on the same parameter allocation. Also, four of the five best models, based on AIC, made an adjacent/non-adjacent distinction (Model 6, Models 8-10). There was little to no support, however, for assigning a separate rate parameter to each length. The single next-best fitting model, Model 3, was a non-focal model which assigned separate rates according to number alone. We address its feasibility below in the participants' analysis.

<< Insert Table 3 here >>

We performed the same model competition analysis by participants. For 15 of 18 participants, the fully saturated  $6\lambda-6\beta-1\delta$  model achieved the lowest AIC/BIC scores (average *AIC*: -104.6,  $w_{\text{AIC}}$ : 0.56; *BIC*: -70.4,  $w_{\text{BIC}}$ : 0.31). Analysis of the parameter values revealed that the pattern of rates was nonetheless qualitatively similar to the average data. Average participant parameters are given in Table 4. Non-adjacent conditions were slower for both singulars and plurals, but the

average by-participant slowing was significantly greater in non-adjacent singular DPs than in plural DPs (Wilcoxon Signed-Ranks test; *Common scaling*:  $W = 476, p < .05$ , *Discriminative*:  $W = 446, p < 0.10$ ). Overall, agreement in plural DPs was processed more quickly than in singular DPs (*Common*:  $W = 137, p < 0.05$ , *Discriminative*:  $W = 129, p < 0.10$ ). These patterns are consistent with the fact that the second-best model for 15 of 18 participants was a focal model (Model 10), which assigned 3 rate parameters: one each to the singular/adjacent condition, the singular/non-adjacent conditions, and the plural conditions. The last finding recalls the fact that, in the average analysis, the 6-2-1 number model was the strongest competitor for the focal model. Average parameter fits thus confirm that a significant majority of participants (15 of 18) showed a large rate disadvantage for non-adjacent singulars coupled with an overall rate advantage for plurals.

<< Insert Table 4 here >>

### **Plausibility control conditions**

We used the plausibility controls to assess whether costs associated with processing longer DPs could be attributed solely to the addition of modifiers, even when there are no number features to keep track of. Four speed-accuracy time series were computed from the plausibility controls (Table 1, (m)-(r)), for single and double adjective conditions and both the singular and plural head nouns. Hit rates for the plausible conditions were scaled against the false alarm rates for implausible conditions, and a competitive model analysis was performed as above. It was not meaningful to scale the adjacent conditions against one another because, without a modifier, both adjacent conditions were plausible.

The results of this analysis are relatively straightforward. For the average data, the best-fitting model of the four conditions was a  $1\lambda-2\beta-1\delta$  model which assigned two parameters: a

faster one to the one-modifier conditions, and a slower one to the two-modifier conditions ( $\beta_1^{-1}$ : 0.741 s,  $\beta_2^{-1}$ : 0.862 s; AIC: -154.4, BIC: -145.6). The by-participants analysis over the fully saturated  $4\lambda-4\beta-1\delta$  model yielded similar conclusions and its results are given in Table 5.

<< Insert Table 5 here >>

The individual rate parameters for the plausibility controls show consistent slowing for one v. two modifier conditions ( $W = 199$ ,  $p < .05$ ) and, crucially, the size of the slow-down is not affected by number ( $W = 58$ ,  $p < .25$ ). Thus, the analysis of individual parameters points to the same conclusion as the average analysis for the plausibility controls: two rates are needed to capture the data. The analysis of individual parameters differs from the average analysis, however, in the estimation of the asymptote parameters. In the individual data, there was a consistent effect of modifier number on accuracy ( $\mu$ :  $-0.27 \pm 0.09$  d';  $W = 93$ ,  $p < .001$ ).

## **Discussion**

### *Summary*

The goal of this experiment was to determine whether information about a phrase's grammatical number is maintained in the focus of attention as that phrase is being processed. The MR-SAT technique was used to measure accuracy at discriminating grammatical sentences from ungrammatical sentences at 17 successive response lags. Results indicate that agreement between non-adjacent heads is processed more slowly than agreement between adjacent heads, a slowing which is exacerbated by singular number values.

We propose the best interpretation of our results is in terms of focal attention. The key assumption is that not all information about a grammatical object can be concurrently maintained while that object is being constructed or analyzed. Once number information is shunted out of focal attention, it must be retrieved in order to complete the task. It is this retrieval that accounts for the rate difference between adjacent and non-adjacent conditions. In the cases we examined, the number encoded on the determiner is only one of several properties about the phrase which the comprehender must decode. For example, the word 'this' signals not only singularity, but also definiteness and that deictic reference is intended. Participants asymptotically showed high sensitivity to the agreement contrasts, regardless of distance ( $\mu: 2.85 \pm 0.06 d'$ ), suggesting that number was generally decoded in the task. However, the intervention of other processing tasks, such as the analysis of modifiers, may have made it impossible to simultaneously maintain all DP properties. Competition between different features for focal attention is likely to be particularly acute if those features are not directly relevant to the modifiers' syntactic and semantic analysis.

*Complexity, composition & reanalysis.* An important question is whether the observed rate difference truly reflects the additional retrieval operation required in the non-adjacent cases. Rate differences are known to arise for other reasons, as, for example, in reanalysis (Bornkessel, Schlesewsky & McElree, 2004, Martin & McElree, 2018). An obvious source of processing difficulty independent of memory retrieval is the analysis of the modifier sequence. Three pieces of evidence mitigate the concern that rate differences observed in our data derive from properties of the modifiers.

Firstly, the observed contrast in processing speed correlates with adjacency and not to how many modifiers there are. There is a binary distinction between adjacent and non-adjacent cases, evidenced by the fact that 2- $\beta$  models outperformed the corresponding 3- $\beta$  models. If it was additional modification that slowed agreement processing, then difficulty would be expected to correlate with the number of modifiers, each with an independent syntactic and semantic contribution. The second piece of evidence comes from the plausibility controls (e.g., *#the clever, risk-taking volcano*). For discriminations based on plausibility, we observed a large rate difference between single and double modification. Rate differences therefore were sensitive to the number of modifiers only for plausibility-based discriminations and not for number-based discriminations<sup>2</sup>. The third piece of evidence comes from the effect of number: the rate

---

<sup>2</sup> It is conceivable that the one- versus two-modifier rate difference in plausibility controls (e.g., “the risk-taking burglars/\*jewels”; “the clever, risk-taking burglars/\*jewels”) could also be attributed to retrieving the first modifier from memory. This is possible and we cannot definitively rule it out. But there are two considerations that push us away from that interpretation. The first is the fact that, in the plausibility controls, the two modifiers were always mutually consistent such that the last modifier was always sufficient to determine (un)acceptability when combined with the head noun. We didn’t have any mixed cases like, “the shiny, risk-taking jewels”, where one modifier was compatible with the head noun, and the other wasn’t. The second is a commitment to early and incremental semantic composition (Brennan & Pytkäinen, 2012), which predicts a partial meaning to be elaborated by the time the head noun is reached. For this reason, we attribute the one versus two modifier cost to the complexity of the meaning. But further work is required here to arrive at a better conclusion.

difference in agreement discrimination was most robust for singular DPs. In the plausibility control conditions, though, we observed that the cost of extra modification was independent of number. There was thus a number asymmetry in agreement contrasts, but no such asymmetry in plausibility contrasts. For these reasons, it seems unlikely that the cost of composing the modifiers with the noun is responsible for the pattern of rates we observe in the agreement contrasts.

A related, but distinct, possibility is that participants initially misanalyzed the modifier sequence, forcing a reanalysis at the noun. It has been observed that singular nouns are incrementally parsed as the head of the determiner's sister NP, even though they could ultimately form the left constituent in a compound (Staub et al. 2007). For example, a sentence like, "I met the elevator mechanic," would have a fleeting parse in which the grammatical object is taken to be "the elevator". The question arises whether a similar fleeting parse was present for our comprehenders, since participles like "risk-taking" can often be gerunds (e.g., "How much risk-taking will you tolerate?"). It is possible that comprehenders adopted a gerund analysis for singular DPs, which would require subsequent reanalysis and could potentially account for the rate disparity between non-adjacent singulars and plurals. Three pieces of evidence speak against this possibility. Firstly, the 2-intervener conditions also showed a slower rate than adjacent conditions. Yet these conditions foreclosed the gerund analysis via the orthographic signal of comma placement ("this clever, risk-taking ..."; recall that the comma was presented in the same chunk as the first modifier, i.e., "clever,"). Secondly, we performed a common scaling analysis which pooled the endorsement rates of singular and plural grammatical conditions into a single hit rate (per length condition and lag-latency) against which the individual false alarms were scaled. If there were an obligatory reanalysis in the (ultimately grammatical) singular phrases,

then the plural conditions would be expected to have slightly higher rate parameters in the common analysis than in the discriminative analysis and the singular conditions slightly lower ones. However, we found the rate parameters differed from the discriminative analysis by less than  $10 \text{ ms}^{-1}$  and both were higher. Finally, in plausibility discriminations, there was no rate difference based on phrase number. If there were an obligatory reanalysis, then we would have expected plural conditions to show slower rates since comprehenders would be forced to revise not only constituent structure but also change the number features of the phrase.

Based on these considerations, it seems unlikely that a reanalysis from gerund to participle, present only in singular conditions, could explain the observed rate differences by number. We conclude that the dichotomous rate difference observed for agreement trials derives not from modification per se, but its derivative effect of separating determiner from noun.

### ***The status of plurals in the focus of attention***

Why should the plural number value be more likely to survive the entire breadth of the phrase than singular? The answer may lie in the observation that plural is very often the marked category for number in two-number systems (Greenberg, 1963, Eberhard, 1997). For example, plurals as a type have a lower frequency of occurrence than singulars. The exponents of plural agreement often do not distinguish the person and gender features which are differentiated in the singular, as is true in English. Finally, they are known to be associated with small but added RT costs in reading (Wagers, Lau & Phillips, 2009).

How the distributional concept of markedness maps onto a representation is complex (Adger & Harbour, 2008). It may simply be another property of marked categories that they are more easily maintained in focal attention. However, a potential explanation lies in the treatment



of unmarked categories as a representational default which is not explicitly encoded. Such a representational scheme is referred to as privative (Trubetzkoy, 1939). For example, in their analysis of pronominal systems, Harley and Ritter (2002) associate plurals with more explicit structure than singulars. Singulars acquire their status as singulars by application of a default rule, whereas plurals (and duals) have distinguished nodes in their feature geometries<sup>3</sup>. This theory of markedness has been called upon to explain the fact that only plural nouns lead to strong agreement attraction effects by positing that only marked feature values are appropriately ‘visible’ (Eberhard, 1997; cf. Badecker & Kuminiak, 2007, Slioussar & Malko, 2016).

It may first appear that the unmarked value for number should be more durable in the focus of attention, since there is literally less to maintain. But if the privative encoding scheme is coupled with the possibility of forgetting, then what the absence of a feature signals is more limited than what its presence does. Suppose that information is shunted from the focus of attention not in wholesale constituent chunks, but on a feature-by-feature basis. The mere presence of [PL] in focal attention provides reliable evidence about previously-encountered plurality, even if other features of the constituent have been lost (and if we assume that features cannot spontaneously populate our representations without evidence). However the absence of [PL] gives rise to two possibilities: either the determiner may have been singular or its [PL] feature was stochastically shunted from focal attention. Crucially, it would be necessary to

---

<sup>3</sup> Whether plurals are actually semantically more complex than singulars is a matter of debate (Sauerland, Anderssen & Yatsushiro, 2005; Farkas, 2006, de Swart & Farkas, 2010). The resolution of this debate is independent of the morphosyntactic facts discussed. If plurals do turn out to be semantically ‘basic’, then they would represent the interesting, less frequent case in which formal morphosyntactic markedness does not align with semantic markedness.

retrieve the determiner's feature set to determine that the absence of [PL] corresponds to singularity.

### *Anti-locality and length-based complexity*

Increasing distance between syntactic dependents in clause-bounded verb-argument relationships has been shown to reduce processing load on the second dependent (Konieczny, 2000, Vasishth & Lewis, 2006, Nakatani & Gibson, 2008, Levy & Keller, 2013), a phenomenon dubbed 'anti-locality.' Although the relationship between a determiner and noun is similar to the verb-argument relationship in the boundedness of its domain, we found no facilitation in any SAT measure for longer dependencies. They were processed neither faster nor more accurately.

One understanding of anti-locality effects comes from probabilistic resource allocation models, like Levy (2008). The comprehender is assumed to have a distribution of confidence or likelihood across the possible analyses of the input. The more highly a category's identity and position is predicted by the context, then the less the comprehender's confidence distribution will have to be reallocated when that category is observed in the input. Consequently, it will be easier to process. Anti-locality obtains when the information that intervenes between two dependents diminishes uncertainty about the identity and location of the second dependent. Thus, because increased material can sharpen and reinforce expectations, it is possible to obtain a length-based facilitation.

To assess the fit of this account to our data, we estimated the length distribution of DPs headed by the four English demonstrative determiners<sup>4</sup>. Under an expectation-based account, we

---

<sup>4</sup> This is only an approximation, since the entire conditional probability of an expression is relevant to the speed of processing on a particular word. For this analysis we assume that the external distribution of the demonstrative DP does not affect how its length varies with number. That may turn out to be false (for example, if plural DPs are shorter in subject position but not in object position).

would expect a noun to be less difficult to process the higher its conditional probability (Levy, 2008). Recall that we found that, for non-adjacent conditions, those that began with a plural demonstrative were always processed faster. This would be predicted under the expectation-based account if the conditional probability of encountering the noun, given a plural demonstrative and one or two modifiers, was greater than encountering the noun, given a singular demonstrative and any modifiers. Plural DPs should therefore, on average, be shorter in word length to explain the consistent rate differences we saw. What we found, in brief, was that singular DPs were on average shorter, and that singular demonstratives led to sharper and earlier expectations for nouns after both single and double modification. For example, the conditional probability of a singular noun, following a singular determiner and two modifiers, was 0.74; compared to 0.71 for plural nouns, following a plural determiner and two modifiers. Yet the singular noun was processed about 204 ms slower in that condition.

Parsed DPs were extracted from the New York Times subsection of the Gigaword corpus (Parker, et al. 2009), using *TGrep2* (Rohde, 2005). Approximately 1.1 million such phrases were found. Singular proximal DPs were the most abundant (*this/these*: 570,699/130,824, *that/those*: 274,658/120,752). For 50 sample sets of the four determiners in our experiment (sample N = 1000 for each determiner), distance from determiner to the head of its sister noun phrase was computed. Table 6 presents the average proportion of DPs at each word length. There was a reliable difference between the singular and plural length distributions for all 50 samples (log-likelihood goodness of fit test; minimum  $G(5) = 36.1$ ,  $p < .001$ ). In contrast, length distributions were closely matched across determiner type (mean  $G(5) = 8.3$ ,  $p < 0.15$ , 95% bounds [1.57, 16.83], reliable in only 14 of 50 samples).

<< Insert Table 6 here >>

Thus, we found difference in the size distribution of DPs based on their grammatical number. In particular, plural DPs tend to be slightly longer (=more words). To compare predictions with each of our 3 experimental length conditions, the proportions were renormalised to represent conditional probabilities at each successive length  $n$  against the distribution of DPs greater or equal to  $n$ . The left panel of Table 7 shows the likelihood of immediately encountering the head noun given the occurrence of the determiner and between zero and two intervening words. The right panel of Table 7 shows how well these predictions match the observed asymptotic accuracies and speed parameters. When compared specifically with the speed measures in our experiment, the conditional probabilities of a noun, given its local DP context, do not align well with an expectation-driven account of processing difficulty. However when compared with the accuracy measures, the conditional probabilities are well aligned. The size distribution of DPs suggests that, comparing across either number values or across lengths, there should be both an advantage for singular phrases and an advantage for shorter phrases.

<< Insert Table 7 here >>

Prior studies of anti-locality have focused on reading times as a dependent measure. However, reading time differences do not necessarily reflect differences in the underlying processing dynamics. The present analysis underscores this fact, by showing how an expectation-based account fares better with the accuracy parameters that model our data than with the speed parameters. This pattern generates clear hypotheses that future work could address namely that, (a), the allocation of confidence to alternative analyses impacts the likelihood that a sequence of parsing operations will ultimately deliver the correct result; but that, (b), the allocation of

confidence to alternative analysis is independent of the speed with which those analyses are delivered.

### *Relationship to previous focus of attention findings and implications*

The non-adjacency-triggered rate difference in agreement discrimination, 94 ms, provides an estimate of the retrieval time for the displaced properties of the determiner. It is worth noting that this estimate is similar in magnitude to the 85 ms figure obtained in McElree, Foraker and Dyer (2003), who studied the subject features displaced by an intervening relative clause; and to estimates provided by Wagers & McElree (2009; 74 ms and 87 ms in two experiments).<sup>5</sup> This observation further implicates focal attention, though admittedly it is a more suggestive piece of evidence than the others presented.

McElree, Foraker and Dyer (2003) found that a clause is sufficient to displace features of the subject from focal attention. That finding is not surprising in light of the present results that an intervening (complex) modifier is enough to displace DP number. Yet, the fact that information can be shunted across such a small domain raises a disconcerting question: how could it be possible to interpret expressions if local information competes so aggressively for focal attention? The apparent smallness or simplicity of DP may in fact be deceptive (Abney, 1987, Leu, 2008). But the more general answer to the question will lie in expanding the present work to other domains and other information sources. Wagers and McElree (2009) have reported that subject information survives across PP attachments that plausibly represent a greater

---

<sup>5</sup> That the similarity of these estimated values is not in some way an artifact of the technique can be appreciated by surveying results from other constructions that cannot be plausibly characterised as a simple difference between whether a feature is available in focal attention or must be retrieved from memory. For example, the difference between single- and double-gap dependency resolution examined in McElree et al. (2003) leads to rate differences on the order of 400 ms. Here, the SAT procedure is tracking differences in the time to retrieve one versus two representations, and, crucially, the respective position (or order) in syntax.

distance than what we have tested here. However, discrimination in that experiment probably relied largely on animacy information. The ability to maintain animacy and number could be associated with different hazard functions. Indeed Ness & Meltzer-Asscher (2019) found that animacy was maintained over relatively long distances as well. The survival rates of different features, we conjecture, will depend on such factors as their diagnostic importance in the construction of future relationships, as well as their respective susceptibility to displacement and interference from on-going processing. Intertwined with these factors is the directionality of the dependency. In the present study, we focused on the maintenance of a feature that was introduced (arguably) by a dependent element: the target of agreement, in this case a demonstrative, whose number value depends on that of its agreement controller, the noun. This contrasts with most other studies of agreement processing, where it is an agreement controller, i.e., the subject, which introduces the feature of interest.

If the scope of what can be maintained in the focus of attention is relativised to the identity of the information, then we arrive at more complicated conception of how complex representations are chunked. There is a conventional, if implicit, view that complex representations are ‘carved-at-the-joints’, having been exhaustively parsed into non-overlapping packets of information according to their structural domain. However, the present data and its comparison to previous findings suggest that the functional capacity of focal attention may not fully correspond to a uniform set of structural domains, but may also be sensitive to the relationships individual features participate in.

**Acknowledgments**

The authors acknowledge the support of NIH grant #HD-056200. Thanks to Kathy Akey for assistance in collecting data. The authors are grateful to the audience of the 22<sup>nd</sup> Annual CUNY Sentence Processing Conference (University of California, Davis) for their helpful comments; to two anonymous reviewers for their insight and critical commentary; to Pranav Anand for his help parsing the Gigaword corpus; and to the following individuals for important contributions to the development of the project: Julie Van Dyke, Ellen Lau, Colin Phillips, Michael Shvartsman & Clare Stroud.

## References

- Abney, S.P. (1987). *The English noun phrase in its sentential aspect*. Dissertation. Cambridge, MA: MIT.
- Adger, D., & Harbour, D. (2008). Why Phi? In D. Harbour, D. Adger, & S. Béjar, eds., *Phi Theory*. New York: Oxford UP.
- Anderson, J.R. & C. LeBiere (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.
- Anderson, M.C., & Neely, J.H. (1996). Interference and inhibition in memory retrieval. In E. Ligon & R. Bjork (Eds.). *Memory, Handbook of perception and cognition* (2<sup>nd</sup> ed., pp. 237-313). San Diego, CA: Academic Press.
- 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 sentence processing conference. San Diego, La Jolla, CA: University of California.
- 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, 65–85.
- Barber, H., & M. Carreiras (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *Journal of Cognitive Neuroscience*, 17, 137-153.
- Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23, 45-93.



- Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from relational structure. *Journal of Memory and Language*, 51, 495-522
- Brennan, J., & Pylkkänen, L. (2012). The time-course and spatial distribution of brain activity associated with sentence processing. *Neuroimage*, 60(2), 1139-1148.
- Broadbent, D. E. (1958). *Perception and communication*. New York: Oxford UP.
- Carroll, J.M. & M. Tanenhaus. (1978). Functional clauses and sentence segmentation. *J Speech Hearing Res*, 21, 793-808.
- Chomsky, N. (1981). *Lectures on government and binding*. The Hague: Mouton de Gruyter.
- Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. New York: Oxford UP.
- Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review*, 119(3), 480–499. <https://doi.org/10.1037/a0027791>.
- Davidson, D. J., Hanulíková, A., & Indefrey, P. (2012). Electrophysiological correlates of morphosyntactic integration in German phrasal context. *Language and Cognitive Processes*, 27(2), 288-311.
- Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, 69(2), 85-103.
- Eberhard, K. (1997). The marked effect of number on subject-verb agreement. *Journal of Memory and Language*, 36, 147-164.

- Elbourne, P. (1999). Some correlations between semantic plurality and quantifier scope. In *North East Linguistics Society* (Vol. 29, No. 1, p. 7).
- Farkas, D.F. (2006). The unmarked determiner. In S. Vogeleer & L. Tasmowski de Rijk (Eds.), *Non-definiteness and plurality* (pp. 81-106). Amsterdam: John Benjamins.
- de Swart, H., & Farkas, D. (2010). The semantics and pragmatics of plurals. *Semantics and Pragmatics*, 3, 6-1.
- Fodor, J.A. & Bever, T.G. (1965). The psychological reality of linguistic segments. *J Verbal Learning Verbal Behavior*, 4, 414-420.
- Fodor, J.A., Bever, T.G., & Garrett, M.F. (1974). *The psychology of language: an introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.
- Franck J., & Wagers, M. (2020). Hierarchical structure and memory mechanisms in agreement attraction. *pLoS ONE* 15(5), e0232163. <https://doi.org/10.1371/journal.pone.0232163>.
- Frazier, L. & Fodor, J.D. (1978). The sausage machine: a new two-stage parsing model. *Cognition*, 6, 291-325.
- Garavan, H. (1998). Serial attention within working memory. *Memory and Cognition*, 26, 263-276.
- Glover, S. & Dixon, P. (2004). Likelihood ratios: a simple and flexible statistic for empirical psychologists. *Psychonomic Bull Rev*, 11, 791-806.
- Gordon, P.C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 1411-1423.

- Gordon, P.C., Hendrick, R., & Levine, W.H. (2002). Memory-load interference in syntactic processing. *Psychological Science, 13*, 425-430.
- Gordon, P.C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*, 97-114.
- Gordon, P.C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32*, 1304-1321.
- Greenberg, J. (1966). *Language universals: with special reference to feature hierarchies*. The Hague: Mouton.
- Gunter, T.C., Friederici, A.D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12*, 556-568.
- Hagoort, P. & Brown, C.M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research, 28*, 715-728.
- Harley, H. & Ritter, E. (2002). A feature-geometric analysis of person and number. *Language, 78*, 482-526.
- Jackendoff, R. (1977). *X-bar syntax*. Cambridge, MA: MIT Press.
- Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language, 94*, 316-339.

- Jarvella, R.J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, *10*, 409-416.
- Johns, C. L., Matsuki, K., & Van Dyke, J. A. (2015). Poor reader'' retrieval mechanism: efficient access is not dependent on reading skill. *Frontiers in psychology*, *6*, 1552.
- Jonides, J., Lewis, R.L., Nee, D.E., Lustig, C.A., Berman, M.G., & K.S. Moore (2008). The mind and brain of short-term memory. *Annu. Rev. Psychol.*, *59*, 15.1-15.32.
- King, R. (2021). *Feature Agreement in Real-Time Language Comprehension: Interaction of Cognitive and Linguistic Constraints*. Dissertation. New York: New York University.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*(6), 627- 645.
- Kush, D., Johns, C. L., & Van Dyke, J. A. (2019). Prominence-sensitive pronoun resolution: New evidence from the speed-accuracy tradeoff procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(7), 1234.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126-1177.
- Levy, R. P., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, *68*(2), 199-222.
- Leu, T. (2008). *The internal syntax of determiners*. Dissertation. New York: New York University.
- Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 375-419.

- Liu, Charles C., & Smith, P.L. (2009). Comparing time-accuracy curves: beyond goodness-of-fit measures. *Psychonomic Bulletin & Review*, *16*, 190-203.
- MacMillan, N.A. & Creelman, C.D. (1991). *Detection theory: a user's guide*. New York: Cambridge UP.
- Marslen-Wilson, W.D. (1975). Sentence perception as an interactive parallel process. *Science*, *189*, 226-228.
- Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, *58*(3), 879-906.
- Martin A.E., & McElree, B. (2018). Retrieval cues and syntactic ambiguity resolution: speed-accuracy tradeoff evidence. *Language, Cognition, and Neuroscience*, *33*(6), 769-738. doi: 10.1080/23273798.2018.1427877.
- McElree, B. (1998). Attended and nonattended states in working memory: accessing categorized structures. *J. Mem. Lang.*, *38*, 225–52.
- McElree, B. (2001). Working memory and focal attention. *J. Exp. Psychol.: Learn. Mem. Cogn.* *27*, 817–35
- McElree B. (2006). Accessing recent events. *Psychol. Learn. Motiv.*, *46*, 155–200.
- McElree, B., & Doshier, B. A. (1989). Serial position and set size in short-term memory: Time course of recognition. *Journal of Experimental Psychology: General*, *118*, 346-373.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory & Language*, *48*(1), 67-91.

- Miller, G.A., & Chomsky, N. (1963). Finitary models of language users. In Luce, R.D., Bush, R.R., & Galanter, E. (eds.), *Handbook of Mathematical Psychology*, Volume II. New York: John Wiley.
- Nakatani, K., & Gibson, E. (2008). Distinguishing theories of syntactic expectation cost in sentence comprehension: Evidence from Japanese. *Linguistics*, *46*, 6386.
- Nee, D.E. & Jonides, J. (2008). Neural correlates of access to short-term memory. *Proceedings of the National Academy of Sciences USA*, *105*, 14228-14233.
- Ness, T., & Meltzer-Asscher, A. (2019). When is the verb a potential gap site? The influence of filler maintenance on the active search for a gap. *Language, Cognition and Neuroscience*, *34*(7), 936-948.
- Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *28*, 411-421.
- Oberauer, K., & Bialkova, S. (2009). Accessing information in working memory: can the focus of attention grasp two elements at the same time? *Journal of Experimental Psychology: General*, *138*, 64-87.
- Ouwayda, S. (2013). Where Plurality Is: Agreement and DP Structure. In Stefan Keine & Shayne Sloggett (eds.), *Proceedings of the 42nd Annual Meeting of the North East Linguistic Society*, Amherst, MA: GLSA Publications.
- Öztekin, I., McElree, B., Staresina, B.P., & Davachi, L. (2008). Working memory retrieval: Contributions of left prefrontal cortex, left posterior parietal cortex and hippocampus. *Journal of Cognitive Neuroscience*, *21*, 581-593.
- Oztekin, I., Davachi, L., & McElree, B. (2010). Are representations in working memory distinct from representations in long-term memory? *Psychological Science*, *21*, 1123–1133.

- Norris, M. (2014). *A theory of nominal concord*. Dissertation. Santa Cruz, CA: UC Santa Cruz.
- Norris, M. (2017). Description and analyses of nominal concord (Pt I). *Language and Linguistics Compass*, 11, e12266.
- Parker, R., Graff, D., Kong, J., Chen, K. & Maeda, K. (2009). *English Gigaword Fourth Edition*. Philadelphia: Linguistic Data Consortium.
- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41, 427-456.
- Pesetsky, David. (2013). *Russian Case Morphology and the Syntactic Categories*. Cambridge, MA: MIT Press.
- Potter, M.C. (1988). Rapid serial visual presentation (RSVP): A method for studying language processing. In D. E. Kieras and M.A. Just (Eds.), *New methods in reading comprehension research*. Hillsdale, NJ: Erlbaum Press.
- Potter, M.C., & Lombardi, L. (1992). The regeneration of syntax in short-term memory. *Journal of Memory and Language*, 31, 713-33.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.
- Rohde, D. (2005). *tGrep2 v. 1.15*. <http://tedlab.mit.edu/~dr/Tgrep2/>
- Sachs, J.S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2, 437-442.
- Sauerland, U., Anderssen, J., & Yatsushiro, J. (2005). The plural is semantically unmarked. In S. Kepser & M. Reis (Eds.), *Linguistic Evidence*, Berlin: Mouton de Gruyter.

- Smith, P. W. (2017). The syntax of semantic agreement in English 1. *Journal of Linguistics*, 53(4), 823-863.
- Slioussar, N., & Malko, A. (2016). Gender agreement attraction in Russian: production and comprehension evidence. *Frontiers in Psychology*, 7, 1651.
- Townsend, D.J. & Bever, T. (2001). *Sentence comprehension: the integration of habits and rules*. Cambridge, MA: MIT Press.
- Trubetzkoy, N.S. (1939). *Grundzüge der Phonologie*. Christiane Baltaxe, trans. *Principles of Phonology*, Berkeley: University of California Press, 1969.
- Unsworth, N. & Engle, R.W. (2009). Speed and accuracy of accessing information in working memory: an individual differences investigation of focus switching. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34, 616-630.
- Vainio, S., Hyönä, J., & Pajunen, A. (2008). Processing modifier-head agreement in reading: Evidence for a delayed effect of agreement. *Memory & Cognition*, 36, 329-340.
- Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A retrieval interference theory of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49(3), 285-316.
- Van Dyke, J.A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55, 157-166.
- Vasishth, S. & Lewis, R.L. (2006). Argument-head distance and processing complexity: explaining both locality and antilocality effects. *Language*, 82, 767-794.



- Vasishth, S., Bruessow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, 32, 685-712.
- Vergauwe, E., Hardman, K. O., Rouder, J. N., Roemer, E., McAllaster, S., & Cowan, N. (2016). Searching for serial refreshing in working memory: Using response times to track the content of the focus of attention over time. *Psychonomic Bulletin & Review*, 23(6), 1818-1824.
- Verhaegen, P. & Hoyer, W.J. (2007). Aging, focus switching and task switching in a continuous calculation task: Evidence toward a new working memory control process. *Aging, Neuropsychology, and Cognition*, 14, 22-39.
- Verhaegen, P., Cerella, J., & Basak, C. (2004). A working memory workout: how to expand the focus of serial attention from one to four items in 10 hours or less. *J. Exp. Psychol.: Learn. Mem. Cogn.* 30, 1322-37.
- Villata, S., & Franck, J. (2020). Similarity-based interference in agreement comprehension and production: Evidence from object agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(1), 170.
- Wagers, M.W., Lau, E. & Phillips, C. (2009). Agreement attraction in comprehension: representations and processes. *Journal of Memory and Language*, 61, 206-237.
- Wagers, M. & McElree, B. (2009). Focal attention and the timing of memory retrieval in language comprehension. Paper presented at *Architectures and Mechanisms for Language Processing Conference*, 15, Sept. 7-9, Barcelona.
- Wickelgren, W.A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41, 67-85.
- Wickelgren, W. A., Corbett, A. T., & Doshier, B. A. (1980). Priming and retrieval from 351

short-term memory: A speed accuracy trade-off analysis. *Journal of Verbal Learning and Verbal Behavior*, 19, 387-404.

Zhang, Y., & Verhaeghen, P. (2009). Glimpses of a one-speed mind: Focus-switching and search for verbal and visual, and easy and difficult items in working memory. *Acta Psychologica*, 131, 235-244.

**Tables**

Table 1. Sample item set

<u>Preamble:</u> <i>The detective was mistaken about the location of _____ .</i>	
<u>Continuation:</u>	
<b>SG Determiner</b>	<b><i>Acceptable/Unacceptable</i></b>
<i>0 interveners</i>	that burglar/*burglars
<i>1</i>	that risk-taking burglar/*burglars
<i>2</i>	that clever, risk-taking burglar/*burglars
<b><i>PL Determiner</i></b>	
<i>0 interveners</i>	those burglars/*burglar
<i>1</i>	those risk-taking burglars/*burglar
<i>2</i>	those clever, risk-taking burglars/*burglar
<b><i>Adj-Noun control</i></b>	
<i>0 interveners</i>	the burglars/jewels
<i>1</i>	the risk-taking burglars/#jewels
<i>2</i>	the clever, risk-taking burglars/#jewels

Table 2. Best-fitting model parameters for average data

	<b>SINGULAR DP</b>			<b>PLURAL DP</b>		
	<i>Adjacent</i>	<i>+1</i>	<i>+2</i>	<i>Adjacent</i>	<i>+1</i>	<i>+2</i>
DISCRIMINATIVE SCALING						
$\lambda$ (d')	3.11	2.87	2.81	3.16	2.75	2.86
$\beta^{-1}$ (s)	0.641	0.735		0.641		
$\delta$ (s)	0.573					
COMMON SCALING						
$\lambda$ (d')	3.11	2.87	2.76	3.17	2.78	2.92
$\beta^{\prime}$ (s)	0.643	0.743		0.643		
$\delta$ (s)	0.573					

*Note.* The best-fitting  $6\lambda-2\beta-1\delta$  model for both discriminative and common scaling  $d'$ . These values correspond to Model 6 in Table 3.

Table 3. Competitive model analysis

MODEL PARAMETERIZATIONS							MODEL FIT							
<u>Number</u>	Singular			Plural										
<u>Distance</u>	0	1	2	0	1	2	DISCRIMINATIVE SCALING				COMMON SCALING			
							AIC*	wAIC	BIC*	wBIC	AIC*	wAIC	BIC*	wBIC
1. 6-1-1	$\beta_1$						-216.8	.01	-195.8	.05	-216.5	.01	-195.5	.05
2. 6-2-1	$\beta_1$			$\beta_2$			-223.3	.20	-199.7	.34	-223.0	.20	-199.4	.34
3. 6-3-1	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_1$	$\beta_2$	$\beta_3$	-217.2	.01	-191.0	~0	-216.9	.01	-190.1	~0
4. 6-6-1	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	-219.6	.03	-185.5	~0	-219.3	.03	-185.1	~0
5. 6-2-1	$\beta_1$	$\beta_2$		$\beta_1$	$\beta_2$		-218.7	.02	-195.0	.03	-218.4	.02	-194.8	.03
6. 6-2-1	$\beta_1$	$\beta_2$		$\beta_1$			-223.7	.24	-200.1	.41	-223.4	.24	-199.8	.41
7. 6-2-1	$\beta_1$			$\beta_1$	$\beta_2$		-215.5	~0	-191.9	.01	-222.9	.19	-194.0	.02
8. 6-4-1	$\beta_1$	$\beta_2$		$\beta_3$	$\beta_4$		-223.2	.19	-194.3	.02	-215.3	~0	-191.6	.01
9. 6-3-1	$\beta_1$	$\beta_2$		$\beta_1$	$\beta_3$		-223.3	.19	-197.0	.09	-222.9	.18	-196.7	.09
10. 6-3-1	$\beta_1$	$\beta_2$		$\beta_3$			-222.0	.10	-195.7	.05	-221.6	.10	-195.4	.05

*Note.* The ten candidate  $6\lambda$ - $1\delta$  models are given with both AIC and BIC goodness-of-fit measures and corresponding model weights ( $w$ ). The left panel schematises how each model's  $\beta$  was distributed according to condition. Models 1 – 4 correspond to the experiment's factorial design. Models 5-10 are all 'focal' models in the sense that they distinguish the two Determiner-Noun adjacent conditions from the other four. Results of model comparison from both discriminative scaling and common scaling are included. For models  $i$  and  $j$ ,  $w_i/w_j$  gives the likelihood ratio favoring model  $i$  over model  $j$ . The best-fitting model is outlined.

Table 4. Average participant parameters for agreement contrasts

	<b>SINGULAR DP</b>			<b>PLURAL DP</b>		
	<i>Adjacent</i>	<i>+1</i>	<i>+2</i>	<i>Adjacent</i>	<i>+1</i>	<i>+2</i>
$\lambda$ (d')	3.16 (0.06)	3.05 (0.13)	2.92 (0.11)	3.16 (0.04)	2.83 (0.10)	2.90 (0.09)
$\beta^{-1}$ (s)	0.580 (0.078)	0.880 (0.202)	0.789 (0.175)	0.518 (0.062)	0.651 (0.096)	0.585 (0.096)
$\delta$ (s)	0.638 (0.044)					

*Note.* Mean parameter values for the  $6\lambda-6\beta-1\delta$  model with standard errors across participants.

Table 5. Average participant parameters for plausibility contrasts

		<b>SINGULAR DP</b>		<b>PLURAL DP</b>	
		<i>+1 Adj</i>	<i>+2 Adj</i>	<i>+1 Adj</i>	<i>+2 Adj</i>
$\lambda$	(d')	3.17 (0.05)	2.91 (0.12)	3.10 (0.06)	2.81 (0.10)
$\beta^{-1}$	(s)	0.574 (0.080)	0.733 (0.179)	0.529 (0.057)	0.772 (0.136)
$\delta$	(s)	0.655 (0.042)			

*Note.* Mean parameter values for the  $4\lambda-4\beta-1\delta$  model with standard errors across participants.

Table 6. Determiner-to-noun distance distribution

		<b>Distance from determiner to head noun</b>					
		<i>In words</i>					
		0	1	2	3	4	$\geq 5$
<b>Sg</b>	'this'	0.883	0.088	0.022	0.005	0.001	0.001
	'that'	0.856	0.112	0.023	0.006	0.002	0.001
	<i>mean</i>	<b>0.870</b>	<b>0.100</b>	<b>0.023</b>	<b>0.005</b>	<b>0.002</b>	<b>0.001</b>
<b>Pl</b>	'these'	0.777	0.165	0.042	0.010	0.002	0.004
	'those'	0.783	0.162	0.040	0.010	0.003	0.001
	<i>mean</i>	<b>0.780</b>	<b>0.164</b>	<b>0.041</b>	<b>0.010</b>	<b>0.003</b>	<b>0.003</b>

*Note.* Proportions represent the average proportions of the number of DPs in each number (Sg, Pl) and type ('this'/'these'; 'that'/'those') across 50 samples (N = 1000, for each determiner type, drawn from Gigaword). In all cases, standard error is no greater than 0.002.



Table 7. Predictions of an expectation-based account and observed parameters

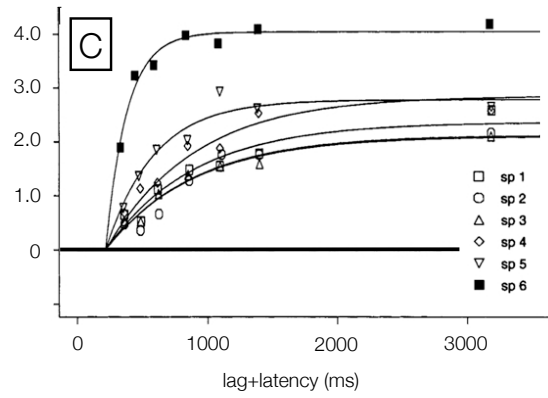
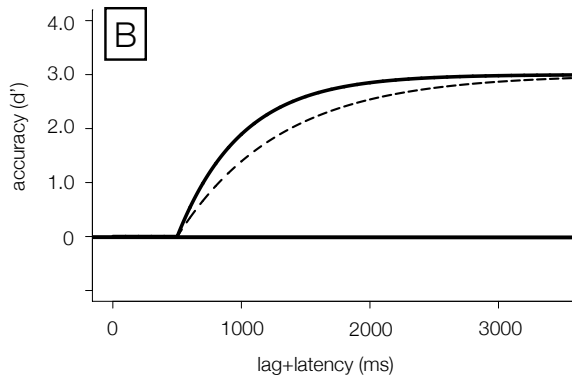
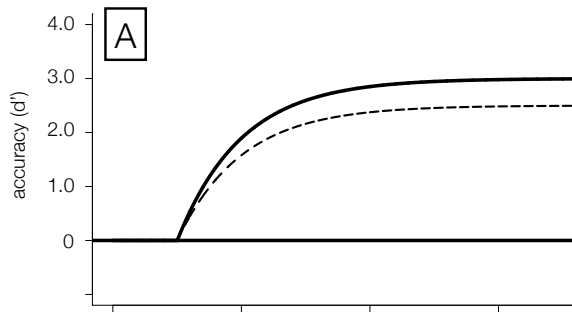
Interveners (i)	P(N   D-i)		PREDICTED FOR NUMBER	$\lambda$ (d')			$\beta^{-1}$ (s)		
	Sg	Pl		SG	PL		SG	PL	
<b>0</b> Det N	0.87	0.78	<i>Advantage for singular</i>	3.11	3.16	✗	0.641	0.641	✗
				3.16	3.16	✗	0.580	0.518	✗
<b>1</b> Det X N	0.76	0.74		2.87	2.75	✓	0.743	0.641	✗
				3.05	2.83	✓	0.880	0.651	✗
<b>2</b> Det X Y N	0.74	0.71		2.81	2.86	✗	0.743	0.641	✗
				2.92	2.90	✓	0.789	0.585	✗
PREDICTED FOR LENGTH	<i>Advantage for shorter</i>		✓	✓*		✓	✗		
			✓	✓*		✓*	✓*		

*Note.* Left-hand columns given the conditional probability of encountering the noun, after having seen the determiner and either 0, 1, or 2 intervening words. Right-hand columns compare these probabilities to the pattern of the SAT parameters (top numbers in each row are from the average discriminative model, bottom numbers from the average value across participant discriminative models). Checks (✓) indicate a correspondence with predictions, crosses (✗) a mismatch, and asterisks (\*) a single reversal in a series. Cells in wavy outline correspond to results that unequivocally mismatch with predictions.

## Figure Captions

Figure 1. The curves are modelled generically as shifted, saturating exponentials (see section 2.4). They can also be fit with the equations from Ratcliff's diffusion model (Ratcliff, 1978), in which case the function's time constant corresponds to diffusion rate.

Figure 2. Average  $d'$  accuracy (symbols) as a function of processing time for judging the acceptability of sentences containing singular and plural determiner phrases. Black symbols represent *Adjacent* conditions; grey symbols the *+1* conditions; and open symbols the *+2* conditions. Smooth curves in each panel show the best fit to of Equation 1, with parameters listed in Table 2. The dotted lines indicate the lag-latency values at which ~63% (or  $100 \times [1 - e^{-1}]$ ) of asymptotic accuracy is achieved. For plural determiners, one lag-latency value characterises all curves. For singular determiners, a second lag-latency value is required for the non-adjacent conditions.



### Panel A

Hypothetical functions differ only by asymptote ( $\lambda$ ).

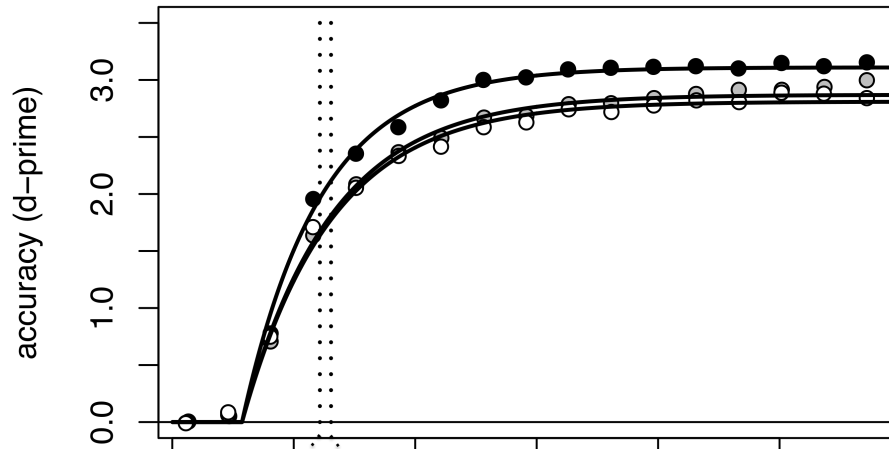
### Panel B

Hypothetical functions differ only by rate ( $\beta$ ).

### Panel C

Data and best fitting functions from a 2-AFC recognition task (McElree & Doshier, 1993). Functions for items in serial positions 1-4 (sp 1-4) vary only in the asymptote. Functions for items in positions 5 and 6 vary in rate as well.

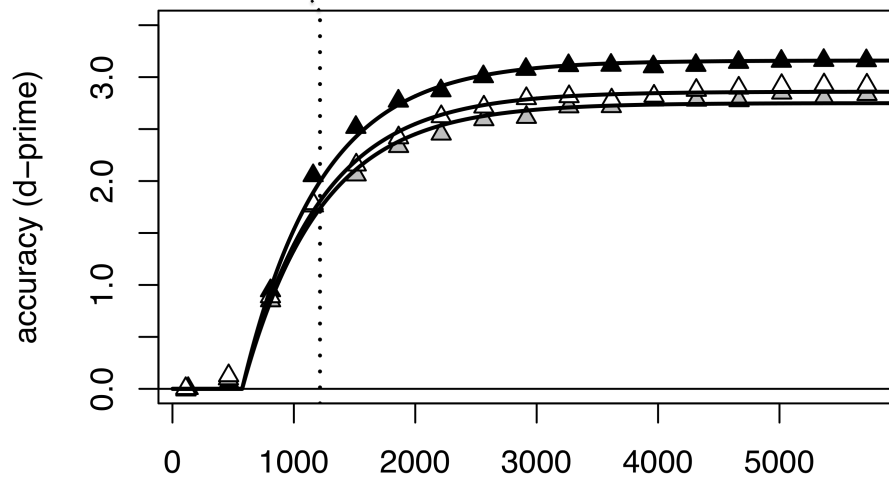
### Singular determiner



1215 ms

1307 ms

### Plural determiner



lag latency (ms)