

# Forgetting in Immediate Serial Recall: Decay, Temporal Distinctiveness, or Interference?

Klaus Oberauer  
University of Bristol

Stephan Lewandowsky  
University of Western Australia

Three hypotheses of forgetting from immediate memory were tested: time-based decay, decreasing temporal distinctiveness, and interference. The hypotheses were represented by 3 models of serial recall: the primacy model, the SIMPLE (scale-independent memory, perception, and learning) model, and the SOB (serial order in a box) model, respectively. The models were fit to 2 experiments investigating the effect of filled delays between items at encoding or at recall. Short delays between items, filled with articulatory suppression, led to massive impairment of memory relative to a no-delay baseline. Extending the delays had little additional effect, suggesting that the passage of time alone does not cause forgetting. Adding a choice reaction task in the delay periods to block attention-based rehearsal did not change these results. The interference-based SOB fit the data best; the primacy model overpredicted the effect of lengthening delays, and SIMPLE was unable to explain the effect of delays at encoding. The authors conclude that purely temporal views of forgetting are inadequate.

*Keywords:* short-term memory, working memory, decay, distinctiveness, interference

The notion of a short-term memory suggests that there are memory traces that vanish quickly over time. This idea matches everyday experience: Often, information that could be repeated immediately after being received is forgotten only seconds later. It is not surprising, therefore, that several theories and models of immediate memory<sup>1</sup> assume that time plays a prominent role in determining how well we remember. Other theories, however, assume that forgetting does not depend on time per se but on interfering events occurring in time.

The role of time in immediate memory is a crucial conceptual issue whose resolution is essential to guiding future theory development. Theories of short-term memory are currently undergoing a marked evolution, with novel approaches being proposed that

arguably represent a fairly radical departure from conventional techniques (e.g., Botvinick & Plaut, 2006) and with existing models being augmented to handle highly constraining new results (e.g., Burgess & Hitch, 2006). There is considerable diversity among theories as to the mechanisms underlying forgetting, with some models citing time as a causal variable (e.g., Burgess & Hitch, 2006), whereas others rely on interference (e.g., Botvinick & Plaut, 2006). Likewise, in the cognate field of working memory research, recent theorizing has variously assumed a central role for time in forgetting (Barrouillet, Bernardin, & Camos, 2004; Towse, Hitch, & Hutton, 2000) or rejected any such role (Saito & Miyake, 2004).

The question whether forgetting in immediate memory is due to the passage of time alone or results from event-based interference is therefore of great theoretical interest. The purpose of this article is to address this fundamental question by adjudicating between two time-based models and one interference-based model of forgetting (Brown, Neath, & Chater, 2007; Farrell & Lewandowsky, 2002; Page & Norris, 1998). We selected these models for the following reasons: (a) They are relatively simple, with no more than four free parameters in their basic versions, and (b) they attribute forgetting to a single cause. These two features together make it relatively easy to trace their behavior to their assumptions about the causes of forgetting. In addition, (c) all three models have been applied to a range of benchmark data with reasonable success (see Lewandowsky & Farrell, in press, for a review).

---

Klaus Oberauer, Department of Experimental Psychology, University of Bristol, Bristol, United Kingdom; Stephan Lewandowsky, School of Psychology, University of Western Australia, Crawley, Australia.

This work was supported in part by Deutsche Forschungsgemeinschaft Grant OB 121/3-3 and in part by departmental funds from the University of Bristol to Klaus Oberauer and by a Large Grant and a Discovery Grant from the Australian Research Council to Stephan Lewandowsky. Stephan Lewandowsky was also supported by an Australian Professorial Fellowship from the Australian Research Council. We are grateful to Gordon Brown and Mike Page for helpful comments on a previous version of this article. We thank Simon Farrell for providing the output timing data from his experiment (Farrell, 2006) and Anupriya Dwivedi and Leo Roberts for assistance during data collection. Supplementary material (i.e., a summary of Experiments 3–5 and the programming code for the primacy model, SOB, and positional decay model) can be found on our Web pages: <http://eis.bris.ac.uk/~psxko/> or <http://www.cogsciwa.com>

Correspondence concerning this article should be addressed to Klaus Oberauer, Department of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, United Kingdom. E-mail: [k.oberauer@bristol.ac.uk](mailto:k.oberauer@bristol.ac.uk)

---

<sup>1</sup> We use *immediate memory* as an umbrella term for short-term memory and working memory, which we regard as largely the same construct investigated with different emphases in different research traditions. As we explain later, our experiments and modeling are relevant for both literatures.

### Three Models of Short-Term Forgetting

#### *Temporal Approaches to Forgetting*

Two notions of the role of time in remembering over the short term have been proposed: time-based decay and temporal distinctiveness. Decay refers to an inexorable decline of memory strength over time, thus necessarily leading to a decreasing chance of correct remembering as time elapses (J. R. Anderson & Lebiere, 1998; Baddeley, 1986; Page & Norris, 1998). Decay theories invariably assume that decay can be counteracted by rehearsal, that is, that memory traces can be reactivated or regenerated by retrieving them before they become inaccessible. The notion of rehearsal is indispensable to decay models because they otherwise could not account for any persistence of immediate memory beyond a few seconds.

The most prominent model of immediate memory that is based on the twin ideas of decay and rehearsal invokes the concept of a phonological loop for verbal material (Baddeley, 1986; Baddeley, Thomson, & Buchanan, 1975). According to this model, verbal material is encoded primarily in phonological representations that decay very quickly, thus preventing reliable recovery of a trace after little more than 2 s unless decay is counteracted by rehearsal. Rehearsal is assumed to be based on articulation, either vocal or subvocal, and its speed is therefore determined by the time it takes to articulate the to-be-remembered material. This assumption elegantly accounts for the fact that people's verbal memory span, that is, the amount of material they can recall perfectly in order of presentation on 50% of trials, equals the amount of material that can be articulated in about 2 s (Baddeley et al., 1975; Schweickert & Boruff, 1986). A computational model of the phonological loop incorporating these ideas, the primacy model, was developed by Page and Norris (1998) and has been shown to capture basic findings in immediate serial recall such as the shape of the serial position curve and the relative frequency of error types.

The notion of temporal distinctiveness, by contrast, proposes that retrieval of an item is driven by the uniqueness of its temporal context, which acts as a cue for retrieval. Each event is associated to a representation of its context, and the context gradually changes over time at a constant rate. A recent formalization of temporal distinctiveness is the SIMPLE (scale-independent memory, perception, and learning) model of Brown et al. (2007; for a critical comment, see Murdock, in press, and the reply by Brown, Chater, & Neath, in press). In SIMPLE, the probability of retrieving an item depends on its distance from all other list items along a temporal dimension. Because the temporal dimension is logarithmically compressed, the psychological distance between items shrinks as they recede into the past. Accordingly, recent events are more readily retrieved than events further in the past. In addition, independent of recency, SIMPLE predicts that items that were widely separated in time at encoding are recalled more easily than items that occurred in close temporal proximity. SIMPLE thereby subsumes what has been known as the ratio rule (Glenberg, Bradley, Kraus, & Renzaglia, 1983), viz., that memory accuracy is roughly proportional to the ratio between interitem intervals and an item's retention interval.

Note that in contrast to decay models, distinctiveness theories such as SIMPLE do not postulate that the quality or strength of memory representations fades over time. By implication, temporal distinctiveness models, unlike decay models, do not require a

compensatory rehearsal process to account for the persistence of any memory beyond a few seconds. The notion of rehearsal is, however, crucial to permitting empirical adjudication between decay theories and interference theories. Any decay model can explain the absence of forgetting, should it arise in the data, by appealing to compensatory rehearsal, and at the same time, it can explain the presence of forgetting, should it occur, by suggesting that rehearsal was withheld or impossible. Control of rehearsal is therefore essential in all research that examines the role of time in short-term memory. We return to this problem in the next section.

#### *Nontemporal Approaches to Forgetting*

Turning to event-based explanations of forgetting, we focus here on the SOB model (Farrell, 2006; Farrell & Lewandowsky, 2002; Lewandowsky & Farrell, in press); SOB stands for serial order in a box, a name that acknowledges the original model's reliance on the brain-state-in-a-box algorithm (e.g., J. A. Anderson, Silverstein, Ritz, & Jones, 1977). The current version of SOB (Farrell, 2006) is an associative network in which items are associated to an event-driven context signal. Unlike the primacy model or SIMPLE, time plays no role in SOB, and the context to which items are associated ticks over not because time has elapsed but because a new event—either a study item or a retrieval from memory—has occurred. In consequence, forgetting in SOB can only occur through interference arising from the encoding of additional information.

An important architectural feature of SOB, one that differentiates it from related precedents, is that encoding is *energy gated* or *novelty sensitive*. The encoding strength of each item is a function of its energy, that is, its novelty or dissimilarity compared with what has already been memorized. Items that are dissimilar to information already in memory are encoded more strongly than items that are similar to existing information. Energy gating has been a core architectural feature of SOB since its inception, derived from the principle that memory avoids encoding of redundant information (see Farrell & Lewandowsky, 2002, for a detailed discussion). Energy gating naturally creates a primacy gradient over a list of items, as long as these are not totally dissimilar. A further prediction from energy gating is that dissimilar items are remembered better when mixed with similar items than in the context of all-dissimilar items; this prediction has been confirmed repeatedly (Farrell, 2006; Farrell & Lewandowsky, 2003; Lewandowsky & Farrell, 2008). Another consequence of this energy-gated encoding is that repetition of interfering information has little additional effect on memory; this consequence turns out to be crucial to accommodating recent data on forgetting in short-term memory.

The main goal of this article is to thoroughly examine the role of time in forgetting from immediate memory and to contrast the explanatory power of temporal and nontemporal approaches to forgetting. To resolve this overarching theoretical question, we present new data within a paradigm that bridges research traditions in short-term memory and working memory. Those data are then used to adjudicate between the three computational models just introduced, namely, the primacy model, SIMPLE, and SOB. We show that both time-based models have difficulty in handling these new data, whereas SOB provides a reasonably good quantitative account.

We next explain the rationale of the experimental design we used to disentangle the effects of time and interference on forgetting and then derive predictions from the three models for this design.

### Studying the Effects of Time on Forgetting

The assumption that memory traces decay rapidly has found much support in the word-length effect: The word-length effect refers to the fact that lists of words that take longer to articulate (e.g., *hippopotamus*, *artichoke*, *difficulty*) are typically recalled less accurately in correct order than lists composed of shorter words such as *bat* or *putt* (Baddeley et al., 1975). The word-length effect has generated a large amount of research whose outcomes form a pattern of great complexity. Instead of reviewing this complex pattern here, we argue that there are strong conceptual grounds on which the word-length effect should be rejected as a suitable empirical tool for adducing support for time-based decay. The full details of those arguments are available elsewhere (Lewandowsky & Oberauer, in press).

Beginning with the work of Baddeley et al. (1975), there have been two kinds of word-length effect in the literature. The first arises when the number of syllables of the to-be-remembered words is varied and gives rise to a robust effect: Generally, lists of multisyllabic words are recalled worse than lists of monosyllabic words. This manipulation of word length is, however, necessarily confounded with phonological complexity—and complexity, not articulation duration, has been shown to be the critical variable when the two factors are disentangled (Service, 1998). It follows that the multisyllabic word-length effect is of little relevance to examinations of the decay notion.

The second, more subtle word-length effect arises from manipulating the articulation duration of words while keeping the number of syllables and phonemes constant, by contrasting words having long phonemes (e.g., *canteen*) with words having short phonemes (e.g., *ticket*). In apparent support of the decay notion, Baddeley et al. (1975) found that lists consisting of short-phoneme words were recalled more accurately than lists of long-phoneme words. This effect is replicable with the same small set of words, but no effect, or even a reversed effect, is obtained with other sets of words constructed by the same principle (Lovatt, Avons, & Masterson, 2000). These conflicting results point to an inevitable difficulty in studies using the natural articulation duration of words to manipulate time: Word length can never be completely separated from possible confounding features of the words being contrasted (for a detailed discussion, see Lewandowsky & Oberauer, in press). We therefore conclude that other means of manipulating time that are less fraught with inevitable confounds should be employed to differentiate between decay, temporal distinctiveness, and interference-based forgetting in immediate memory.

Although it is straightforward to manipulate time, for example, by simply pacing the rate at which items are presented (Lewandowsky, Brown, Wright, & Nimmo, 2006; Nimmo & Lewandowsky, 2005), this opens up the problem of how to prevent rehearsal. Expanding an unfilled interval between presentation of individual items or between study and test may encourage participants to rehearse already-presented material, thus masking any effect—for example, of decay—that might otherwise have occurred during that time. A common method for preventing re-

hearsal is by use of a distractor task, such as counting backward from some random number. The problem with this approach is that every distractor task involves representations and processes that could arguably interfere with memory, and any finding of decreased memory performance over longer distractor-filled delays is open to the interpretation that forgetting was produced by interference rather than decay or diminished temporal distinctiveness.

Lewandowsky, Duncan, and Brown (2004) introduced a paradigm that prevents articulatory rehearsal and largely deconfounds time and amount of interference. In one of their studies, participants orally recalled a list of letters while repeating an irrelevant word (*super*) aloud between retrievals. This manipulation is known as *articulatory suppression* (AS from here on) and, according to Baddeley (1986), engages the component of the phonological loop that would otherwise perform subvocal rehearsal, thereby preventing the reactivation of memory traces. Lewandowsky et al. found that recall performance was unaffected by the number of times the distractor was repeated between memory retrievals, implying that time per se does not cause forgetting in serial recall. Lewandowsky et al. buttressed their conclusion by a model-based analysis involving SIMPLE: The results of every participant were fit worse by a completely time-based model than by an event-based version of SIMPLE that relied on a positional (rather than temporal) dimension. Conversely, eliminating the temporal dimension from the model did not appreciably worsen the fit for any participant, confirming that forgetting in immediate memory must involve a process other than the passage of time.

The conclusions by Lewandowsky et al. (2004) hinge on the assumption that AS eliminated rehearsal during retrieval. There is broad consensus in the literature that AS interferes with rehearsal (Baddeley, 1986, pp. 37, 86; Baddeley & Lewis, 1984; Page & Norris, 1998, pp. 764, 770), implying that manipulating the duration of AS between retrieval events constitutes a manipulation of retention time without possibility of compensatory rehearsal. Moreover, because repetition of the same word two or three times does not involve any new representations beyond those generated for its first utterance, increasing time through repetition arguably involves very little, if any, additional interference or demands for attentional resources, which are often implicated in explaining the limitations of working memory (e.g., J. R. Anderson, Reder, & Lebiere, 1996).

We therefore based our model comparison on data from six new experiments (two of which we present in detail) that extended the paradigm of Lewandowsky et al. (2004). One goal of these experiments was to obtain a broader database for modeling. Another goal was to bridge the findings of Lewandowsky et al., which emerged from the short-term memory literature, with relevant work in the working memory arena, in which the issue of time-based forgetting is also hotly debated (Barrouillet et al., 2004; Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Oberauer & Kliegl, 2006; Saito & Miyake, 2004; Towse et al., 2000). These two literatures have been surprisingly isolated from each other given that they use similar paradigms and struggle with similar issues.

In this light, it is interesting to note the similarity between the paradigm of Lewandowsky et al. (2004) and the complex span procedure, the paradigm most frequently used to study working memory (Conway et al., 2005). In the complex span task, a

processing task (e.g., reading a sentence or solving an arithmetic equation) alternates with encoding of to-be-remembered items. In one version of the complex span task studied by Barrouillet et al. (2004), the processing task consisted of repeating an utterance (“baba”) several times, thus effectively alternating AS with encoding of list items. To link this paradigm to the studies by Lewandowsky et al., the present Experiment 1 interspersed AS not only between items at retrieval but also between encoding events, thus bridging research on the effects of time in immediate serial recall and in working memory.

The present Experiment 2 further extended the bridge between the short-term memory and the working memory literatures. Barrouillet et al. (2004, 2007) have argued that memory performance in the complex span task is determined by the interplay of time-based decay and rehearsal. Specifically, they argued that memory traces decay over time but can be refreshed—without requiring articulatory rehearsal—by retrieving them. Retrieval of memory items for refreshing competes for an attentional bottleneck with the processing task that must be performed between presentations of the list items. The bottleneck is assumed to switch rapidly between the two demands, implying that even small pauses in which the distractor task makes no processing demands can be used to refresh memory traces. In consequence, the time-based resource-sharing model of Barrouillet and colleagues predicts that memory performance is a function of the rate at which the distractor task demands operations rather than of the total elapsed time. This prediction was confirmed in several studies using the complex span paradigm with a computer-paced processing task (Barrouillet et al., 2004, 2007; Barrouillet & Camos, 2001; Lepine, Barrouillet, & Camos, 2005). When the pace was varied independently of total elapsed time, the determinant of performance turned out to be rate of processing, not absolute time.

One important feature of the time-based resource-sharing model is that its presumed mechanism for refreshing of memory items does not depend on subvocal articulation. Rather, the bottleneck is conceptualized as a domain-general attentional mechanism that supports all central cognitive processes, including retrieval and response selection (for a similar idea, see the concept of refreshing in the MEM framework of Johnson, 1992; Raye, Johnson, Mitchell, Reeder, & Greene, 2002; the idea of a second rehearsal mechanism besides articulatory rehearsal has also been discussed by Vallar & Baddeley, 1982). Evidence for such a bottleneck comes from extensive research on dual-task performance, in particular with the psychological refractory period (PRP) paradigm (Pashler, 1994; Pashler, Johnston, & Ruthruff, 2000; Ruthruff, Pashler, & Klaassen, 2001). Work with this paradigm has firmly established that response selection in speeded choice tasks requires the bottleneck (Pashler, 1994), as does retrieval from long-term memory (Rickard & Bajic, 2004).<sup>2</sup>

Repeated articulation of the same utterance is unlikely to occupy the bottleneck except for a brief period during which the speech plan is established (Naveh-Benjamin & Jonides, 1984). Therefore, from the perspective of Barrouillet et al.’s (2004) time-based resource-sharing model, one could argue that saying “super” three times in a row does not engage the bottleneck for longer than saying “super” only once and that people may therefore be able to refresh memory items concurrently with AS not by conventional articulatory rehearsal but by an attentional form of rehearsal that we refer to as refreshing (cf. Raye, Johnson, Mitchell, Greene, &

Johnson, 2007). It follows that the results of Lewandowsky et al. (2004) might be compatible with time-based decay after all if we assume that refreshing can occur concurrently with repeated articulation, thus counteracting forgetting during the AS-filled delays. This perspective motivated the present Experiment 2 in which we again manipulated time between items at retrieval and encoding while preventing both articulatory rehearsal and refreshing, using two distractor tasks (AS and a speeded choice task), each of which is known to disrupt one of the hypothetical mechanisms by which decaying traces could be revived.

The remainder of the article is organized as follows. First, we present predictions of the three models under investigation for the design of Experiment 1, which involved AS-filled delays between items during recall or during encoding. Next, we present the first experiment, which tested these predictions. We then seek remedial modifications that might enable the primacy model and SIMPLE to accommodate the present data, and we also optimize the quantitative fit of SOB to the same data for comparison. Next, we present Experiment 2, which, in addition to using AS to block articulatory rehearsal, employed a speeded choice task to block attention-based rehearsal. These data permit a test of the three models with stronger constraints on rehearsal. We finish with an in-depth investigation of the interplay of decay and rehearsal in decay-based models.

### Model Predictions

Figure 1 illustrates the predictions of the primacy model (Page & Norris, 1998), SIMPLE (Brown et al., 2007), and SOB (Farrell, 2006) concerning the effects of time between items at encoding and at retrieval when rehearsal is eliminated by AS. Each panel shows serial position curves for three conditions: a baseline condition in which items are presented and serially recalled without any temporal delays, two conditions in which delays filled with AS are present during encoding only (either one or three repetitions of an irrelevant word), and two conditions in which delays filled with AS are present during retrieval only (one or three repetitions). The latter two conditions correspond to Experiment 2 of Lewandowsky et al. (2004).

These predictions were generated on the basis of published parameter values. For the primacy model and SIMPLE, we used the parameter values originally used for the simulations of dissimilar lists in Experiment 1 of Henson, Norris, Page, and Baddeley

<sup>2</sup> In the PRP paradigm, two speeded choice reactions to two stimuli are required in close temporal succession. Reaction times to the second stimulus increase with decreasing stimulus onset asynchrony (SOA) between the two stimuli. This effect is interpreted by assuming that the central component of the second task has to wait until the bottleneck has completed the central component of the first task. Increasing the difficulty of the sensory component of the second task (e.g., by blurring the second stimulus) affects reaction times to that stimulus only at long SOAs, suggesting that at short SOAs, sensory processing of the second task can occur concurrently with central processes on the first task such that even the prolonged sensory processing of the second task is finished before the bottleneck can start working on the second task’s central processes. In contrast, manipulations that affect the difficulty of response selection of the second task (e.g., arbitrary vs. compatible stimulus–response mappings) are additive with SOA, suggesting that response selection has to wait for the bottleneck. These observations provide evidence that central but not sensory processes require the bottleneck (Pashler, 1994).

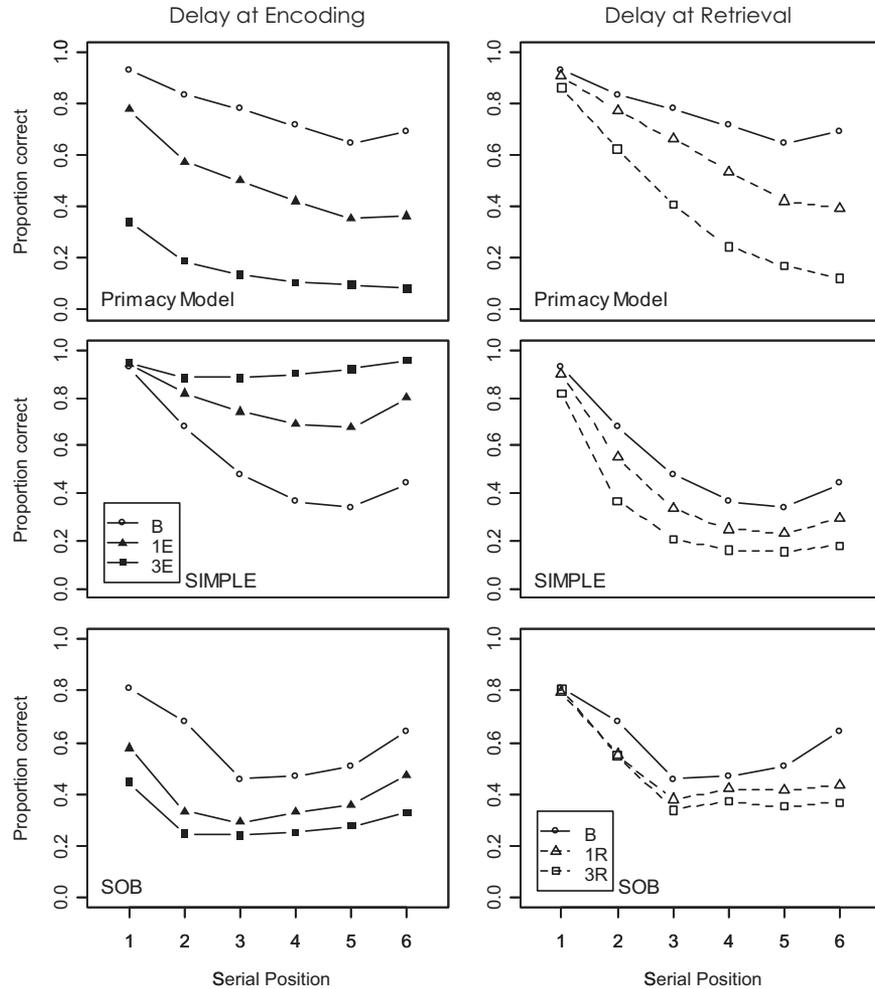


Figure 1. Predictions of the primacy model (top), SIMPLE (middle), and SOB (bottom) for delays at encoding (left) and delays at retrieval (right). Predictions for all conditions were generated using published parameter values. Parameter values for the primacy model were decay rate  $D = 0.27$ , noise  $N = 0.23$ , retrieval threshold  $T = 0.49$ , threshold noise  $M = 0.74$ . Parameter values for SIMPLE were generalization gradient  $c = 8.90$ , omission threshold = 0.49, omission slope = 8.12. SIMPLE used only the temporal dimension to compute distinctiveness (i.e., the  $wt$  parameter introduced by Lewandowsky, Duncan, & Brown, 2004, was set to 1). Parameter values for SOB were  $\phi_e = 484$ ,  $\phi_s = 1.21$ ,  $N_O = 1.43$ ,  $c = 0.30$ . B = baseline condition without distractors; 1E = one distractor word at encoding; 3E = three distractor words at encoding; 1R = one distractor word at retrieval; 3R = three distractor words at retrieval.

(1996), as published in Page and Norris (1998) and in Brown et al. (2007), respectively.<sup>3</sup> For SOB, we used the parameter values published by Farrell (2006) for recall of dissimilar lists (short-delay condition), in an experiment very similar to that of Henson et al. Parameter values are given in the caption of Figure 1. Readers unfamiliar with the three models are referred to the later section on quantitative model comparison, which provides a detailed description of each model and its parameters.

The two time-based models, SIMPLE and the primacy model, require additional assumptions about the timing of events. The experiment of Henson et al. (1996) used a presentation time of .5 s per item. Following Page and Norris (1998), we assumed that people used only part of this time (250 ms) for encoding, thus leaving the remaining 250 ms for rehearsal after each item. Ac-

cording to the primacy model, people use this time to rehearse, as long as the time for a complete rehearsal of the items encoded so far does not exceed the available time window. The time for rehearsing an item was set to 250 ms, according to Table 1 in Page and Norris; thus, only the first list item could be rehearsed once after its encoding. Recall times were not recorded by Henson et al.,

<sup>3</sup> The published application of SIMPLE to that experiment involved a weighted combination of two dimensions of distinctiveness, viz., time and phonological similarity. Because here we are interested in predictions from temporal distinctiveness only, we set the parameter that weights the contribution of the temporal dimension,  $wt$ , to 1, thus obtaining predictions from the unidimensional time-based version of SIMPLE.

and the authors of the two time-based models made different assumptions about recall times: Page and Norris (Table 1) assumed that people recalled at a rate of .5 s per letter. Brown et al. (2007) made the more realistic assumption that recall took 1 s per letter. Because the published parameter estimates are contingent on the respective assumptions about recall times, we maintained the assumptions made in the context of each published model for deriving the predictions. For both time-based models, the duration of saying each distractor word aloud during AS was set to 350 ms, on the basis of the data of Lewandowsky et al. (2004).

The event-based SOB modeled the AS-filled delays by encoding each repetition of the distractor word (“super”) into memory using the energy gating that is intrinsic to the model. We used the exact version of SOB presented by Farrell (2006). SOB uses distributed representations of items, which enables modeling their similarity as the cosine of the vectors representing each item. We set interitem similarity to .70 (see below for more details about how similarity was implemented), a value chosen to approximate the average similarity among vectors in a recent application of SOB (Lewandowsky & Farrell, 2008) that relied on a behavioral multidimensional scaling solution of the actual stimuli (i.e., letters). AS was modeled by associating each distractor token to the immediately preceding context marker. Thus, depending on condition, a single copy of “super” or three identical copies of “super” were encoded between each pair of study items or between retrievals. To capture the fact that “super” differs considerably from a list of letters, the representation of the distractor item was chosen to be less similar (by a factor of .5) to the list items than they were to each other.

As is apparent from the figure, for delays at retrieval, both time-based models predict performance to decline with additional repetitions of the distractor. In the primacy model, this decline is generated by decay, whereas, in SIMPLE, it results from an increasing loss of distinctiveness because the logarithmic compression of time moves events closer together in psychological space as they recede into the past. Moreover, both models predict that the serial position curves fan out because the cost of increasing the delays between successive recall attempts cumulates toward the end of the list. Importantly, the primacy model predicts that the degree of fanning is proportional to the added delay so that the difference between one and three distractors at retrieval is larger than that between the baseline and a single distractor (this effect is mellowed in the recency portion by a floor effect on the three-distractor condition). This prediction arises because the amount of decay is a function of the length of the delay, and three distractors take 3 times as long to articulate as a single distractor. SIMPLE, by contrast, predicts a large effect from inserting a single distractor word between retrievals and a comparatively small additional effect from adding another two distractors. The first distractor diminishes distinctiveness more than the additional distractors because, on the logarithmic time scale, additional delays have increasingly smaller effects on the psychological distances.

SOB predicts a large decrement of memory by a single distractor word between retrievals but a much smaller additional effect of adding two more distractor words, and that effect is confined primarily to the later serial positions. The diminishing effect of additional distractors beyond the first one arises from the energy-gated encoding that is central to SOB: Once the distractor has been associated to a given context, subsequent occurrences of the same distractor in the same context are necessarily given a very small encoding weight.

The predictions of all three models seem to deviate from the existing results of Lewandowsky et al. (2004), who found no interaction between serial position and the number of distractors. However, the serial-position curves of their Experiment 2 did show a small degree of fanning that, although not statistically reliable, might be quantitatively compatible with the limited fanning predicted by SIMPLE and SOB for the contrast of one versus three distractors.

For delays at encoding, the predictions of the two time-based models differ substantially. The primacy model predicts that performance declines with increasing delays and that the effect is fairly evenly distributed over serial positions. Again, this prediction reflects the effect of decay, which increases with the delay interval. SIMPLE, by contrast, predicts positive effects of adding distractors at encoding. These predictions arise because temporal distinctiveness increases with greater spacing between items. This benefit is larger than the loss of temporal distinctiveness incurred by the increased lag between encoding and recall of each item, which reduces distinctiveness due to the logarithmic compression of the temporal dimension. The spacing advantage is largest at the end of the list because the negative effect of delay is smallest for terminal list items.<sup>4</sup>

The encoding conditions therefore differentiate between the strong and opposing predictions by the two time-based models. SOB predicts that adding a single distractor word at encoding strongly impairs performance across all serial positions. Adding two more distractors should lead to a further but comparatively small drop in performance. As noted earlier, the diminishing effect of additional distractors beyond the first one arises from the energy-gated encoding in SOB.

To highlight the theoretically most relevant differences between the models’ predictions, we computed two indices, one representing the effect of increasing delays at encoding and one representing the effect of increasing delays at retrieval. Our index for delays at encoding simply measured the main effect of delay on accuracy. Thus, the *time effect at encoding* (TEE) was computed as  $PC(\text{long delays; 3 distractors}) - PC(\text{short delays; 1 distractor})$ , averaged across serial positions ( $PC$  stands for proportion correct). Negative values thus reflect a loss of accuracy incurred by longer delay, whereas positive values reflect an advantage for long delays. Our index for delay effects at retrieval was the *time-loss slope* (TLS), that is, the amount of accuracy lost per unit time. To compute the TLS, we first measured loss of accuracy as  $PC(\text{long delay; 3 distractors}) - PC(\text{short delay; 1 distractor})$ ; a performance loss with longer delays is thus again reflected in a negative value. This difference was obtained for each serial position. We also computed the difference in time between the two delay conditions. Time was measured from the end of list presentation until completion of recall for each serial position. These times were the same as those we assumed for the simulations for the primacy model and SIMPLE; for SOB, we used the recall times measured in the experiment of Farrell (2006) for the baseline condition augmented by the distractor times observed by Lewandowsky et al. (2004) as required. The TLS is the accuracy difference divided by the time difference (in seconds). A TLS score was computed for each serial position and then averaged across all positions.

<sup>4</sup> SIMPLE’s predictions differ when temporal discriminability between items is scaled to the total temporal envelope of the study–test event; we explore this option later.

To investigate the range of predictions within a plausible region of the parameter space, we generated distributions of predicted TEE and TLS indices from the three models as follows. Starting from the parameters that generated the predictions in Figure 1, we varied each parameter value in both directions until the model's deviation from its original predictions for the baseline condition, assessed by the root-mean-square deviation (RMSD), reached 0.08. The RMSD was defined as

$$\text{RMSD} = \sqrt{\sum \frac{(\text{pred} - \text{obs})^2}{N}} \quad (1)$$

where  $N$  refers to the number of data points being fitted. The parameter values obtained in this way were then considered to be the vertices of a parameter space, which we explored by grid search. Five equidistant points were found within the range of each parameter (including the vertices), resulting in five values for each parameter. We obtained model predictions for all combinations of the five values for each parameter and computed the TLS and TEE indices from these predictions. This procedure generated a distribution of index values from  $5^k$  simulation runs (where  $k$  is the number of parameters) within plausible ranges of parameter values. The distributions of the TEE index for the three models are shown in Figure 2, and those of the TLS index are shown in Figure 3. The figures show that the two indices together clearly discriminate between the three models.

The vertical lines in Figure 2 and 3 represent the mean indices observed in all experiments we conducted so far with the paradigm of Lewandowsky et al. (2004, Experiment 2) and its extensions. We present two of these experiments in detail below, the data of which were used for evaluating the quantitative model fits. The TEE and TLS values of the remaining experiments are summarized in Table 1, which also provides a summary of the methodological differences between these experiments. The two experiments presented in detail were selected by two criteria, namely, coverage (i.e., including delays at both encoding and retrieval) and tightness of control over rehearsal opportunities. The selection was made without consideration of the outcome of the modeling. Details of the unpublished experiments can be found at the Web addresses provided in the author note.

The figures and the table underscore the quantitative approach taken in this article. Rather than relying entirely on the binary logic of significance testing, we aimed at measuring the effects of time during encoding and retrieval as precisely as possible, and we evaluate the theoretical hypotheses by how well they serve, in the context of a model, to quantitatively fit these effects.

Only the predictions of SOB overlapped with the observed TLS indices at retrieval, whereas none of the models predicted the TEE indices satisfactorily. After presenting Experiment 1, we investigate how well modified versions of these three models can account for the data.

### Experiment 1

The first experiment tested the preceding predictions of the three models. Participants said aloud a distractor word once or several times to manipulate the interval between items at encoding or retrieval in a serial recall task. We tested participants in five conditions: a baseline condition without any delays at either en-

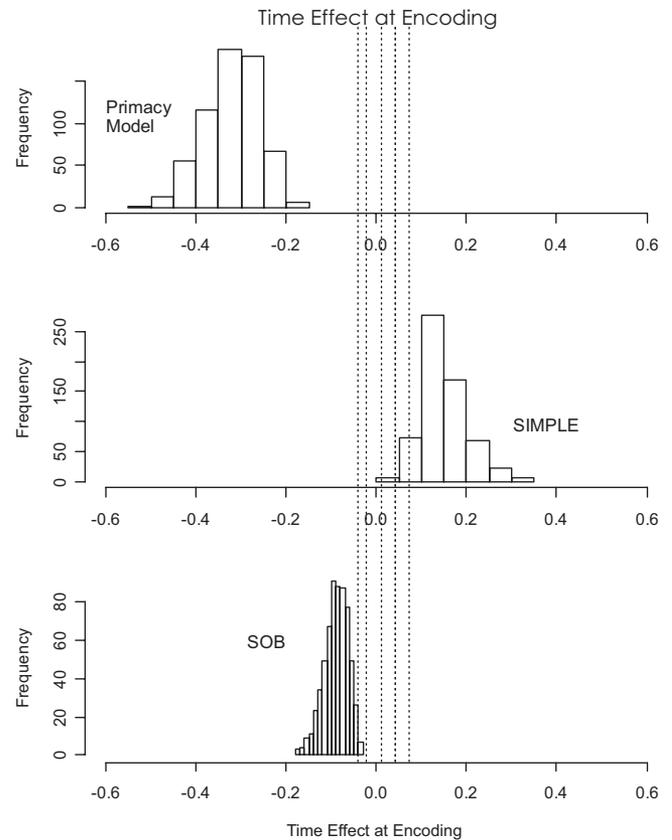


Figure 2. Time effects at encoding represent the loss of proportion correct due to long versus short delays at encoding. Distribution of predictions from the primacy model (top), SIMPLE (middle), and SOB (bottom). See text for details about how the distributions were obtained. Each dotted vertical line represents the mean time effect at encoding measured in one of the experiments summarized in Table 1.

coding or retrieval, two conditions with one versus three repetitions of the distractor preceding each item at retrieval, and two conditions with one versus three repetitions of the distractors following each item at encoding.

### Method

**Participants.** Participants were 20 students from the University of Potsdam, Potsdam, Germany, who took part for course credit.

**Materials and procedure.** Participants had to recall lists of six consonants in correct order. Each list was generated by randomly ordering the consonants H, J, M, Q, R, and V. At the beginning of each trial, the condition was displayed for 2 s (e.g., “3 × super at retrieval”). Participants then saw a sequence of six screens. In the conditions without delays at encoding, each screen contained only the letter to be remembered, printed in red at the top. In the conditions with delay at encoding, “super” was displayed either once or three times below the letter (in blue font, on separate rows). Participants read the letter and the distractors (if any) aloud, and as soon as they were finished, the experimenter pressed the space bar, proceeding to the next screen. This experimenter-paced

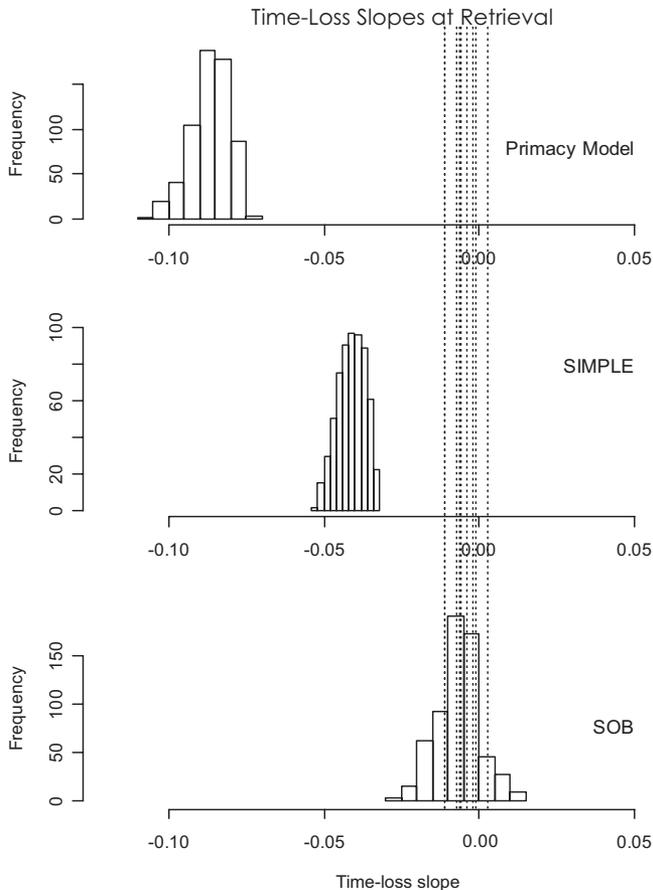


Figure 3. Time-loss slopes represent the decline in proportion correct per second due to long versus short delays at retrieval. Distribution of predictions from the primacy model (top), SIMPLE (middle), and SOB (bottom). See text for details about how the distributions were obtained. Each dotted vertical line represents the mean time-loss slope measured in one of the experiments summarized in Table 1.

procedure was chosen because research with the complex span paradigm, which is similar to the present procedure, has shown that experimenter-paced timing effectively reduces slack time and thereby diminishes the opportunity for rehearsal (Friedman & Miyake, 2004). To further reduce rehearsal opportunities, we instructed participants to read the letters and the “supers” quickly and without pauses.

After the screen presenting the last list item, the German word for *recall* was displayed for 1 s, after which the first of six recall screens was displayed. Each recall screen showed the serial number of the item to be recalled accompanied by a question mark. In the conditions with delays at retrieval, either one or three rows with “super” were shown above this recall prompt. Participants were instructed to first read each “super” (if any) aloud and then to recall the letter in the requested serial position aloud. As soon as the participant has recalled a letter, the experimenter pressed the space bar to move on to the next screen and recorded the recalled letter.

After a practice block with six trials, participants completed four blocks of 15 trials each. The order of conditions was randomized within the set of 60 test trials, with the constraint that 12 trials were presented for each condition.

## Results

The top panel of Figure 4 shows the serial position curves for correct recall in position, averaged across participants, for the baseline condition and the two conditions with delay at encoding. The bottom panel of Figure 4 shows the serial position curves for the two conditions with delay at retrieval, complemented again with the baseline condition for comparison.

The data were analyzed by a within-subjects analysis of variance (ANOVA) with serial position (6) and condition (5) as independent variables. Serial position was coded as the linear contrast over serial position. Condition was coded by four orthogonal contrasts. The AS contrast compared the baseline with the four conditions having delays filled by AS. The encoding-retrieval contrast compared the two conditions with delays during encoding with the two conditions with delays during retrieval, averaging across the number of repetitions of “super”. The delay-at-encoding contrast compared one “super” versus three “supers” at encoding. Finally, the delay-at-retrieval contrast compared one versus three “supers” at retrieval. The ANOVA results are summarized in Table 2.

Of critical interest are the delay-at-encoding and the delay-at-retrieval contrasts. The main effect of delay at encoding tests the effect of time at encoding, which is predicted to be positive by SIMPLE, slightly negative by SOB, and strongly negative by the primacy model. This is the effect captured by the TEE index in Figure 2. The main effect of the delay-at-retrieval contrast and its interaction with the linear contrast of serial position together test the fanning-out pattern at retrieval, which is predicted to be a small effect by SOB, a somewhat larger effect by SIMPLE, and a large effect by the primacy model. The degree of fanning out is captured by the TLS index in Figure 3.

Figure 4 shows that all four conditions with delays resulted in much worse recall than the baseline condition; this is reflected in the large effect of the AS contrast. Delays damaged memory to the same extent at encoding and at retrieval, as shown by the absence of an effect of the encoding-retrieval contrast. The number of distractors at encoding had no effect at all—neither the delay-at-encoding contrast nor its interaction with serial position approached significance. The number of distractors at retrieval, in contrast, had a small but significant negative effect on recall, as reflected by the delay-at-retrieval contrast. The interaction of this contrast with serial position reflects the fanning-out pattern predicted by all three models, albeit to different extents.

Table 3 shows the mean times for encoding and retrieving of an item, which include the times for reading the distractors in the conditions with delays (five individual latencies exceeding 10 s were removed as outliers).<sup>5</sup> The times for encoding of an item were around 900 ms in all three conditions in which there was no delay at encoding. This time increased by 320 ms with a single “super” and by another 850 ms when three “supers” were required. Thus, each “super” added a delay of about 400 ms at encoding.

Looking at the retrieval times, we first notice that retrieval took longer in the conditions with delays at encoding than in the baseline condition. This increase in retrieval times likely reflects the increased difficulty of recall when there had been a delay at

<sup>5</sup> The criterion for defining outliers in our experiments was determined from plots of the latency distributions, following the advice of Ulrich and Miller (1994) to eliminate not more than 1% of the data.

Table 1  
*Mean Time Effects at Encoding and Time-Loss Slopes From All Experiments Conducted in Our Labs*

Experiment	TEE (SD)	95% CI	TLS (SD)	95% CI
LDB1			-.0111 (.020)	[-.0245, .0023]
E1 (AS-only)	-.0222 (.108)	[-.0727, .0282]	-.0040 (.083)	[-.0430, .0350]
E2, AS-only	.0425 (.091)	[.0001, .0849]	.0026 (.022)	[-.0076, .0129]
E2, AS+CRT	-.0404 (.095)	[-.0849, .0041]	-.0112 (.015)	[-.0181, -.0044]
E3 (AS-only)	.0415 (.123)	[.0144, .0973]	-.0021 (.051)	[-.0246, .0204]
E4, AS-only			-.0059 (.027)	[-.0175, .0006]
E4, AS+CRT			-.0073 (.023)	[-.0175, .0029]
E5, AS-only	.0726 (.114)	[.0179, .1274]		
E5, AS+CRT	.0111 (.134)	[-.0537, .0758]		
E6, AS+CRT			-.0012 (.008)	[-.0047, .0023]
E7 (AS-only)			-.0063 (.205)	[-.0929, .0803]
Meta-analysis	.0173	[-.0140, .0490]	-.0048	[-.0075, -.0021]

*Note.* Time-loss rates are defined as accuracy (on a scale from 0 to 1) lost per second. Timing data were available for only 5 participants of Experiment 2, so we could not compute TLS for it. E1 and E2 are the present Experiments 1 and 2, respectively; E6 is Experiment 4 in Lewandowsky, Geiger, and Oberauer (in press). E7 combines the “simple-burst” conditions of Experiments 1 and 2 in Lewandowsky et al. (in press); the remaining experiments are unpublished. Experiment 3 (E3): like E1, but with separate screens for each distractor and fixed presentation times (1 s) for each letter and each distractor at encoding. Experiment 4 (E4): like E2, with .5 s presentation time of letters, and CRT with stimulus–response compatible mapping (left/right keypress in response to left/right asterisk). Experiment 5 (E5): like E4, but with distractors at encoding and instruction to read letters aloud. Experiment 6 (E6): like E2, with 1 s presentation time of letters. Experiment 7 (E7): like E1, but with the identity of the to-be-articulated distractor word changing across trials and participants. AS = articulatory suppression; CRT = choice reaction task; LDB1 = Lewandowsky, Duncan, and Brown (2004, Experiment 1); TEE = time effects at encoding; TLS = time-loss slopes.

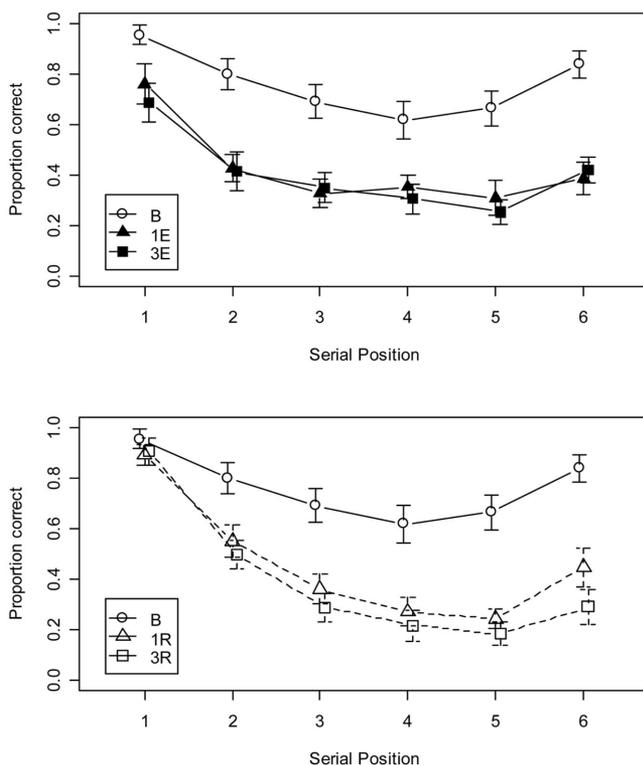


Figure 4. Serial position curves observed in Experiment 1. Top panel: Baseline and conditions with delay at encoding. Bottom panel: Baseline and conditions with delay at retrieval. Error bars represent 95% confidence intervals for within-subject comparisons, computed by the method of Bakeman and McArthur (1996). B = baseline (no delay); 1E = one distractor word at encoding; 3E = three distractor words at encoding; 1R = one distractor word at retrieval; 3R = three distractor words at retrieval.

encoding. Times per item in the conditions with delays at retrieval were even longer, reflecting the time required for reading the distractor words. The difference between the condition with one “super” at retrieval and the baseline amounts to 870 ms, suggesting that saying “super” at retrieval took about twice as long as during encoding. This impression is probably misleading because delays at retrieval increased task difficulty about as much as delays at encoding—as shown by mean accuracy of recall—and it is likely that this increased difficulty also affected the time for retrieving an item. If this is true, it would be more adequate to compare the time spent on retrieval in the condition with one “super” at retrieval with the time spent on retrieval in the condition with one “super” at encoding, and likewise for the two conditions with three distractors. From this comparison, we calculate 410 ms for saying one “super” at retrieval, and 1,050 ms for saying three “supers.” Thereby, the time for saying each “super” at retrieval can be estimated to be between 350 and 400 ms, similar to the time estimated for saying “super” during encoding. We compared these times with the speech rate measured by Allen and Hulme (2006), who asked participants to repeat monosyllabic words 10 times as fast as possible. Their mean speech rate of 3.3 words per second translates into 300 ms per word. Our participants took only 50–100 ms longer to pronounce a bisyllabic word. We conclude that participants in Experiment 1 complied well with the instruction to read the distractor words as quickly as possible and without generating extensive pauses for articulatory rehearsal.

### Discussion

The present experiment yielded a small but significant effect of long versus short delays at retrieval, reflected in the limited fanning out of serial position curves. Although this finding is in apparent contradiction to the experiments of Lewandowsky et al. (2004), who found no significant fanning, the two sets of results are roughly comparable in magnitude. Specifically, the TLS ob-

Table 2  
Summary of Analysis of Variance Results for Experiment 1

Contrasts	<i>F</i>	<i>p</i>	Partial $\eta^2$
Serial Position	330.1	<.001	.95
Articulatory Suppression (AS)	133.9	<.001	.88
Encoding Retrieval	0.4	.52	.02
Delay at Encoding	0.9	.37	.04
Delay at Retrieval	10.1	.005	.35
Serial Position $\times$ Articulatory Suppression	48.4	<.001	.72
Serial Position $\times$ Encoding Retrieval	20.3	<.001	.52
Serial Position $\times$ Delay at Encoding	0.6	.45	.03
Serial Position $\times$ Delay at Retrieval	7.8	.01	.29

*Note.* Serial position refers to the linear contrast of serial position. Degrees of freedom were 1, 19.

served in this study ( $-.004$ ; see Table 1) is actually slightly smaller than that observed in one of the experiments of Lewandowsky et al. ( $-.01$ ), underscoring the need to interpret these data quantitatively rather than relying exclusively on the binary outcome of significance tests.

The need to consider the data quantitatively is further mandated by the fact that all three models qualitatively predict fanning with delays at retrieval; however, at a quantitative level, the primacy model predicts larger fanning and a smaller drop in performance from the baseline to the condition with one “super” at retrieval than was observed (see Figures 1 and 3). SIMPLE correctly predicts a large drop of accuracy from baseline in both conditions with delayed retrieval, but it overpredicts the amount of fanning, as captured by the TLS index (see Figure 3). By contrast, the predictions of SOB match the relative amounts of forgetting for delays at retrieval fairly well.

In addition, the experiment yielded two new findings. First, increasing delays between presentation of items did not result in worse performance: With three distractors at encoding, memory was as good as with one distractor word. Assuming that rehearsal was effectively blocked by AS, this finding is incompatible with time-based decay because longer delays should lead to more decay.

The second new finding is that, in comparison with the baseline (not used by Lewandowsky et al., 2004), any disruption of either encoding or retrieval by a short or long period of repetitive articulation led to a dramatic loss of memory for serial order. This effect can be interpreted in several ways. First, it could be due to the attentional demands of executing a second, unrelated task or of switching between the two tasks. Second, the interleaved processing events could have disrupted some mechanism that links events immediately following each other in memory (Jones, Beaman, & Macken, 1996). These two explanations are not part of any of the

models considered here, but they could probably be added to all of them, provided that more parsimonious accounts fail to handle the data. Third, this effect could be explained by interference between the representation of the word “super” and the memory representations. This interference would be the same for saying “super” once or three times because, in both cases, a single additional representation is encoded into working memory. SOB incorporates one possible implementation of an interference account. We next show that this account is sufficient to explain the pattern of effects, thus obviating the need to add complexity to the models by introducing assumptions about attentional resources or linking mechanisms.

### Quantitative Model Comparison

Although the first experiment differentiated between the predictions of the three models, our analyses so far have been based on a priori predictions from the models and parameter estimates based on fits to different experiments. It is therefore possible that other parameter values, or some minor plausible modifications of model assumptions, might improve the models’ ability to accommodate our data. Accordingly, we now make an effort to enhance the published versions of the three models by fine-tuning their assumptions—without, however, questioning their theoretical core—and estimating the parameters from the data of Experiment 1. We show that even with reasonable amendments, the primacy model continues to predict too much forgetting over increased delays at both encoding and retrieval, and SIMPLE still fails to account for the pattern of effects of delays at encoding. SOB, in contrast, accounts well for the pattern of effects of delay and also provides the best quantitative account of the full set of serial position curves. The MATLAB codes for our simulations using the primacy model and SOB are available on the Web pages provided in the author note.

### The Primacy Model

The main obstacle for the primacy model in handling the present data seems to be the assumption that rehearsal is articulatory and hence blocked by AS. Therefore, the model in its published version must assume that there was no rehearsal during the filled delays in any of the conditions of Experiment 1. Here, we relax that assumption by introducing the probability of rehearsal as a free parameter. At the same time, we provide a realistic instantiation of the rehearsal process. Explication of those modifications requires a more detailed consideration of the model’s architecture.

*Primacy gradient.* In the primacy model, items are represented by localist units. At encoding, the unit representing a list item  $i$  is quickly activated toward an asymptote  $A$ , the rate of activation

Table 3  
Mean Latencies for Encoding and for Retrieving an Item and Reading Distractor Words in Experiment 1

Process measured	Baseline	1 distractor at encoding	3 distractors at encoding	1 distractor at retrieval	3 distractors at retrieval
Encoding	0.88 (0.09)	1.20 (0.19)	2.05 (0.31)	0.94 (0.12)	0.94 (0.14)
Retrieval	0.94 (0.34)	1.40 (0.47)	1.62 (0.83)	1.81 (0.64)	2.67 (0.86)

*Note.* Times are given in seconds. Numbers in parentheses are standard deviations. Encoding times include times for reading the item and times for any distractor words read after it. Retrieval times include times for recalling the item and times for any distractor words read before the recall prompt.

being determined by the input parameter  $I$ . After the input is set to zero, activation decays according to the decay parameter  $D$ :

$$\frac{dx_i}{dt} = -Dx_i + (A_i - x_i) \times I. \quad (2)$$

The asymptotic activation for each item declines linearly over input position according to the equation

$$A_i = s \times \left(1 - \frac{i}{P}\right). \quad (3)$$

Here,  $P$  is the peak parameter representing the maximum possible activation value, and  $s$  is the activation of a start node that is set to  $P$  at the beginning of a list and then decays according to Equation 2. Two parameters were fixed for all our simulations,  $P = 11.5$  (as in Page & Norris, 1998) and  $I = 10$ .

The decline of asymptotic activation  $A_i$  generates a primacy gradient, with each successive item being activated less than the preceding one. At recall, the item with the highest activation is selected for output and then suppressed to zero. Suppression is essential for the model—after the first list item is recalled and suppressed, the second item is the one with the highest activation and therefore will be picked for recall next. Recall errors arise from Gaussian noise added to the activation of each unit at recall; the standard deviation of this noise,  $N$ , is a free parameter. Moreover, the item selected as the one with the highest activation is produced only if its activation surpasses a threshold  $T$ . That threshold has its own Gaussian noise with standard deviation  $M$ .

*Rehearsal.* According to Page and Norris (1998), people rehearse only when they anticipate that the interval between two encoding or retrieval events is long enough to rehearse the complete list currently represented as a primacy gradient. This restriction is necessary because rehearsal is assumed to replace the entire primacy gradient by a new one. An obvious alternative would be to permit partial rehearsal, for example, by boosting the activations of a subset of items within the existing primacy gradient. This implementation of rehearsal, however, would disrupt the primacy gradient, with disastrous consequences for the model's performance.<sup>6</sup>

One potentially critical simplification in the published application of the model to forward serial recall is that rehearsal is implemented as error free. This is at odds with the notion that rehearsal represents an act of retrieval and reencoding because retrieval cannot be guaranteed to be error free. If a retrieval error occurs in a rehearsal attempt, the new primacy gradient generated by rehearsal will reflect that error and will no longer be a veridical copy of the original list. To take this possibility into consideration, we modeled rehearsal during encoding as follows. First, the available time for rehearsal is estimated by taking the average actual times between successive items in each experimental condition and subtracting 250 ms for encoding the item and a further 100 ms as a safety margin. Second, the expected time needed for rehearsing the entire primacy gradient is estimated as  $r \times i$ , where  $r$  is the time for rehearsing an individual item and  $i$  is the number of items in the current primacy gradient. If the expected time needed is smaller than the anticipated available time, rehearsal is initiated. Rehearsal begins by setting  $s$  to  $P$ , thereby initiating a new primacy gradient. Rehearsal, then, consists of an attempt to recall the items in forward order in the same way as during overt recall. At each

step, the item with the highest activation is chosen after adding noise to the current primacy gradient, and this item is regarded as recalled if it surpasses the threshold  $T$ . If an item is recalled, it is activated in the new primacy gradient according to Equations 2 and 3, and it is suppressed in the old primacy gradient. After the entire current list is rehearsed, the new primacy gradient replaces the old one, and the next presented item is added to the new primacy gradient. In this way, rehearsal encodes the list anew, thereby neutralizing all decay that has occurred so far, but rehearsal is not guaranteed to be error free. If items are recalled in the wrong order, the new primacy gradient represents the wrong order, and if an item falls below the recall threshold, it is omitted in the new primacy gradient. If, after one sweep of rehearsal, there is still sufficient time, a further sweep is initiated—this can happen, for instance, after encoding the first or second item, when the current list is still short.

In an analogous manner, rehearsal can operate between two retrieval events if rehearsal of all remaining to-be-recalled items fits into the anticipated available time. Rehearsal at retrieval becomes relevant only for the conditions in which retrieval is delayed; in the conditions without delay at retrieval, there is no point in engaging in covert rehearsal when one could as well recall overtly.

We relaxed the assumption that the distractor activity blocks rehearsal by introducing a free parameter  $R$  that determines the probability with which rehearsal is initiated when the model decides that sufficient time is available. By letting  $R$  vary between 0 and 1, we cover the whole space of possible assumptions about the effectiveness of the distractor activity in blocking rehearsal. Because overt articulation is incompatible with concurrent articulatory rehearsal, any nonzero value of  $R$  is likely to reflect a contribution from attentional refreshing.<sup>7</sup>

<sup>6</sup> One problem with partial rehearsal concerns the role of the activation asymptote  $s$ , which decays along with the list elements. Consider a list ABCD, of which the first two elements are rehearsed. Boosting their activation requires boosting  $s$  as well. As a consequence, the next encoded item E will receive an activation level higher than D because D has not been boosted but the asymptote controlling the activation level of E has been. With E activated higher than D, the order of D and E is misrepresented. Another problem concerns response suppression. After encoding ABCD, rehearsal might start with the first item; this involves retrieving Item A as a first step. Now the model faces a dilemma: Rehearsal of Item A involves giving A an activation boost, but to continue rehearsal of Item B requires applying response suppression to Item A. Without response suppression, the model can only repeat rehearsing Item A.

<sup>7</sup> One might object that the time interval between encoding one letter and the next is not completely filled with AS. After the participant finishes speaking, it will take the experimenter a few hundred milliseconds to press the space bar. During this silent gap, rehearsal probability could be higher than during AS. If that were the case, performance would be better in all conditions than predicted by our simulations, but the relation between conditions would hardly change. Any effect of additional rehearsal during the experimenter's reaction time gap must in any case be small because the reaction time is certainly less than 500 ms, leaving time for just one item to be rehearsed (at the assumed rehearsal rate of 250 ms per item). Thus, the primacy model could use this time gap to rehearse only the very first list item once, immediately after its presentation, at a point in time where it has hardly decayed, so rehearsal would make little difference.

In summary, we introduced two modifications to the primacy model: We instantiated rehearsal by the same error-prone process that governs recall, and we introduced a parameter that governs the probability with which that rehearsal occurs when time permits. In all other respects, we instantiated the primacy model exactly.

*Fitting the primacy model to Experiment 1.* We investigated the performance of the augmented primacy model by a grid search, varying six parameters (lower end, step size, and upper end in brackets): the decay rate  $D$  [0, 0.04, 0.4], the activation noise  $N$  [0.06, 0.06, 0.6], the threshold  $T$  [0.1, 0.1, 0.6], its noise  $M$  [0.1, 0.1, 0.8], the time spent on rehearsing each item  $r$  [0.1 s, 0.1 s, 0.6 s], and the probability of rehearsal  $R$  [0, 0.2, 1]. We ran 500 replications for each combination of parameter values and compared the model predictions with the data of Experiment 1. Model performance was evaluated at each parameter combination by computing the RMSD. The RMSD statistic (defined in Equation 1) reflects the average deviation of the predictions from the data on the original measurement scale (in this case, proportion correct between 0 and 1). The initial grid search was followed by a refined search with smaller steps in the region of the parameter space that gave the best results in the first search.

The best fitting parameter values were  $D = 0.13$ ,  $N = 0.36$ ,  $T = 0.05$ ,  $M = 0.05$ ,  $r = 0.5$ , and  $R = 0.3$ . With these parameters the model achieved a reasonably good fit (RMSD = 0.092, evaluated with 5,000 repetitions). The model predictions are shown as lines in Figure 5, together with the data as points. Similar to its initial predictions (see Figure 1), despite the relaxed constraints on rehearsal and despite optimization of parameter values, the primacy model still overpredicts the difference between short and long delays at encoding, and it underpredicts the difference between baseline (no delays) and short delays at encoding. Likewise, the model overpredicts the amount of fanning between short and long delays at retrieval and underpredicts the negative effect of having a short delay at retrieval relative to the baseline condition.

The estimated rehearsal probability ( $R = 0.3$ ) suggests that the model uses rehearsal only sparingly to accommodate the data. One reason for this is that whenever rehearsal compensates for the effect of decay, it diminishes not only the difference between short and long delays but also the difference between the baseline condition and the short-delay condition. The best fit is obtained with a compromise between underpredicting the baseline condition and overpredicting the short-delay condition. Nonetheless, re-

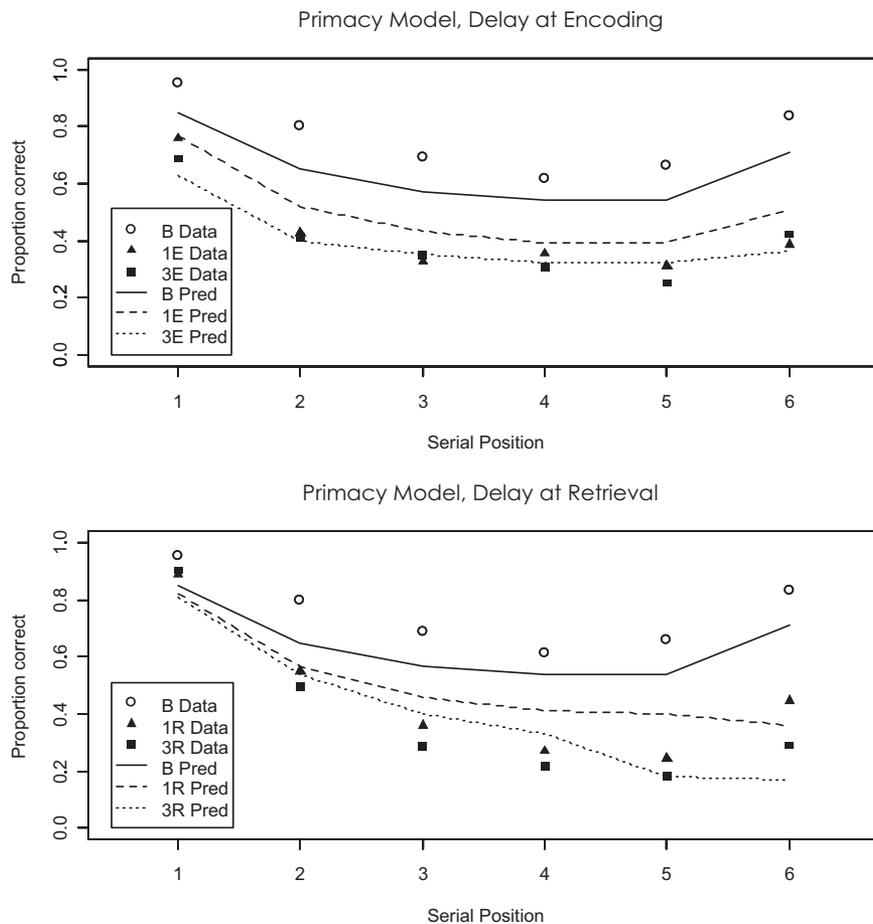


Figure 5. Fit of the primacy model to the data of Experiment 1. Model predictions are represented by lines, observed values by points. B = baseline (no delay); 1E = one distractor word at encoding; 3E = three distractor words at encoding; 1R = one distractor word at retrieval; 3R = three distractor words at retrieval; Pred = prediction.

hearsal does help the primacy model fit the data. The best fit achieved with  $R = 0$  yielded an RMSD = 0.111, and the associated predictions exacerbated the problems apparent in Figure 5, with an even larger predicted difference between conditions with short and with long delay at encoding, albeit not at retrieval. We revisit the effects of rehearsal in the primacy model in a later section of this article.

### SIMPLE

SIMPLE is built on three core assumptions: (a) Items are represented by their position within a potentially multidimensional psychological space, with one of those dimensions necessarily representing time. (b) The similarity between any two items in memory is a declining function of the distance separating them in that psychological space. (c) The probability of recalling a given item is inversely proportional to that item's summed similarity to all other potentially recallable items.

*Encoding in a multidimensional space.* Memory representations are organized along a temporal dimension that reflects the (logarithmically transformed) time elapsed since encoding. The distance ( $D$ ) between any two items  $i$  and  $j$  is thus given by

$$D_{i,j} = |TD_i - TD_j|, \quad (4)$$

where  $D_{i,j}$  is the psychological distance between stimulus  $i$  and stimulus  $j$  and  $TD_i$  is the (logarithmically transformed) temporal distance of stimulus  $i$  from the time of retrieval.

*Similarity-distance metric.* The similarity of any two items in memory is a reducing exponential function of the distance between them in psychological space:

$$\eta_{i,j} = e^{-cD_{i,j}}, \quad (5)$$

where  $\eta_{i,j}$  is the similarity between items  $i$  and  $j$  and  $D_{i,j}$  is the distance between them. The free parameter  $c$  governs the rate at which similarity falls off with distance, with the steepness of the similarity gradient increasing with the magnitude of  $c$ . In consequence, items in close psychological proximity have a similarity approaching unity, whereas items that are more psychologically distant have a similarity that, in the extreme, tends toward zero.

*Similarity determines recall.* Given its location within the representational space as cue, the probability of recalling an item  $P(R_i)$  is inversely proportional to the summed similarity of that item to every other potentially recallable response:

$$P(R_i) = \frac{1}{\sum_{k=1}^n (\eta_{i,k})}, \quad (6)$$

where  $n$  is the number of items that could be recalled (here, the number of list items).

Omissions are assumed to arise from thresholding of low retrieval probabilities (as calculated by the preceding equation). Brown et al. (2007) implemented this threshold mechanism using a sigmoid function such that if  $P$  is the recall probability as just defined, it is transformed to become the final predicted recall probability  $P^*$  as follows:

$$P^* = \frac{1}{1 + e^{-s(P-t)}}, \quad (7)$$

where free parameters  $t$  and  $s$  refer to the absolute value and to the noisiness of the threshold, respectively (see Brown et al., 2007, for a complete account).

In addition, output interference was modeled by assuming that the memory representations of to-be-recalled items become progressively less distinctive as recall proceeds. In the simulations, this is implemented by the parameter  $o$ , which reduces the value of  $c$  for the  $k$ th item recalled by multiplying  $c$  by  $o^{k-1}$ . To adapt SIMPLE to the present experimental paradigm, we assumed that distractors at retrieval could introduce additional noise. To give SIMPLE the opportunity to represent this possibility and to thereby achieve a better fit, we introduced a separate output interference parameter  $o_r$  to replace  $o$  in the two conditions with delays at retrieval.

*Adapting discriminability to the time scale.* The predictions we derived from SIMPLE for Experiment 1 (see Figure 1) used a single value of  $c$  for all conditions. These predictions are severely at odds with the data because SIMPLE predicts that long delays at encoding result in performance better than baseline. The assumption of a constant value of  $c$ , however, is potentially questionable because the baseline condition and the conditions with delays at encoding differ in the range of time covered by a list. Brown et al. (2007) argued that  $c$  may vary between experiments in a systematic way reflecting the time range of relevant events. Moreover, in the context of applying SIMPLE to data from an absolute identification paradigm, Neath and Brown (2006) proposed a principled way of linking  $c$  to the range of values on a dimension—in the present case, time—on which items must be discriminated. The idea is to adapt  $c$  such that discriminability of items is scale invariant, implying that a proportional increase or decrease of all temporal distances between stimuli does not change performance.

We therefore used the procedure of Neath and Brown (2006) to compute a separate value of  $c$  for each condition, as follows: We defined the temporal range of each condition as the difference between the shortest and the longest  $TD_i$  values, measuring  $TD_i$  from presentation of item  $i$  until onset of the first recall screen (we disregarded the retention interval during recall of other items because the value of  $c$  must be set at the beginning of recall at the latest). For the condition with the shortest time range—in the present case, this was the baseline—a value for  $c$  was freely estimated. For the remaining conditions, each value of  $c$  was a function of the ratio of the time ranges between that condition and the baseline. Specifically, the freely estimated baseline value of  $c$  was divided by the ratio of time ranges, thus ensuring scale invariance of performance by reducing  $c$  in synchrony with an expanding time scale (see Neath & Brown, 2006, for further detail).

*Fitting SIMPLE to Experiment 1.* SIMPLE was fit to the data of Experiment 1 using the standard simplex algorithm as implemented in MATLAB. The best fitting parameter values were  $c_{\text{baseline}} = 2.08$ ,  $t = 0.26$ ,  $s = 17.63$ ,  $o = 1.00$ , and  $o_r = 0.997$ ; they yielded an RMSD = 0.141. Figure 6 shows the predictions. It is clear from the figure that even after its twofold adaptation to the present paradigm (by scaling  $c$  and by allowing different extents of output interference) and optimization of its parameter values, SIMPLE drastically mispredicts the data. Linking  $c$  to the list's

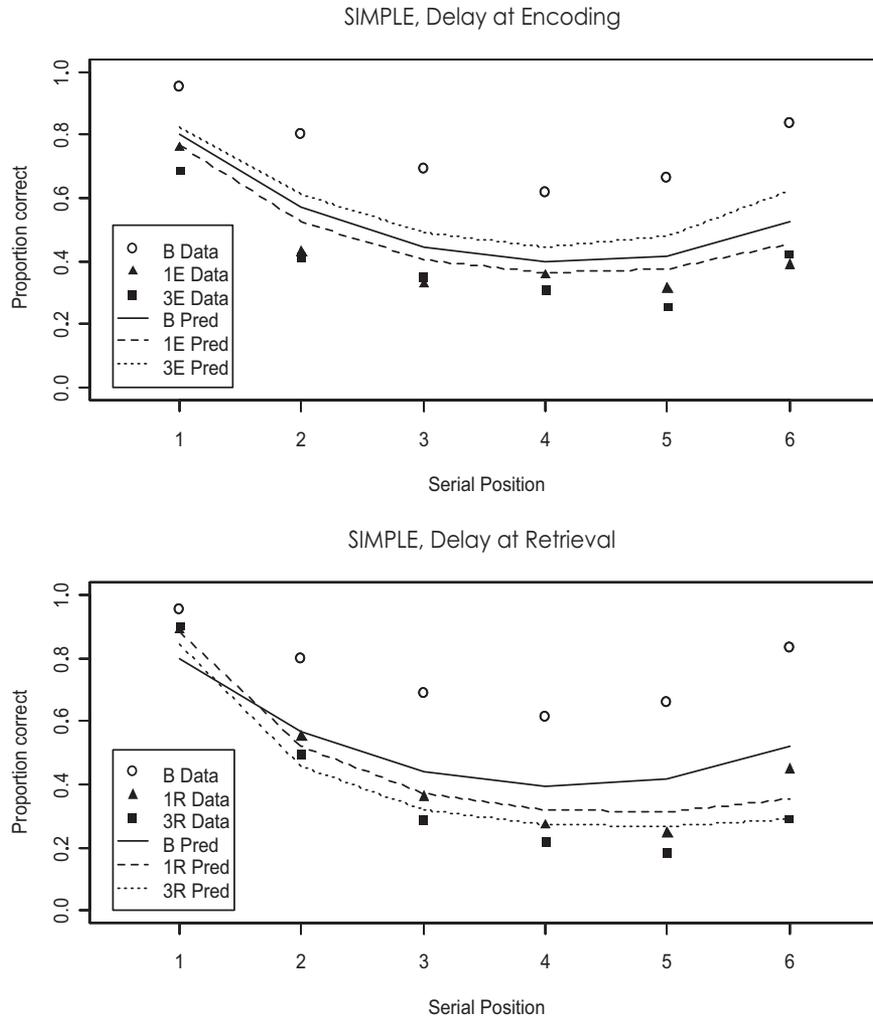


Figure 6. Fit of SIMPLE to the data of Experiment 1. Model predictions are represented by lines, observed values by points. B = baseline (no delay); 1E = one distractor word at encoding; 3E = three distractor words at encoding; 1R = one distractor word at retrieval; 3R = three distractor words at retrieval; Pred = prediction.

time range, as suggested by Neath and Brown (2006), helped avoid the most dramatic misprediction for the condition with long delays at encoding. Scale invariance over time, however, can only neutralize the effect of delays at encoding; it does not enable SIMPLE to predict the large drop of accuracy that was caused by the distractors at encoding, relative to baseline. Moreover, despite having an extra free parameter,  $\sigma_r$ , to account for interference from distractors at retrieval, SIMPLE lost its ability to capture the pattern obtained for delays at retrieval that was apparent in the initial predictions (see Figure 1). Obviously, optimizing parameters for fitting all five conditions simultaneously led to a solution that distributed the misfit across all conditions.

When Lewandowsky et al. (2004) applied SIMPLE to their experiments, they introduced ordinal list position as a second dimension of discrimination besides time. The relative weight of the positional and the temporal dimensions was determined by a free parameter  $wt$ . Lewandowsky et al. found that SIMPLE could fit their data—which exhibited virtually no difference between one

and three distractors at retrieval—with a value of  $wt$  close to zero, implying complete reliance on the positional dimension and no role for time. We tested whether SIMPLE, extended by the positional dimension, was able to fit the results of the present Experiment 1 as well. The best fitting solution was again obtained with  $wt = 0$ , and it achieved an acceptable fit (RMSD = 0.093). Thus, a distinctiveness model based on positional distinctiveness, with no role for time, can account for the present data better than if temporal distinctiveness is present. It should be added, however, that this purely position-based version of SIMPLE can predict differences between the baseline condition and the delay-at-encoding conditions only through the adaptation of  $c$  to the time range. Adapting  $c$  to the time range while not using time as a dimension of discrimination is at odds with the idea that  $c$  is adapted to achieve scale invariance on the relevant dimension of discrimination. Therefore, a purely positional version of SIMPLE is plausible only with a constant value for  $c$ . With  $c$  held constant across conditions, however, the positional SIMPLE inevitably

predicts the serial position curves of baseline and the two delay-at-encoding conditions to lie on top of each other, in stark deviation from the data. Thus, despite our efforts to adapt SIMPLE to the present paradigm, we found this model unable to capture the data of Experiment 1.

## SOB

SOB rests on three principal architectural commitments: (a) Energy-gated encoding as described at the outset, (b) implementation of response suppression by the same process that governs encoding, and (c) dynamic iterative deblurring of partial images retrieved from memory. Following other recent applications of SOB (Farrell, 2006; Lewandowsky & Farrell, in press), we retained the first two commitments but relaxed the third one for reasons of computational constraints (i.e., available equipment could not readily support the computing demands of dynamic iterative deblurring).

*Architecture.* The network consists of two layers: an input layer ( $N = 16$ ) used to represent positional markers and an output layer ( $N = 150$ ) representing list items, where  $N$  refers to the number of units in each layer. The two layers are fully interconnected by a weight matrix  $\mathbf{W}$ .

*Positional markers.* SOB incorporates a positional marker for each list position, as is frequently assumed by models of memory (e.g., Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999). The similarity between any two positional markers is an exponential function of their absolute separation in list positions, that is,

$$\cos(\mathbf{p}_i, \mathbf{p}_j) = t_c^{(|i-j|)}, \quad (8)$$

where  $i$  and  $j$  are the positions of the  $i$ th and  $j$ th items,  $\mathbf{p}_i$  and  $\mathbf{p}_j$  are distributed vectors representing positional markers at those positions, and  $t_c$  is a fixed parameter determining the degree of overlap of successive positional markers ( $t_c = 0.50$  throughout, as in Farrell, 2006). By implication, SOB represents only positional information and is unaware of the temporal separation between items at encoding. This is in line with much recent research that has shown temporal isolation to be irrelevant in serial recall (e.g., Lewandowsky et al., 2006; Nimmo & Lewandowsky, 2005).

*Encoding.* At list presentation, incoming items are associated with successive positional markers using standard Hebbian learning (see, e.g., J. A. Anderson, 1995):

$$\Delta \mathbf{W}_i = \eta_e(i) \mathbf{v}_i \mathbf{p}_i^T, \quad (9)$$

where  $\mathbf{W}$  is the weight matrix connecting markers to items,  $\mathbf{v}$  is the vector representing the  $i$ th presented item, and  $\mathbf{p}$  is the positional marker for the  $i$ th serial position. The learning rate  $\eta_e$  for the  $i$ th association,  $\eta_e(i)$ , was calculated using the energy between the to-be-learned association and the information captured by  $\mathbf{W}$  up to that point:

$$\eta_e(i) = \begin{cases} 1, & i = 1 \\ -\phi_e / E_i, & i > 1 \end{cases}, \quad (10)$$

where  $\phi_e$  is a free parameter and  $E_i$ , the energy of the  $i$ th association, is given by

$$E_i = -\mathbf{v}_i^T \mathbf{W}_{i-1} \mathbf{p}_i \quad (11)$$

(see Farrell & Lewandowsky, 2002). The use of energy to compute the weighting of incoming information is a core principle of SOB that turns out to be responsible for its ability to accommodate the present data.

Energy is a measure of the consistency between novel information and what has already been learned or, in psychological terms, of the novelty of an item in the context of a particular positional marker. By reducing the encoding strength of items that are not entirely novel, energy gating limits redundancy and, by implication, enhances memory for unique information (see, e.g., Lewandowsky & Farrell, in press). In the present context, energy-gated encoding also limits the adverse effect of additional repetitions of a distractor.

*Item and distractor representation.* The use of energy-gated encoding mandates an accurate representation of the similarity structure among input items. Accordingly, all study items in the present simulations were derived from a common prototype (a binary vector that was randomly sampled for each replication) by retaining each feature with probability  $p_c$  and resampling it (from a symmetric binary distribution) with probability  $1 - p_c$ . As for the generation of initial predictions, the value of  $p_c$  was fixed to .70, thus capturing the empirically determined similarity between letters (Lewandowsky & Farrell, in press). For each simulation replication, a single distractor item (i.e., “super”) was derived from the prototype. To reflect the obvious fact that the distractor differed more from the ensemble of list items than they differed among each other, the distractor was created using a value of  $s_c$  fixed to half of the value of  $p_c$  (i.e.,  $s_c = 0.35$ ). When modeling a condition that involved delay, the association between the nearest preceding context marker and the distractor item was added to  $\mathbf{W}$  either once or thrice, as determined by condition, using the same mechanics as for study items. Thus, when modeling delay at encoding, each “super” was associated to the context marker of the preceding item, either once or three times. When modeling delay at retrieval, the marker for the just-recalled item was used (in the case of the first distractor, which preceded recall of the first item, the marker of the last input item was used because it represented the nearest preceding context).

*Retrieval.* Retrieval in the network consists of stepping through the positions and using the positional markers to cue for their associated items. For position  $i$ , retrieval is cued by placing positional marker  $\mathbf{p}_i$  across the input layer and updating the item unit activations:

$$\mathbf{v}_i' = \mathbf{W}_i \mathbf{p}_i, \quad (12)$$

where the resultant vector,  $\mathbf{v}_i'$ , is a noisy version of  $\mathbf{v}_i$ , containing a blend of  $\mathbf{v}_i$  and the other items on the list owing to the overlap among positional markers. The probability of recalling an item  $\mathbf{v}_j$  was then determined by matching  $\mathbf{v}_i'$  to the experimental vocabulary. Formally, the probability of recalling  $\mathbf{v}_v$  from the experimental vocabulary is

$$P(\mathbf{v}_v) = \frac{s(\mathbf{v}_i', \mathbf{v}_v)}{\sum_{k=1}^n s(\mathbf{v}_i', \mathbf{v}_k)} \quad (13)$$

where  $n$  is the number of items in the experimental vocabulary and the function  $s$ , the similarity between the noisy vector and an item vector, is calculated as

$$s(\mathbf{v}_i', \mathbf{v}_k) = \exp[-cD(\mathbf{v}_i', \mathbf{v}_k)^2]. \quad (14)$$

The Euclidian distance measure  $D$  is weighted by the free parameter  $c$  and—for computational reasons—is normalized by subtracting the minimum distance, across all the  $n$  items, from the distance for each item. A discrete response is then determined by random sampling according to the preceding equation.

The recalled item is then suppressed by adjusting the weights between the positional marker layer and the item layer, according to

$$\Delta \mathbf{W}_j = \eta_s(j) \mathbf{v}_{o,j} \mathbf{p}_j^T, \quad (15)$$

where  $j$  is the output position and  $\mathbf{v}_{o,j}$  is the item recalled at the  $j$ th position.

To ensure that the extent of response suppression approximately matches that of learning, the learning rate for suppression,  $\eta_s(j)$ , is also determined from the energy of the recalled item  $j$  according to

$$E_j = -\mathbf{v}_{o,j}^T \mathbf{W}_{j-1} \mathbf{p}_j. \quad (16)$$

The response suppression rate is given by

$$\eta_s(j) = \frac{-E_j}{\phi_s E_1}, \quad (17)$$

where  $\phi_s$  is a free parameter and  $E_1$  is the energy of the item recalled at the first output position (see Farrell & Lewandowsky, 2002); this term reduces to  $-\phi_s^{-1}$  for the first output position (where  $j = 1$ ). Suppression thus uses the same Hebbian learning principles that are used at encoding, except that the negative learning rate (note the minus sign in the numerator of the preceding equation) implements *anti-learning* (J. A. Anderson, 1991) and hence attenuates an item's representation.

Reflecting a fairly general consensus in the short-term memory literature (e.g., Brown et al., 2000; Cowan, Saults, Elliott, & Moreno, 2002; Fitzgerald & Broadbent, 1985; Oberauer, 2003b) and paralleling the earlier implementation of SIMPLE, retrieval of an item is accompanied by output interference. Output interference is modeled by adding Gaussian noise with standard deviation  $N_o$  (a free parameter) to each weight in  $\mathbf{W}$  after each retrieval.

*Fitting SOB to Experiment 1.* SOB was fit to the results of Experiment 1 with two changes from the off-the-shelf version that generated the predictions in Figure 1: First, the representations of the distractors drifted across input and output positions (but not within immediate repetitions of the distractors at a given serial position), reflecting the assumption that although people nominally uttered the same word throughout the experiment, the internal representation of those words likely changed somewhat across serial positions. Accordingly, the item–distractor similarity  $s_c$  was used to derive a prototype for the distractors (rather than the single copy that was used for the predictions in Figure 1). The distractor at each serial position was then derived from that prototype by retaining features with probability  $J_c$  and resampling them at random with probability  $1 - J_c$ , where  $J_c$  was another freely estimated parameter.

A second change was made to accommodate recent findings from the working memory literature that strongly point toward the existence of an attentional refreshing mechanism (Barrouillet et al., 2004; Hudjetz & Oberauer, 2007; Raye et al., 2007). There is no fully fleshed rehearsal or refreshing mechanism yet for SOB,

and implementing one would require substantial changes and additional free parameters, contrary to our purpose to keep the models simple and close to their published versions. We therefore implemented a very simple form of refreshing: We assume that the last item encoded remains in the focus of attention (McElree & Doshier, 1989; Oberauer, 2003a). Speaking repeated distractor words aloud does not displace the last item from the focus because this process does not make demands on the attentional bottleneck. This assumption is supported by the finding of Garavan (1998) that overt articulatory rehearsal does not change the content of the focus of attention in working memory. Therefore, the last item can be reencoded after the encoding of a distractor word into memory. After each distractor during encoding, we encoded the preceding item again, computing its energy anew every time (Equations 9–11). The reencoding adds only a small amount of strength to each item because an already encoded item yields a low absolute value of energy; this value is not zero, however, because the intervening distractor adds noise to the weight matrix. At retrieval, refreshing does not operate because the item that resides in the focus of attention during the distractor activity is the item just recalled (and suppressed), and it would be counterproductive to encode it again.

This final application of SOB therefore used five freely estimated parameters ( $\phi_e$ ,  $\phi_s$ ,  $N_o$ ,  $c$ , and  $J_c$ ) in addition to the fixed parameters ( $t_c = 0.50$ ,  $p_c = 0.70$ , and  $s_c = 0.35$  for all applications in this article). The best fit to the data of Experiment 1, found through grid search, was obtained with  $\phi_e = 300$ ,  $\phi_s = 0.61$ ,  $N_o = 1.38$ ,  $c = 0.54$ , and  $J_c = 0.31$ . The model fit the data moderately well and slightly better than the primacy model, with RMSD = 0.078. Figure 7 shows the predictions together with the data. The model correctly reproduced the large decrease in performance with a single distractor at either encoding or retrieval, relative to baseline, together with virtually no further decrease with three distractors. With its best fitting parameters, the adapted version of SOB stopped showing even the small differences between one and three distractors that were originally predicted by the model (see Figure 1); the TEE and TLS indices generated by the adapted model version are virtually zero.

There are two discernible sources of misfit. One is that there is no fanning with delays at retrieval in the predictions, whereas there is some degree of fanning in the data. As the initial predictions derived from SOB show, the model can produce a small amount of fanning with other parameter settings. We analyzed the conditions under which fanning is predicted through a series of simulations, each setting one parameter to its value used for the initial predictions in Figure 1, while holding all others at the best fitting values. None of these simulations produced fanning, but when we set both  $N_o$  and  $J_c$  to the values used to produce the initial predictions ( $N_o = 1.28$ ,  $J_c = 1$ ), a small degree of fanning emerged, albeit at the expense of creating a very bad fit overall (due to underestimating the overall effect of distractors). This analysis confirms the conclusion from the predicted distribution of TLS values (see Figure 3): Limited fanning (i.e., negative TLS values) can be predicted by SOB under certain circumstances, but this is not a strong prediction that arises from core assumptions of the model.

A second source of misfit is the shape of the serial position functions in the conditions with delay at retrieval. SOB underpredicts the depth of the trough at Serial Positions 3–5. We defer

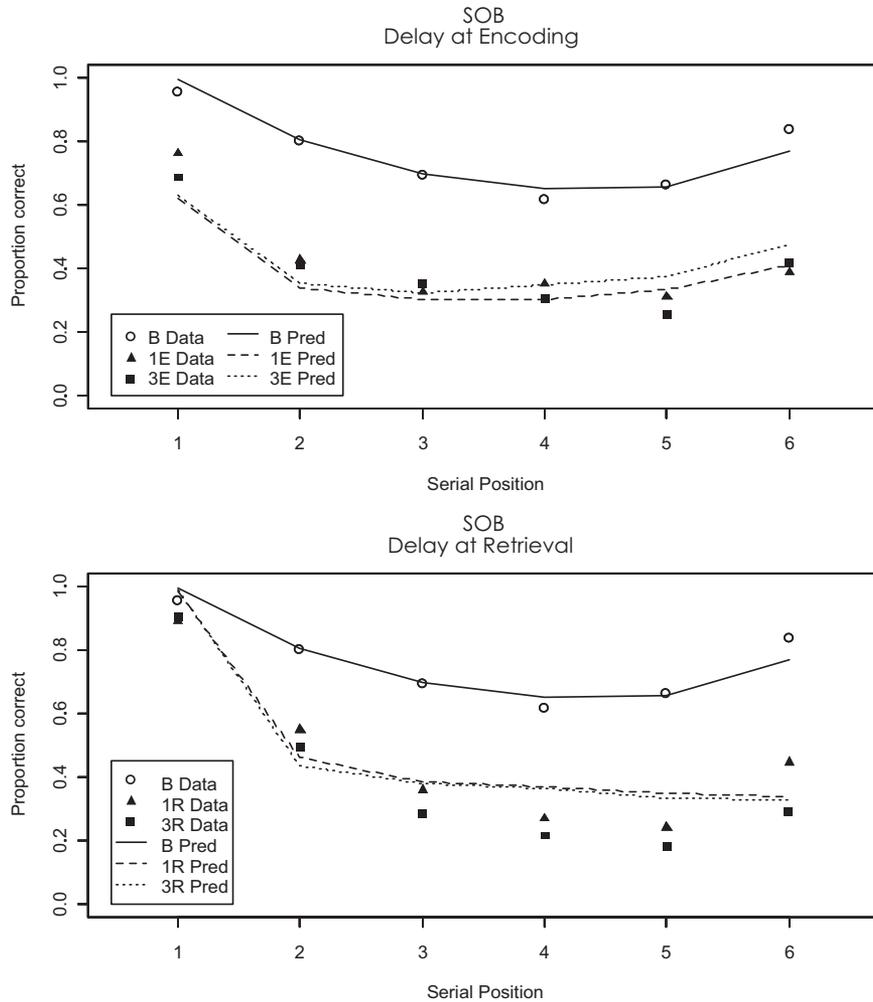


Figure 7. Fit of SOB to the data of Experiment 1. Model predictions are represented by lines, observed values by points. B = baseline (no delay); 1E = one distractor word at encoding; 3E = three distractor words at encoding; 1R = one distractor word at retrieval; 3R = three distractor words at retrieval; Pred = prediction.

discussion of this observation until after presenting the model fits to Experiment 2.

### Summary of Modeling

The modeling revealed that a temporal distinctiveness theory, SIMPLE, cannot handle the effects of time at encoding and retrieval that were observed in our first experiment. SIMPLE failed to handle the data even when its principal parameter was scaled separately for each condition and when the model was augmented by a positional dimension, which on previous occasions (Lewandowsky et al., 2004, 2006) has enabled the theory to overcome otherwise challenging results.

The modeling also confirmed that the decay-based primacy model, as well as the interference-based SOB, provided acceptable rival accounts of the data. SOB fit the data slightly better than the primacy model, and although the numeric difference in fit was small, the way in which the two theories deviated from the data was informative of their respective strengths and weaknesses.

Despite being able to rehearse concurrently with AS, the primacy model predicts larger differences between short and long delays than observed. SOB correctly reproduced the relative difficulties of baseline, short-delay, and long-delay conditions, but this came at the cost of somewhat distorted serial position curves.

### Experiment 2

The first experiment used AS to prevent rehearsal. AS has been the accepted technique for preventing rehearsal in the working memory community, based on the assumption that rehearsal of verbal material is articulatory in nature, as postulated by the phonological loop model and the primacy model. In recent years, an attentional refreshing mechanism has been theoretically postulated by Barrouillet et al. (2004), and there is mounting empirical evidence in support of its existence (Hudjetz & Oberauer, 2007; Raye et al., 2002, Raye et al., 2007). Arguably, the attentional mechanism might be engaged very little or not at all by repetitive articulation of the same word, so that in Experiment 1 and other

related precedents (e.g., Lewandowsky et al., 2004), the memory contents might have been refreshed concurrently with AS. Indeed, our account of Experiment 1 within SOB assumed that this is exactly what people were doing.

Experiment 2 therefore tested the effect of delays during which articulatory rehearsal and attentional refreshing were both blocked. We added a choice reaction task (CRT from here on) to the articulation of a distractor word because response selection in a CRT has been identified as a process that occupies the central attentional bottleneck (Pashler, 1994), which is the presumed locus of refreshing. Experiment 2 crossed two independent variables, the number of “supers” to be spoken in the delay period (one vs. four) and whether or not the distractors were accompanied by a choice trial. The inclusion of choice trials constrains the predictions of the primacy model and of SOB.

Considering the augmented primacy model first, for the condition with “supers” only (AS-only from here on), the model predicts only a moderate difference between short and long delays because refreshing can partially compensate decay (recall that  $R > 0$  for the fits to Experiment 1). For the condition combining each “super” with a CRT (AS+CRT), the augmented primacy model must assume that both rehearsal and refreshing are prevented (i.e.,  $R = 0$ ), and therefore this condition should finally reveal the full effect of decay. Thus, if AS blocks only articulatory rehearsal while leaving attentional refreshing uncontrolled, we should find a substantially larger amount of forgetting in the AS+CRT conditions than in the AS-only conditions. This translates into possibly small TEE and TLS scores in the AS-only condition but larger and substantial negative indices in the AS+CRT conditions.

Turning to the adapted version of SOB, the model predicts that adding a CRT to the distractor activity prevents reencoding of items in the delay periods. This assumption is based on the finding that a CRT competes with consolidation of information in short-term memory, probably because both place demands on an attentional bottleneck (Jolicoeur & Dell’Acqua, 1998). Therefore, we assume that, different from our AS condition, the added CRT prevents reencoding in SOB during the delay periods. This leads to the prediction that memory is better in the AS-only conditions than in the AS+CRT conditions when the distractors are inserted during encoding. During retrieval, in contrast, the reencoding mechanism does not operate anyway, so there should be little difference between AS-only and AS+CRT. Finally, the critical contrast with the predictions made by the primacy model concerns the effect of the duration of the delays in all conditions. Irrespective of attentional reencoding, SOB assumes no time-based decay and accounts for the effects of additional distractors entirely on the basis of energy-gated interference. It follows that as in Experiment 1, SOB predicts only a very small difference between short and long delays, and the TEE and TLS values should differ little from zero.

## Method

*Participants.* Twenty psychology undergraduates of the University of Western Australia (Crawley, Western Australia) participated for course credit or remuneration at the rate of Australian \$10/hour (approximately U.S. \$9.11). The experiment was conducted in three 1-hr sessions.

*Design.* Participants were given lists of five consonants for immediate serial recall. The lists were assembled by drawing items at random without replacement from a pool of 19 consonants (excluding Q and Y). The design crossed the placement of distractors (at encoding or at retrieval) with the kind of distractor events (AS-only or AS+CRT) and the number of distractor events in each delay period between items (one or four). In addition, there was a baseline condition without distractors. Across the three sessions, there were 21 test trials in each of the nine conditions; trials of all conditions were mixed in a random order.

*Materials and procedure.* The experiment was programmed using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Each trial started with the presentation of a cue indicating the condition (e.g., “4 × super + choice recall” or “1 × super study,” and so on) for 3.5 s, followed by a fixation cross for 1.5 s. After offset of the fixation cross, items were presented sequentially in the center of the screen at a rate of 0.85 s per item (750 ms on, 100 ms off) in black on a white background. In contrast to Experiment 1, we instructed participants to read the items silently. In the conditions with distractors at encoding, each item was followed by one or four distractor events. In the AS-only conditions, each distractor event involved saying “super” once; in the AS+CRT conditions, each distractor event involved saying “super” once and performing a single CRT simultaneously.

The recall phase followed immediately after the last study item or the last distractor in the encoding phase. It started with the presentation of a question mark to prompt recall of the first item (in conditions without delays at retrieval) or else with the first distractor event. When delays were introduced at retrieval, they preceded each item. Once a question mark was displayed after each set of distractor events, participants spoke the to-be-recalled item and pressed the up arrow key to proceed to the next screen.

In the AS-only conditions, the stimulus for each distractor event consisted of an asterisk displayed centrally in the upper third of the screen. In response to this stimulus, participants said “super” aloud and pressed the up arrow key. The next stimulus, presented after a 100-ms delay, could be another asterisk (in the condition with four distractor events), the next consonant (during encoding), or a question mark (during recall). The encoding phase therefore consisted of a sequence of silent letter reading alternating with utterances of “super,” accompanied by the same keypress every time. The recall sequence consisted of a sequence of vocal utterances (either saying “super” or recalling a consonant), each of which was accompanied by pressing the same key; the keypressing task thus involved no response selection.

In the AS+CRT conditions, the stimulus for each distractor event was & or % displayed centrally. In response to this stimulus, participants had to say “super” aloud and press the left or the right arrow key, depending on the symbol displayed (left for & and right for %). Symbols were selected at random with equal probability. Following the CRT response, the trial continued exactly as in the AS-only conditions. The encoding and recall sequences in the AS+CRT condition thus required stimulus-contingent selection between altogether three keys, the two response keys of the CRT and the up arrow key for continuing after encoding or recalling an item. These selection demands engaged the attentional bottleneck during each event. To ensure that successive asterisks and symbols were perceived as different events, the exact location of each

distractor stimulus was shifted vertically by a random amount (about 1 cm).

Each session consisted of 81 trials. Of these, the first 27, 18, and 9 were practice trials in Sessions 1, 2, and 3, respectively. Data from the practice trials were not analyzed. Trials were separated by a blank interval of 3.5 s. A short break was provided after every 20 trials. At the beginning of the first session, participants completed 100 practice trials on the CRT alone to familiarize themselves with the stimulus–response mapping.

## Results

The serial position curves for the nine conditions are shown in Figure 8. The data were analyzed through two ANOVAs—one for distractors at encoding and one for distractors at retrieval—with serial position and delay condition (baseline, 1AS-only, 4AS-only, 1AS+CRT, or 4AS+CRT) as independent variables. Serial position was coded as linear contrast, and delay condition was decomposed into four orthogonal contrasts: The distraction contrast compares the baseline condition with the mean of all four distraction conditions. The CRT contrast compares the AS-only condition with the AS+CRT condition, averaging across the number of distractors. The delay-with-AS contrast compares one versus four distractors in the AS-only condition, and the delay-with-CRT

contrast compares one and four distractors in the AS+CRT condition. The results of these ANOVAs are summarized in Table 4.

The large effect of the distraction contrast at both encoding and retrieval shows that, once again, introducing delays filled with a distractor task impaired memory substantially, even when the distractor activity consisted merely of speaking a constant word aloud. At encoding, the CRT contrast showed that the distracting effect of AS+CRT was markedly stronger than that of AS-only. Looking at the CRT contrast at retrieval, AS+CRT also disrupted memory more than AS-only, but the difference was relatively small compared with the overall distraction contrast, which considered the effects of both types of distractors jointly, relative to baseline.

The number of distractors (one vs. four) had hardly any effect at encoding. The just-significant delay-with-AS contrast reflects the fact that, with AS-only, there was even a trend toward better performance with more distractors; with AS+CRT, there was a trend in the opposite direction, reflected by the delay-with-CRT contrast that just failed the significance criterion. At retrieval, the number of distractor events had no effect at all in the AS-only condition, as reflected in the delay-with-AS contrast. This result differs from Experiment 1, where we found a moderate degree of fanning between these two conditions. Here, we obtained fanning only in the AS+CRT condition; it is reflected in the significant interaction of the delay-with-CRT contrast with the linear trend of serial position.

Table 5 summarizes the mean times participants took for the distractor activity between items at encoding and at retrieval. The times at encoding were measured from the onset of the asterisk in the first distractor event to the last response before the next item was displayed. The times at retrieval were measured from the onset of the first asterisk until the keypress given in response to the question mark that prompted recall of the next item. These data show that speaking a single “super” at encoding took longer in Experiment 2 than in Experiment 1 (780 ms for a single “super” and about 500 ms for every following “super” in a series of four, compared with 400 ms in the previous experiment), probably because, in Experiment 2, participants had to press a key to continue after each “super.” The duration of a complex distractor event, consisting of an utterance of “super” and a choice reaction, was 1,100 ms for a single distractor event and about 700 ms for every further distractor event in a series of four. When the retrieval times in the conditions with delays at encoding, which reflect only the time to recall an item, are subtracted from the retrieval times in the corresponding conditions with delays at retrieval, roughly the same estimates are obtained for distractor events at retrieval as for distractor events at encoding.

## Discussion

This experiment replicated and extended the main findings of Experiment 1. With AS as the only means to prevent rehearsal, there was again a substantial loss in memory accuracy compared with the baseline condition, but the number of distractors and therewith the extent of temporal delay had no further detrimental effect on memory. The modest amount of fanning that resulted from longer delays at retrieval in Experiment 1 ( $TLS = -.004$ ; see Table 1) was absent in this experiment ( $TLS = .0026$ ).

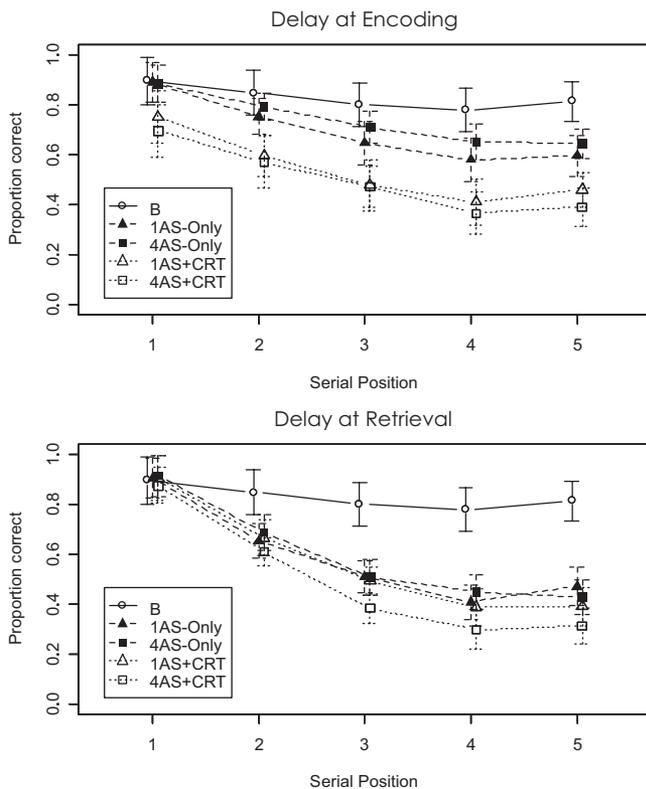


Figure 8. Serial position curves observed in Experiment 2. Top panel: Baseline and conditions with delay at encoding. Bottom panel: Baseline and conditions with delay at retrieval. Error bars represent 95% confidence intervals for within-subject comparisons. B = baseline; AS = articulatory suppression; CRT = choice reaction task.

Table 4  
Summary of Analysis of Variance Results for Experiment 2

Contrast	Delays at encoding			Delays at retrieval		
	<i>F</i>	<i>p</i>	Partial $\eta^2$	<i>F</i>	<i>p</i>	Partial $\eta^2$
Serial Position	80.2	<.001	.81	160.9	<.001	.89
Distractor	61.4	<.001	.76	151.9	<.001	.55
CRT	72.2	<.001	.79	23.6	<.001	.55
Delay With AS	4.4	.05	.19	0.2	.64	.01
Delay With CRT	3.6	.07	.16	14.1	.001	.43
Serial Position $\times$ Distractor	62.3	<.001	.77	173.6	<.001	.90
Serial Position $\times$ CRT	1.1	.30	.06	15.8	.001	.46
Serial Position $\times$ Delay With AS	2.3	.15	.11	1.7	.21	.08
Serial Position $\times$ Delay With CRT	0.2	.69	.01	1.4	.26	.07

Note. Serial position: linear contrast of serial position. Distraction contrast: baseline versus all distractor conditions; CRT contrast: AS-only versus AS+CRT; delay-with-AS contrast: one versus four distractors in AS-only condition; delay-with-CRT contrast: one versus four distractors in AS+CRT condition. Degrees of freedom were 1, 19. AS = articulatory suppression; CRT = choice reaction task.

Indeed, with AS-only at encoding, we found even a slight advantage for saying “super” four times over saying it only once after each new item. Although only just significant, this beneficial effect of longer delays at encoding should not be dismissed as spurious—we found this effect, which translates into a positive TEE index, in three experiments (see Table 1).

The new conditions in which attentional refreshing was blocked by an added CRT impaired memory even more than AS alone, particularly at encoding. Crucially, however, even with both hypothetical revival mechanisms blocked, long delays at encoding did not engender more forgetting than short delays. Only in the AS+CRT condition at retrieval did we find some evidence for a detrimental effect of longer delays, manifested in a moderate degree of fanning of the serial position curves, and a negative TLS. Table 1 shows that the present TLS (−.0112) had the largest absolute value of nine estimates, which cluster closely together in a region ever so slightly below zero. To place the results into context, this largest empirical TLS is equivalent to a one-percentage-point decline in performance with every additional second of elapsed time; the average value across all nine estimates represents less than half that amount of forgetting.

Modeling of Experiment 2

We applied the same models that we fit to Experiment 1 to the data from Experiment 2, making only a few adjustments to implement the effects of the CRT during the filled delays.

The Primacy Model

This application relied on the same six free parameters that were used to model Experiment 1, with the constraint that rehearsal was precluded in the AS+CRT conditions. Rehearsal still occurred with probability *R* in the AS-only conditions.

The modified primacy model fit the data poorly (RMSD = 0.139). The best fitting parameter values were decay rate *D* = 0.11, rehearsal duration *r* = 0.6 s, noise *N* = 0.28, threshold *T* = 0.05, threshold noise *M* = 0.10, and AS-only rehearsal probability *R* = 0.3. The predictions are shown as lines, together with the data as points, in Figure 9. The model’s failure can clearly be traced to one main source: The model predicts much larger effects of temporal delays than were observed, both at encoding and at retrieval.

Could the primacy model explain the data if we allowed some rehearsal even in the AS+CRT conditions? To explore this possibility, we introduced a further parameter, *R*<sub>CRT</sub>, for the probability of rehearsal in the AS+CRT condition. The best fit found through a grid search in the seven-parameter space was obtained with *D* = 0.11, *r* = 0.6, *N* = 0.28, *T* = 0.05, *M* = 0.10, *R* = 0.3, and *R*<sub>CRT</sub> = 0.2. With RMSD = 0.114, this model version reached a closer fit than the six-parameter version, but it still overpredicted the effects of long versus short delays at both encoding and retrieval.

SIMPLE

We applied the same augmented version of SIMPLE to Experiment 2 as to Experiment 1; since SIMPLE has no role for

Table 5  
Mean Latencies for Distractor Activity and Item Recall in Experiment 2

Experimental condition and process measured	Baseline	1 AS-only	4 AS-only	1 AS+CRT	4 AS+CRT
Distractors at encoding					
Encoding times		0.78 (0.24)	2.19 (0.63)	1.10 (0.27)	3.28 (0.61)
Retrieval times	0.69 (0.26)	0.91 (0.42)	0.91 (0.40)	1.04 (0.47)	1.11 (0.45)
Distractors at retrieval					
Retrieval times	0.69 (0.26)	1.75 (0.49)	3.13 (0.78)	2.17 (0.43)	4.44 (0.75)

Note. Times are given in seconds. Numbers in parentheses are standard deviations. Encoding times reflect times for distractor activities only; retrieval times include the time for recalling the item. AS = articulatory suppression; CRT = choice reaction task.

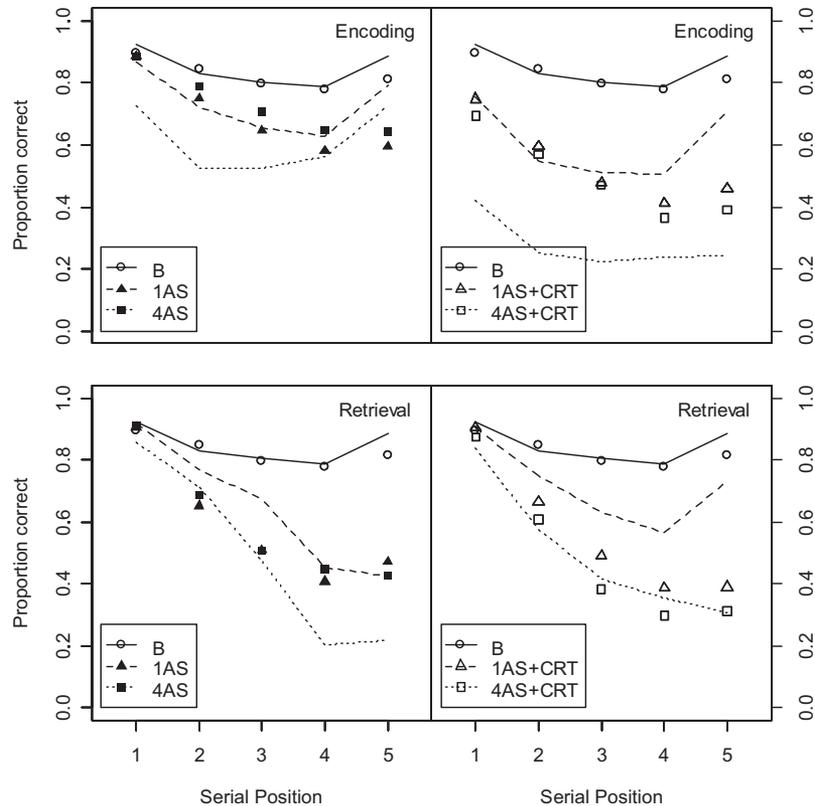


Figure 9. Fit of the primacy model to Experiment 2. Top panels: Delays at encoding. Bottom panels: Delays at retrieval. Left panels: AS-only. Right panels: AS+CRT. Data are represented by points, predictions by lines. B = baseline; AS = articulatory suppression; CRT = choice reaction task.

rehearsal, no adaptations to the new AS+CRT condition were required. Judged by its numerical fit, SIMPLE reproduced the data with moderate success ( $\text{RMSD} = 0.116$ ). The best fitting parameters were  $c_{\text{baseline}} = 8.07$ ,  $t = 0.27$ ,  $s = 3.24$ ,  $o = 0.78$ , and  $o_r = 0.67$ ; Figure 10 shows, however, that the predictions miss the important characteristics of the data. SIMPLE predicts hardly any difference between baseline and the conditions with delays at encoding, and it underpredicts the difference between baseline and the conditions with delays at retrieval.

### SOB

We fit SOB to Experiment 2 using the same five free parameters as before. As before, we added a vector representing the word “super” to the weight matrix once for each distractor event. In the AS+CRT conditions, we added a second vector that represented the CRT activity (i.e., the symbol and the selected response). This vector was created at random, without similarity to either items or the “super” vector. For simplicity, we ignored the fact that there were two different symbols and responses and added the same vector at each distractor event within a trial. A further difference between the AS-only conditions and the AS+CRT conditions was that, in the former but not in the latter, we assumed that the last encoded item can be reencoded once after each distractor event (if distractors were present at encoding). In the AS+CRT condition,

the last item cannot be reencoded because the required attentional bottleneck is occupied by the CRT.

The best fit was obtained with  $\phi_e = 341$ ,  $\phi_s = 0.57$ ,  $N_o = 2.52$ ,  $c = 1.87$ , and  $J_c = 0.48$ . As shown in Figure 11, the model fit the data reasonably well, with  $\text{RMSD} = 0.085$ . In particular, the model predicts the correct relationship between the nine conditions: At encoding, AS-only led to a moderate amount of forgetting relative to baseline, and AS+CRT led to substantially more forgetting. Differences between short and long delays were small. SOB even reproduced the small advantage of four over one AS-only delays at encoding. This effect arises from the reencoding mechanism: Because of energy gating, saying “super” four times after each item adds hardly any further interfering changes to the weight matrix. At the same time, reencoding the preceding item four times improves the signal-to-noise ratio in the weight matrix. When the benefit of reencoding supersedes the loss from interference, longer delays have a beneficial effect on memory.

At retrieval, the model correctly predicts a large amount of forgetting in all four delay conditions. Predicted performance was somewhat worse in the AS+CRT condition than in the AS-only conditions because two distractor vectors rather than one were encoded with each distracting event. For both AS-only and AS+CRT, SOB predicts no effect of long versus short delays. Thus, the model does not reproduce the small amount of fanning

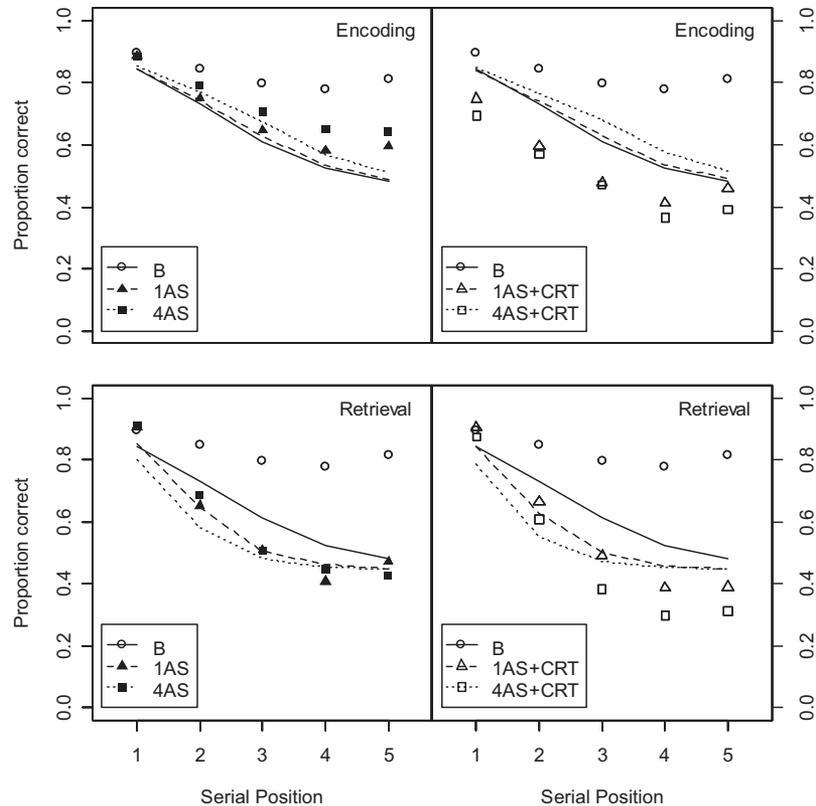


Figure 10. Fit of SIMPLE to Experiment 2. Top panels: Delays at encoding. Bottom panels: Delays at retrieval. Left panels: AS-only. Right panels: AS+CRT. Data are represented by points, predictions by lines. B = baseline; AS = articulatory suppression; CRT = choice reaction task.

in the AS+CRT condition. One way to produce some fanning of one versus four AS+CRT events is to add different vectors for the two stimuli presented in the CRT, but this would mean introducing a further free parameter controlling their similarity for only a minimal gain in fit.

The remaining deviation from the data arises mainly from SOB's inability to capture the exact shape of the serial position curves and, in particular, the interaction of distractor interference with serial position. In Experiment 2, more so than in Experiment 1, the amount of forgetting in the delay conditions relative to baseline increased with serial position. SOB does not predict such an increase for delays at encoding, and it predicts a smaller increase than obtained in the data for delays at retrieval.

This deviation between predictions and data is informative about the nature of interference, or of mechanisms to combat it, in immediate memory. SOB makes the simple assumption that each distractor is encoded in the same way. This generates roughly the same amount of interference for each item across the serial position curve (except for the first item when delays are introduced at retrieval because the distractors preceding that item are associated to the last position marker). The data suggest that interference from the distractor events builds up during encoding as well as during retrieval or, alternatively, that mechanisms to prevent interference weaken toward the end of the list. One possible way of capturing that trend is to avail the model with a more powerful refreshing mechanism. So far, SOB has only a minimal mechanism of reen-

coding the last item just presented. A more extended mechanism would involve cumulative refreshing of all currently held items from the start of the list (Palmer & Ornstein, 1971). Because such a mechanism would have to rely on retrieval and reencoding of list items, rather than on just reencoding the current content of the focus of attention, it could be applied in temporal gaps during recall as well as during list presentation. We do not explore this possibility here because the focus of this article is on distinguishing between three general hypotheses about the nature of short-term forgetting. Our results show that interference is the most promising of the three hypotheses; working out its specific mechanism is an obvious next step for future research.

### Interim Summary of Modeling

A summary of our efforts in augmenting the three models and fitting them to the data of two experiments is presented in Figure 12. The figure plots the TEE and TLS indices produced by the three models with their best fitting parameters for both experiments. The vertical lines in each plot represent the empirical indices summarized in Table 1. The primacy model generates TEE and TLS indices that are far too negative in all conditions. With one exception (TLS in Experiment 1), SIMPLE reproduces the indices reasonably well, but as we have seen, this comes at the cost of missing the effect of distractors, compared with baseline. The

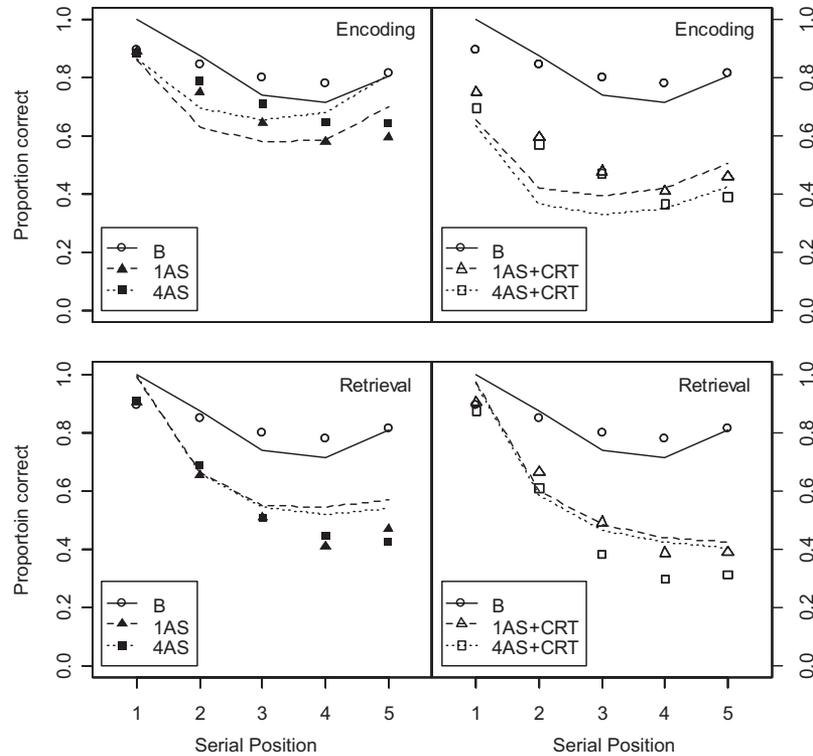


Figure 11. Fit of SOB to Experiment 2. Top panels: Delays at encoding. Bottom panels: Delays at retrieval. Left panels: AS-only. Right panels: AS+CRT. Data are represented by points, predictions by lines. B = baseline; AS = articulatory suppression; CRT = choice reaction task.

indices generated by SOB are all close to or within the range of the data.

### A Positional Decay Model

So far, we have conducted two comparisons among three models of serial recall, incorporating decay, temporal distinctiveness, and interference, respectively. In both comparisons, the interference-based model has provided the preferred explanation of the experimental findings. A critical question at this point is whether the relative success of the three models allows more general inferences beyond these particular models about the three hypothetical sources of forgetting. We chose to address our fundamental question about the role of time in memory within simple and transparent models that enabled us to identify which assumption led to which aspect of each model's behavior. We have shown that the primacy model's problems in accounting for the data arise from the decay assumption, which implies larger effects of temporal delay than observed. Likewise, we have shown that SIMPLE's problem arises from the prediction of a distinctiveness gain with increasing delays at encoding, which can be neutralized but not reversed by adapting the  $c$  parameter to the experiment's time range. Because these mispredictions arose from deep properties of the models, we submit that the problems we uncovered for the primacy model and SIMPLE are inherited by any other instantiation of the decay or temporal distinctiveness notion, respectively.

Nevertheless, the possibility remains that other models may find a way to circumvent these problems. Whereas there are no obvious

alternatives to SIMPLE for implementing temporal distinctiveness, there are other models besides the primacy model that assume decay as one source of forgetting in immediate memory (e.g., Burgess & Hitch, 1999; Daily, Lovett, & Reder, 2001). One feature these models have in common, which sets them apart from the primacy model, is the use of position markers rather than a primacy gradient to encode serial order. This feature could be critical for a decay-based model because it gives a rehearsal mechanism more flexibility. As discussed earlier, in the primacy model, rehearsal must recreate the whole primacy gradient encoded at any given moment, and if the available rehearsal time is not sufficient for rehearsing all currently encoded items, then none can be rehearsed. In a model that uses associations of items to positional markers to maintain order, in contrast, partial rehearsal is possible. Therefore, even short intervals of time can be used to rehearse one or two items.

Partial rehearsal is important not only because it makes rehearsal potentially more powerful but also because the time-based resource-sharing model of working memory (Barrouillet et al., 2004) assumes—with considerable empirical support; see our earlier discussion of the effects of rate of processing rather than absolute time—that even small pauses between the cognitive operations of a distractor task can be used to rehearse memory contents. This assumption, in turn, demands the possibility of partial rehearsal because small time gaps between two cognitive operations typically permit rehearsal of just one or two items. For this reason, the primacy model would not be a natural candidate for

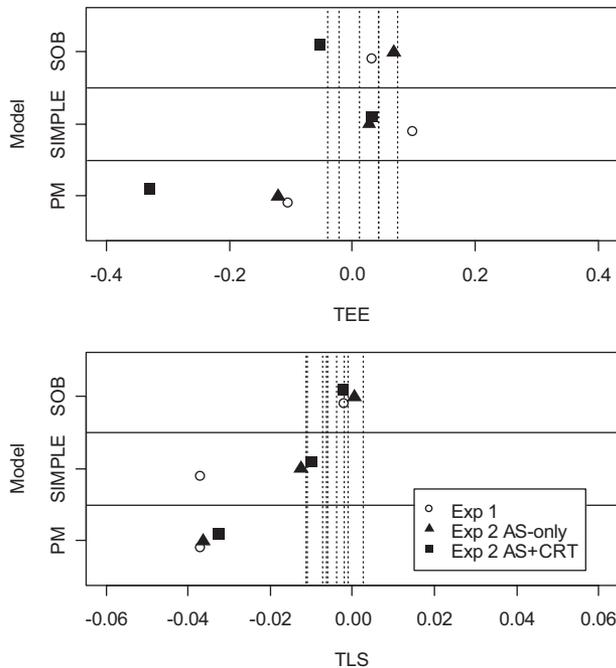


Figure 12. Time effect at encoding (TEE) and time-loss slope (TLS) indices predicted by the augmented versions of the primacy model (PM), SIMPLE, and SOB, using the best fitting parameter values, for Experiment 1 and the two types of distractors in Experiment 2. Vertical dotted lines are the empirical indices from Table 1 (as in Figures 2 and 3). Exp = experiment; AS = articulatory suppression; CRT = choice reaction task.

a computational instantiation of the time-based resource-sharing model. Instead, such an implementation must rely on an architecture that supports partial rehearsal. A positional model of serial recall is therefore the most promising starting point for developing a computational model of the complex span paradigm along the lines of Barrouillet and colleagues (2004).

To investigate whether a positional decay model can account for our data better than the primacy model, we built a simple model implementing decay and rehearsal in a position-based architecture, drawing on ideas of Burgess and Hitch (1999), Daily et al. (2001), and Barrouillet et al. (2004). We opted for testing a generic positional decay model rather than testing an existing model because the existing models all have specific features needed to explain phenomena that are not the focus of our present work (such as phonological similarity and grouping effects), and these features add complexity and potentially obscure the effect of the decay and rehearsal assumptions. Our positional decay model is therefore reduced to the essential mechanisms needed to make the model work for the present experimental paradigm, and thereby, its behavior can be traced to its core assumptions more easily than in a more complex model. The MATLAB code for the positional decay model is available on the Web pages provided in the author note.

### Architecture

The positional decay network consists of two layers, an input layer used to represent positional markers and an output layer with

one unit for each item in the experimental vocabulary. The positional markers are modeled after the moving windows scheme of Burgess and Hitch (1999): Each position marker consists of a section of  $n$  contiguous units, starting with the leftmost section of the input layer for the first position; for each successive position, the activated section moves  $d$  units further to the right. Each position marker overlaps with the neighboring markers to the left and the right, and the proportion of overlapping features between neighbors is determined by a free parameter  $c$ . The length  $n$  of the position markers and their distance  $d$  are set such that all position markers needed for a given list length fit into the 16 units of the input layer with the required proportional overlap  $c$ . The two layers are fully interconnected by a weight matrix  $\mathbf{W}$ .

### Encoding and Decay

When an item is presented, the activation of its corresponding unit in the output layer is set to one, while all other items in that layer remain zero. In the input layer, the current position's marker is activated. The weight matrix  $\mathbf{W}$  is then updated according to the same Hebbian learning rule as used in SOB (see Equation 9), which creates an association between the positional marker and the item. The learning rate  $\eta_e$  in that equation is not modulated by energy but is simply set to one. The weight matrix  $\mathbf{W}$  decays exponentially over time:

$$\frac{dw_{ij}}{dt} = -Dw_{ij}, \quad (18)$$

where  $D$  is a decay parameter.

### Retrieval and Rehearsal

Retrieval proceeds by reinstating the positional marker of the to-be-recalled position in the input layer and forwarding this activation through the weight matrix  $\mathbf{W}$  to the output layer, using SOB's Equation 12. The activation in the output layer is divided by the number of units in each position code,  $n$ , to render activation strength independent of  $n$  and, thus, of the free parameter  $c$ . Gaussian noise with mean zero and standard deviation  $N$  is added to each output unit. The item corresponding to the unit with the highest activation in the output layer is then selected as the recalled item. Recall is followed by response suppression using Hebbian anti-learning. As in SOB, anti-learning consists of subtracting from the weight matrix the association of the currently instantiated context vector with the vector for the recalled item (Equation 15). The rate of anti-learning,  $\eta_s$ , equals  $-1$  times the activation level of the highest activated unit in the output level, reflecting the recall strength of the recalled item.

Rehearsal consists of recalling an item and reencoding the recalled item. Although response suppression following recall eradicates the association between the position marker and the recalled item in the weight matrix, reencoding creates these associations again, thus replacing the already decayed weights with fresh ones. This, of course, reverses the effect of response suppression. Unlike the primacy model, the positional decay model can proceed to the next item nonetheless because response suppression, though helpful, is not essential for moving on to recalling and rehearsing the next item.

Whenever the available time interval exceeds the rehearsal duration  $r$  for rehearsing a single item, rehearsal is initiated with probability  $R$ . This is a critical difference from the primacy model, in which the available time must suffice for rehearsal of the entire list. As with the primacy model, in the encoding phase, the available time for rehearsal is the time between the onset of the current item and its successor minus 250 ms for encoding of the current item. At retrieval, the available time is the interval between recall of the current item and the next one, as measured by keypresses, minus the time for recalling one item, as estimated by the conditions without delays at retrieval. Thus, during recall, there is time for rehearsal only when distractor-filled delays are introduced at retrieval. Whenever rehearsal is initiated, it starts at the beginning of the currently encoded list (during retrieval, already recalled items are excluded from that list, such that only the yet-to-be-recalled items are rehearsed) and proceeds until time runs out.

In summary, the positional decay model has five free parameters: decay rate  $D$ , noise  $N$ , proportional overlap of position markers  $c$ , rehearsal duration  $r$ , and rehearsal probability  $R$ . When fit to Experiment 2, we applied  $R$  as a free parameter to the AS-only condition but fixed it to zero for the AS+CRT condition to reflect the assumption that no rehearsal is possible concurrently with the pair of distractors in that condition.

### *Fitting the Positional Decay Model to Experiments 1 and 2*

We fit the model to both experiments, using a grid search with boundaries determined from previous explorations of the parameter space. The fit to Experiment 1 was followed up by a simplex search using the best fitting parameters from the grid search as starting values. The positional decay model achieved a reasonable fit to Experiment 1, with RMSD = 0.1007. The best fitting parameters were  $D = 0.211$ ,  $N = 0.132$ ,  $c = 0.60$ ,  $r = 0.312$  s, and  $R = 0.393$ . The predictions with these parameter values are illustrated in Figure 13. Similar to the primacy model, the positional decay model underpredicts the amount of forgetting caused by a single distractor, relative to baseline. The model also predicts an effect of longer versus shorter delays at encoding, especially in the earlier serial positions, which was not present in the data. In contrast, the model captures well the small amount of fanning at retrieval.

The fit of the positional decay model to Experiment 2 was less convincing, with RMSD = 0.1479. The best fitting parameters were  $D = 0.13$ ,  $N = 0.125$ ,  $c = 0.52$ ,  $r = 0.31$ , and  $R = 0.59$ . Figure 14 shows that the model predicts a large effect of longer versus shorter delays at encoding, contrary to the data. The model also overpredicts the amount of fanning with delays at retrieval. This was true not only for the AS+CRT condition, in which rehearsal was precluded, but also for the AS-only condition, in which the model chose to rehearse on roughly 60% of all occasions.

As with the primacy model, we explored a version of the positional decay model with more liberal assumptions about rehearsal in Experiment 2. We added a further free parameter,  $R_{\text{crt}}$ , that represented the probability of rehearsal in the AS+CRT conditions. The best fit (RMSD = 0.1025) was achieved with  $D = 0.213$ ,  $N = 0.125$ ,  $c = 0.394$ ,  $r = .318$ ,  $R = 0.81$ , and  $R_{\text{crt}} = 0.61$ .

The introduction of the  $R_{\text{crt}}$  parameter permitted the model to move into a different region of the parameter space, with a larger decay rate and a higher rehearsal probability in the AS-only as well as the AS+CRT conditions. As a consequence, the model predicts less fanning at retrieval (not shown in detail here) compared with the five-parameter version. The positional decay model still largely overpredicts the difference between one and four distractors at encoding, both with AS-only and with AS+CRT.

To summarize, with regard to both its numerical fit and its behavior, the positional decay model was remarkably similar to the primacy model. Like the primacy model, the positional decay model failed to give a satisfactory account of the data because it predicts too large an effect of long versus short delays, compared with the effect of short delays versus baseline. This behavior can be immediately traced to the decay assumption: Decay generates forgetting that is roughly proportional to the time that has passed since encoding. In the experiments, the temporal delay introduced by a single distractor word relative to baseline was smaller than the additional delay caused by several further distractor events. This necessarily makes it hard for any decay-based model to predict a large amount of forgetting caused by a single distractor event, relative to baseline, together with only a small amount of forgetting caused by two or three further distractor events. Even permitting rehearsal during the distractor events does not help the models out of this dilemma: Rehearsal can counteract forgetting during the distractor events and thereby reduce the gap between performance with short and with long delays (i.e., one vs. many distractors). At the same time, however, rehearsal also reduces the gap between conditions with one distractor and the baseline. Even very liberal assumptions about rehearsal concurrent with distractor activity therefore are of limited use for bringing decay-based models into better agreement with our data, and this fact provides one explanation for why the models' best fitting values for rehearsal probability turned out to be fairly low in most cases (with the exception of the six-parameter positional decay model applied to Experiment 2). In the next section, we turn to a closer analysis of rehearsal in the primacy model and the positional decay model, providing a detailed explanation for this unexpected finding.

### Rehearsal Disenchanted

Whenever forgetting is absent in an experiment, advocates of decay are quick to invoke compensatory rehearsal as an explanation. This invocation persists even when the delay period is filled by a distractor task because decay proponents can then argue that rehearsal operates concurrently with the distractor activity or in small time gaps between individual steps of that activity. For instance, from their finding of very little forgetting during an AS-filled delay, Vallar and Baddeley (1982) concluded that "we must assume either that the articulatory loop is not necessary for rehearsal in this task, or else that suppression is insufficient to prevent such subvocal rehearsal" (p. 57). It was in anticipation of this line of reasoning that we explored the performance of the primacy model and the positional decay model when rehearsal was assumed to be possible during the delay periods (by allowing estimates of  $R > 0$ ).

Looking back at those model fits, one surprising result was that the primacy model hardly made use of the possibility to rehearse during the delay periods. The rehearsal probability parameter  $R$

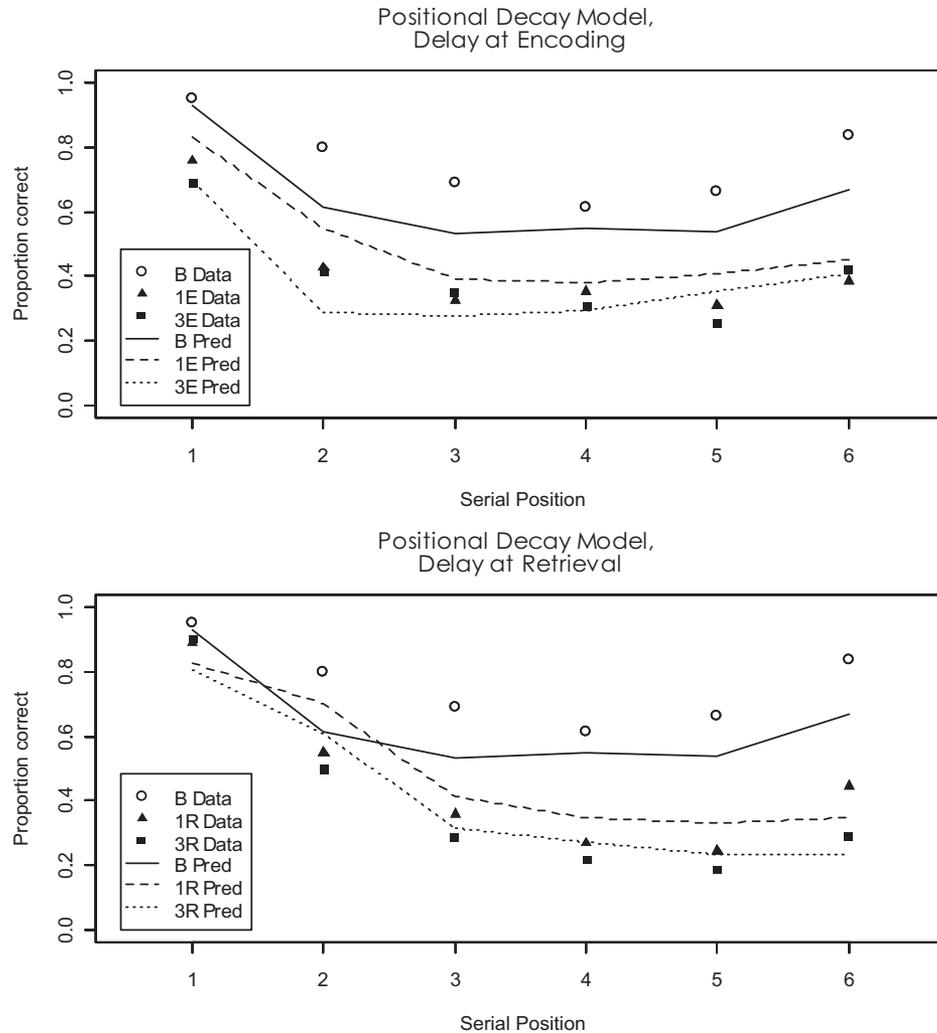


Figure 13. Fit of the positional decay model to the data of Experiment 1. Model predictions are represented by lines, observed values by points. B = baseline (no delay); 1E = one distractor word at encoding; 3E = three distractor words at encoding; 1R = one distractor word at retrieval; 3R = three distractor words at retrieval; Pred = prediction.

was estimated to be 0.3, and in addition, the fairly long estimate of rehearsal duration (500 ms per item) precluded rehearsal most of the time, except when the currently remembered list was short and the available rehearsal time was very long. Our positional decay model made somewhat more use of rehearsal, with  $R$  estimated to be about 0.6 and rehearsal duration estimated to be about 300 ms per item. Still, one may wonder why the model did not achieve its best fit with  $R = 1$  to make the most of the opportunity to counteract decay by rehearsal.

We investigated the reasons for the models' hesitation to use rehearsal by plotting mean accuracy in the five experimental conditions of Experiment 1 as a function of  $R$ , holding the other parameters constant at their best fitting values (see Figure 15). We found that increasing rehearsal probability had a beneficial effect on performance when rehearsal opportunities occurred only during encoding (i.e., when the distractor events were inserted into the encoding phase; see the filled symbols in Figure 15). When most

of the rehearsal opportunities occurred during retrieval, however, rehearsal was not beneficial and could even harm performance (i.e., in the condition with three distractor events at retrieval; open squares in Figure 15). The reason for this harmful effect is that rehearsal involves recall, and when this recall is erroneous, an erroneous list is reencoded. Retrieval errors during rehearsal cumulate and therefore reduce accuracy. When the probability of recall errors is small, the beneficial effect of rehearsal (i.e., counteracting decay) offsets the harmful effect from retrieval errors, but as the probability of recall errors increases, the harmful effect becomes dominant. Rehearsal at retrieval involves recall of a memory trace not better than that underlying overt recall in the baseline condition. Thus, rehearsal at retrieval cannot be less error prone than overt recall in a condition without delays. Therefore, the decay-based models have no way to avoid the harmful effect of rehearsal at retrieval when the lists are difficult enough to yield below-ceiling performance in the baseline condition.

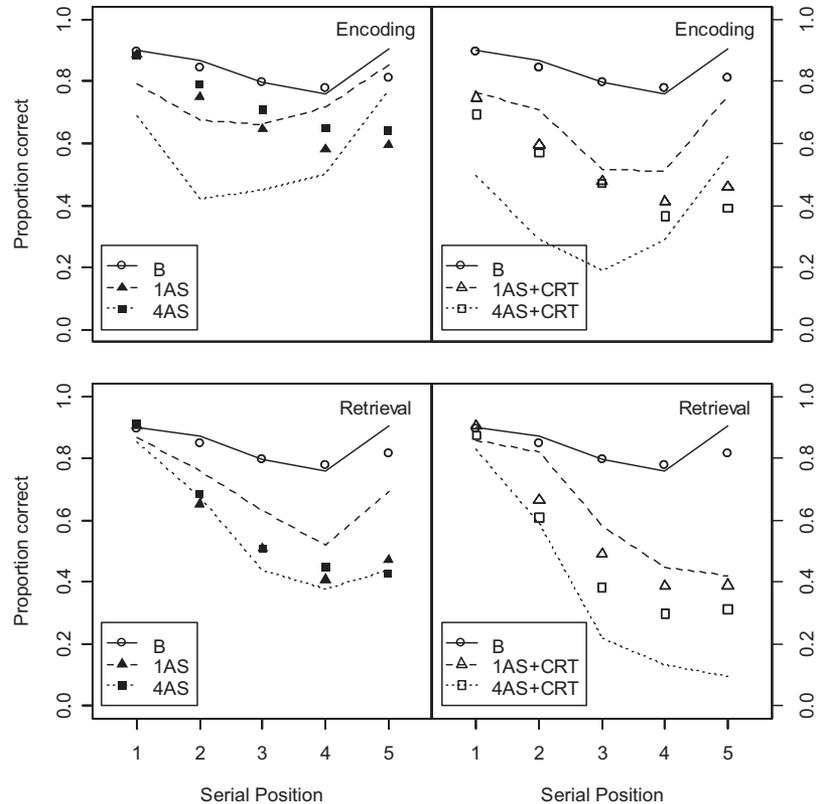


Figure 14. Fit of the positional decay model to Experiment 2. Top panels: Delays at encoding. Bottom panels: Delays at retrieval. Left panels: AS-only. Right panels: AS+CRT. Data are represented by points and predictions by lines. B = baseline; AS = articulatory suppression; CRT = choice reaction task.

We also explored the effect of varying  $R$  at other values of the decay parameter and of rehearsal duration (maintaining the best fitting values for the remaining parameters). Unsurprisingly, in both models, rehearsal was most harmful when decay rate was small and most beneficial in the presence of strong decay. Perhaps more surprising is an observation made in the predictions of the primacy model (but not the positional decay model): When rehearsal duration is shortened from 500 ms to smaller values, rehearsal tends to become more harmful. This trend is very slight between 500 ms and 200 ms, but when moving from 200 ms to 100 ms per item, rehearsal impairs memory dramatically.<sup>8</sup> Shorter rehearsal durations mean that rehearsal is possible in shorter intervals between successive encoding events or retrieval events, and apparently, very short rehearsal durations afford rehearsal at points in time where little is gained from it but much can be lost due to retrieval errors. This finding shows that invoking a very fast rehearsal mechanism that uses even small gaps between distractor events, as assumed in the time-based resource-sharing model, is not necessarily beneficial to memory.

To conclude, if rehearsal is modeled explicitly, it turns out to be anything but the magic cure for decay that it is often purported to be. Rehearsal can be harmful to performance as well as helpful, and its net effect depends on the specific assumptions and parameters of the model in which it is embedded and the task phase during which it is engaged. Importantly, the assumption of rehearsal does not alter the critical mispredictions of the primacy

model and of the positional decay model, namely, that recall is expected to be worse with longer than with shorter delays at encoding or retrieval.

## General Discussion

The goal of this article was to test three hypotheses about the sources of forgetting in immediate memory, using serial recall as a model case. The three hypotheses—time-based decay, temporal distinctiveness, and interference—were represented by three formal models of the serial recall task: the primacy model, SIMPLE, and SOB, respectively. The three models were applied to a serial recall paradigm with filled delays between items either during encoding or during retrieval. To generalize our conclusions about decay, we additionally investigated a positional decay model in which items were associated to context markers.

To disentangle the effects of time from the effects of interference, we used a three-level manipulation, with baseline (no delay), a short delay, and a long delay. The contrast between baseline and short delay reflects the joint effects of increased time and of

<sup>8</sup> This pattern was found in all conditions except with three distractor events at retrieval, where, at the 100-ms rehearsal duration, performance first decreases as  $R$  is increased but then is boosted sharply as  $R$  moves beyond 0.5.

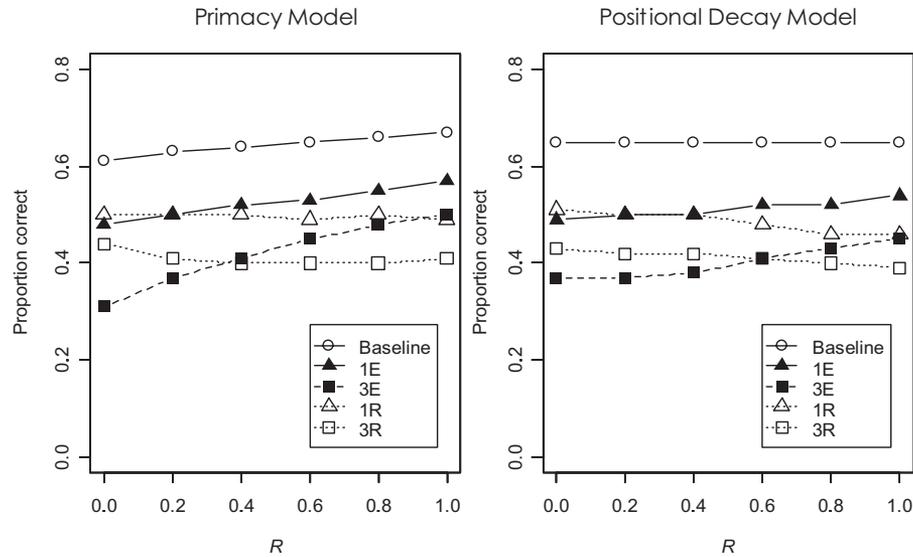


Figure 15. Effect of rehearsal probability  $R$  on recall accuracy as predicted by the primacy model (left) and the positional decay model (right) with best fitting parameters and times obtained from Experiment 1. B = baseline (no delay); 1E = one distractor word at encoding; 3E = three distractor words at encoding; 1R = one distractor word at retrieval; 3R = three distractor words at retrieval.

interference by the distractor task. The contrast between short and long delay, however, was assumed to reflect the relatively pure effect of time because the distractor task was designed to be as repetitive as possible, introducing little additional potentially interfering information in the long-delay condition relative to the short-delay condition.

### Summary of Findings

Our summary of empirical results is based on a large number of experiments, summarized in Table 1, with altogether seven conditions varying delays at encoding and nine conditions varying delays at retrieval. The results of Experiments 1 and 2 are fully representative of the pattern in the table. We conducted a small meta-analysis on the TEE and TLS indices computed from the data in Table 1 (Hunter & Schmidt, 1990), the results of which are shown in the last row of the table. A highly consistent pattern emerged across these experiments.

Introducing short delays at encoding or at retrieval reduced memory performance substantially compared with the baseline condition. Longer delays at encoding did not reduce performance further, and in some conditions, longer delays seemed even to improve performance compared with shorter delays. This improvement is reflected in positive TEE indices in Table 1 and Figure 2. Positive TEE indices occurred in conditions where only AS was used to prevent rehearsal (Experiments 2, 4, and 5 in Table 1). This unexpected finding, if replicable, could point to a process of refreshing or consolidating memory traces that can operate concurrently with overt articulation (cf. Hudjetz & Oberauer, 2007).

Longer delays at retrieval engendered a small additional decline in accuracy at final serial positions, thereby producing fanning in the serial position curves of short versus long delays. As reflected by the TLSs in Table 1 and Figure 3, fanning occurred in some but

not all experiments that we conducted with this paradigm, and we were not able to detect systematic predictors of when it occurs, suggesting that it is always present but its magnitude is so small that it escapes detection in all but the most powerful experiments. Clearly, irrespective of whether it is on occasion statistically detectable, the magnitude of fanning (less than one half of a percentage point of performance loss for each additional second of time) falls far short of what is expected on the basis of decay models, in particular those that inherit the guiding proposition of Baddeley's phonological loop that memory span is linked to the amount of material that can be articulated in about 2 s (Baddeley et al., 1975; Schweickert & Boruff, 1986).

We conclude that extending the delay between retrievals of items sometimes has an adverse effect on memory but that the effect is small—much smaller than the effect of introducing a distractor task (i.e., the contrast between no delay and short delays). Importantly, the effect of extending retrieval delays remained unchanged when we increased the cognitive demand involved in the distractor task from articulating a word (Experiment 1) to articulation combined with a variety of choice tasks (Experiment 2 in this article and Experiments 3–6 in Table 1). Increasing cognitive demands resulted in increasing time demands of the distractor task but no notable increase in the amount of fanning of the serial position curves; Table 1 shows that the average TLS for AS-only conditions was  $-.0045$ , whereas it was  $-.0066$  for the AS+CRT conditions. Together, these results point to a large role of interference in forgetting over the short term and a very small role at best for time.

### Analysis of Models

The model-based analyses confirm this conclusion. The most elaborate contemporary model incorporating temporal distinctive-

ness, SIMPLE, could handle the retrieval effects, but it grossly mispredicts the results of manipulating delays at encoding. The fundamental problem for SIMPLE is that introducing delays between presentation of items at encoding increases temporal distinctiveness. This should lead to improved performance relative to the baseline, contrary to the observed decrement in accuracy. SIMPLE can avoid that prediction by assuming that the similarity gradient, controlled by the parameter  $c$ , is adjusted to the time scale of list presentation. Although this assumption follows precedent in absolute identification (Neath & Brown, 2006) and although it permits SIMPLE to predict that the conditions with delay at encoding are not recalled better than baseline, the model still fails to account for the large decrement observed with delays at encoding.

SIMPLE's inability to explain the pronounced drop in performance from the baseline condition to the condition with a single distractor at encoding persists even when it is augmented by an ordinal dimension that represents the serial positions of items. Inclusion of that ordinal dimension enabled the model to accommodate the null effects of delays at retrieval (Lewandowsky et al., 2004). On the ordinal dimension, distinctiveness between two items is a function of the number of intervening list items regardless of their temporal separation. Thus, when attention is shifted away from time and is focused on ordinal position, the model ceases to be sensitive to any temporal spacing manipulation among items. It follows that SIMPLE correctly predicts no effect of one versus many distractors but also predicts, incorrectly, no effect of a single distractor relative to baseline. These problems are not specific to SIMPLE—any model using temporal or ordinal distinctiveness as a major source of errors in serial recall (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998) will have the same problems explaining the effects of distractors at encoding.

The primacy model as a prototypical decay-based model fared better than SIMPLE. We augmented the primacy model by modeling rehearsal in a more explicit, realistic way than in the published version and by relaxing the assumption that AS fully prevents concurrent rehearsal. The augmented primacy model reproduced many aspects of our data but nonetheless failed to capture a few diagnostic features. Its main problem is that it must predict substantially more decay with long than with short delays. Therefore, the primacy model predicts worse performance with long than with short delays at encoding and substantial fanning of serial position curves for short versus long delays at retrieval. Rehearsal can be used to counteract decay, thereby diminishing the predicted differences between short and long delays, but this also diminishes the predicted difference between short delays and the baseline condition. Therefore, the model has to compromise between assuming powerful rehearsal to bring the short- and long-delay conditions close together and assuming no rehearsal to account for the drop in performance from baseline to short delays. As a result, the best fitting estimate for the rehearsal probability  $R$  was quite low.

A second reason for the low estimate of  $R$  was identified by investigating the costs and benefits of rehearsal. In the original primacy model (Page & Norris, 1998), rehearsal is implemented as regenerating the primacy gradient with perfect accuracy. At the same time, Page and Norris (1998) assumed rehearsal to consist of retrieval and reencoding of the current contents of working memory. When we modeled rehearsal as retrieval and reencoding, we

discovered—not surprisingly—that rehearsal is far from perfect and therefore is not beneficial in all circumstances. Rehearsal during delays at retrieval has more costs than benefits because it introduces errors into the primacy gradient that outweigh the beneficial effect of raising the items' activation. This observation should caution against invoking rehearsal as a mechanism that can compensate decay under any circumstances without modeling these putative processes in detail. Modeling rehearsal as retrieval and reencoding is, of course, not without alternatives—other mechanisms of rehearsal can be conceived, but whether they are effective in saving representations from decay without corrupting them needs to be demonstrated by realistic simulations, rather than simply assumed.

The primacy model represents but one way of implementing decay as a mechanism of forgetting in a formal model of serial recall. It is impossible to prove that no other decay-based model that can account for the present data exists. We were able to show, however, that the key predictions of the primacy model—a large detrimental effect of long versus short delays both at encoding and at retrieval—generalize to an alternative decay model with a different architecture and a different rehearsal strategy. Our positional decay model uses positional coding of order rather than a primacy gradient, and it uses partial rather than exhaustive rehearsal. The latter difference is potentially most promising in the present context because the primacy model's ability to counteract decay by rehearsal may have been hampered by the constraint that rehearsal was possible only when all memorial information could be rehearsed in the time available. Nevertheless, despite allowing partial rehearsal, the positional decay model mispredicts the data in the same way as the primacy model: The positional decay model predicts an excessive TEE and—except for the six-parameter version applied to Experiment 2—an excessive amount of fanning at retrieval. We traced this problem directly to the decay assumption. Therefore, our conclusion that decay is unlikely to be a major source of short-term forgetting generalizes beyond the primacy model and transcends the constraint of exhaustive versus partial rehearsal.

We do not claim to have ruled out any role of decay in immediate memory, but our findings and the informative failure of our efforts to explain them with decay-based models have significantly diminished the plausibility of the decay assumption. The case against decay is particularly strong because it rests on an experimental manipulation of time, whereas an important piece of evidence often cited in favor of the decay hypothesis, the word-length effect, is purely correlational and therefore of little diagnostic value (Lewandowsky & Oberauer, in press).

As an example of an interference-based model, we applied SOB (Farrell, 2006; Farrell & Lewandowsky, 2002) to our experimental paradigm. SOB gave a reasonably good account of the data. It reproduced all the diagnostic qualitative features—a large drop in performance from the baseline condition to the conditions with short delays at encoding or retrieval and hardly any further decrease from short to long delays. SOB also gave the best quantitative account of the data from both experiments. There is room for improvement of SOB in that the serial position curves of the predictions do not perfectly match the shape of the serial position curves in the data, especially in Experiment 2 (see Figure 11). The assumption of SOB that is key to its present success is the notion of energy-gated encoding: New information is encoded into mem-

ory to the degree that it is novel in comparison with the current content of immediate memory. We interpret our findings first and foremost as support for this assumption, not necessarily for all the details of the current version of SOB. Our data thus add to the growing body of evidence for energy-gated encoding (Farrell, 2006; Farrell & Lewandowsky, 2003; Lewandowsky & Farrell, 2008).

### *Implications for Decay and Rehearsal*

Our analysis of decay models, in which rehearsal was explicitly implemented as retrieval followed by reencoding, revealed two problems for this class of models. One arises from the fact that rehearsal is not guaranteed to be error free. This has no dramatic consequences during the encoding phase, when memory traces are still fresh and likely to be retrieved accurately. When delays are introduced at retrieval, however, rehearsal needs to access a memory trace that—by definition—cannot be any better than it would be for immediate overt recall. Thus, if the list is sufficiently difficult to generate recall errors in immediate recall, a comparable number of retrieval errors are to be expected for each sweep of rehearsal during the retrieval phase. When several delays are introduced during retrieval, raising the need for several sweeps of rehearsal, such errors necessarily accumulate, thus increasingly damaging the memory representation. Depending on the model's parameter values, the net effect of rehearsal in the retrieval phase can therefore often be harmful. This result of our modeling compromises blanket appeals to rehearsal as a convenient compensatory variable in any situation in which no forgetting is observed. Our modeling results are consistent with empirical evidence that at least overt rehearsal can be detrimental to recall (Estes, 1991).

The second problem for any model based on the twin assumptions of decay and rehearsal arises from the precise way in which filled delays affected the results of our experiments. At both encoding and retrieval, a single distractor between every item led to a large amount of forgetting, but there was hardly any further forgetting when more of the same distractor events were added. To explain these two findings simultaneously, it is not sufficient to invoke rehearsal as a mechanism to compensate decay. A powerful rehearsal mechanism might indeed fully reverse the effect of decay during each distractor event, such that the number of distractors has no effect on memory. However, if rehearsal is so powerful that it can neutralize decay even during extended continuous articulation, combined with CRTs, then it is most likely powerful enough also to neutralize decay during a single utterance of "super," thus reverting memory strength back to the baseline level. Yet just saying "super" once between every item at encoding or at retrieval reduced accuracy by about 20 percentage points in Experiment 1. Thus, every decay model faces a dilemma: It must assume that rehearsal is largely prevented by the distractor activity to explain the large difference between baseline and the short-delay conditions, but it must assume that rehearsal is hardly prevented at all to explain that short-delay and long-delay conditions are virtually indistinguishable.

### *Implications for the Complex Span Paradigm*

Our experiments, especially the conditions with delays at encoding, resemble the complex span paradigm, which is the para-

digm most frequently used to study individual differences in working memory capacity (Conway et al., 2005). Researchers interested in working memory capacity have discussed largely the same set of hypotheses concerning forgetting in the complex span paradigm as those that are discussed in the literature on immediate serial recall. Towse et al. (2000) as well as Barrouillet et al. (2004) assumed time-based decay as the source of forgetting, whereas Saito and Miyake (2004) argued for representational interference. No detailed formal models comparable to those tested here have been developed for the complex span paradigm as yet. Our applications of SIMPLE, the primacy model, the positional decay model, and SOB to the conditions with delay at encoding can therefore be taken as a start in that direction. The failure of the first three models and the relative success of SOB suggest that interference, not decay or distinctiveness, determines performance in the complex span paradigm.

This conclusion is in conflict with the successful time-based resource-sharing theory of the complex span paradigm proposed by Barrouillet et al. (2004). These authors assumed that information in working memory decays during the distractor period, with the effects of that decay being counteracted by an attention-based rehearsal mechanism. The attentional mechanism is engaged by the distractor task, but according to the model, people can use even small pauses between successive operations to switch attention to refreshing of the memory representations.

From the viewpoint of this theory, an explanation for our findings could be constructed as follows: Continuously speaking the same word, as in Experiment 1 and the AS-only condition of Experiment 2, does not place high demands on the attentional mechanism, which therefore could engage in refreshing during the distractor periods. Support for the notion that rehearsal may occur concurrently with continuous articulation has been obtained by Hudjetz and Oberauer (2007). Therefore, the time-based resource-sharing model does not necessarily predict any effect of increasing the delay during retrieval in these AS-only conditions. To test this idea, we prevented refreshing as well as articulatory rehearsal in the AS+CRT condition of Experiment 2 (and also Experiments, 4, 5, and 6 in Table 1). We reasoned that the dual demand of selecting a response to a visual stimulus while speaking a word aloud would prevent all modes of rehearsal sufficiently to make decay observable. The results did not support this prediction. Adding a CRT to the distractor event resulted in more forgetting relative to baseline, and it gave rise to marginally larger effects of longer versus shorter delays, but the resulting TEE and TLS indices still clustered together with those obtained with AS only, all of which were close to zero (see Table 1 and Figures 2 and 3).

A small effect of increasing the number of distractors is to be expected on the assumption of interference because the representations engaged in the CRT necessarily differ from one distractor event to the next. Decay-based models predict much larger TEE and TLS indices, in particular when rehearsal is entirely blocked by the AS+CRT distractors. Even when the assumption of complete blocking of rehearsal is relaxed and rehearsal is permitted concurrently with AS and the execution of a choice task, the two decay-based models considered here overpredict the effects of increasing delays. This result poses a challenge for the time-based resource-sharing model, as well as any other model trying to explain capacity limits in working memory as emerging from the balance between time-based decay and rehearsal.

An additional problem for the time-based resource-sharing model arises from the dilemma explained above. On the one hand, the model has to explain the pronounced drop in performance when moving from simple span paradigms (i.e., immediate serial recall without distractors) to complex span paradigms (i.e., serial recall with distractors between items at encoding). On the other hand, Barrouillet et al. (2004) showed that the number of distractor events between items at encoding has no effect on memory performance in the complex span paradigm, as long as their pace is held constant. The problem to be solved is how to balance the amount of decay during each distractor event and the amount of regained memory strength through rehearsal between distractor events such that a single distractor event results in substantial memory loss but further distractor events do not incur additional forgetting.

Can we explain the available data on forgetting in the complex span paradigm without decay? The time-based resource-sharing model has gained much support from the finding that performance in the complex span paradigm is a function of the rate of processing operations in the distractor periods (Barrouillet et al., 2004; Barrouillet & Camos, 2001). That is, stretching out the same distractor activity over more time improves memory performance. We can think of two explanations for this finding. One is in terms of increased temporal distinctiveness, but the present results rule out such an account. The other is that longer pauses between processing operations indeed allow more rehearsal, as proposed by Barrouillet et al. (2004). If decay plays no significant role in immediate memory, what could be the role of rehearsal?

A first answer to this question is that rehearsal can be used to (partially) reverse the effect of interference as well as of decay. In support, our minimalist implementation in SOB of attentional refreshing as reencoding of the most recently presented item demonstrably helps to overcome the noise created by the encoding of distractors. A second answer could be that rehearsal contributes to encoding or consolidating the contents of short-term or working memory in long-term memory (McCabe, 2008; Phaf & Wolters, 1993) and that long-term memory traces contribute to immediate recall (Unsworth & Engle, 2007). Further modeling work is needed to explore these and other possibilities before a firm conclusion can be reached.

### Conclusions

Our data and model analyses support five main conclusions. First, neither temporal distinctiveness nor decay models hold much promise for explaining why we forget information from immediate memory and, by implication, why working memory capacity is limited. Second, at least one mechanism of interference, the obligatory encoding of representations used in intermittent distractor activity into immediate memory, can explain why and how a concurrent processing task disrupts immediate recall. Third, in the context of this form of interference, energy-gated encoding provides an explanation for why adding further highly similar distractor events has little additional effect beyond the first distractor event. Fourth, decay models face the challenge of simultaneously explaining a large amount of forgetting arising from a single distractor and hardly any further forgetting during additional distractor events. This data pattern moves the debate beyond the impasse created by the arms race between rehearsal blocking

methods on the one side and the assumption of ever more powerful and flexible rehearsal mechanisms on the other side. Finally, rehearsal is highly overrated—when modeled as a realistic mechanism, it does not reliably compensate for decay.

### Future Outlook

The present research has addressed a very general theoretical question—what causes forgetting in immediate memory?—by testing instantiations of alternative answers in formal models. The gain in explicitness and precision achieved by modeling comes at a cost. Formal models often explain in some detail people's behavior in one or a limited set of experimental paradigms. The field of immediate memory is no exception: On the one hand, there are broad and encompassing verbal theories of working memory (Baddeley, 2001; Cowan, 2005), and on the other, there are computational models that apply to a small range of immediate memory tasks, primarily forward serial recall.

The present work makes some initial steps toward extending the scope of computational models to the complex span paradigm, but there is a rich inventory of immediate memory tests that remains to be explored. Among those tests are close relatives of serial recall, such as the probed recall task, in which one or several items are probed for recall at random (e.g., Oberauer, 2003b; Sanders & Willemsen, 1978); the order reconstruction task (Lewandowsky, Nimmo, & Brown, 2008); and the running memory paradigm (Bunting, Cowan, & Sauls, 2006; Geiger & Lewandowsky, in press; Palladino & Jarrod, in press). In addition, immediate memory is studied with free-recall tasks (Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005) and with a broad palette of recognition tasks (e.g., Luck & Vogel, 1997; Smith & Jonides, 1997; Sternberg, 1969). Moreover, there is a growing number of more complex tasks that have been established as measures of working memory capacity alongside the complex span paradigm, such as the memory-updating paradigm in which participants remember a small number of items and update them through cognitive operations (Oberauer & Kliegl, 2001).

An often implicit assumption in experimental studies is that these different paradigms reflect at least in part the operation of the same short-term or working memory system. Factor analytic studies have provided support for that assumption by showing that many, if not all, of these paradigms share a large amount of variance (Cowan et al., 2005; Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000), suggesting that they share the cause or causes of forgetting as the performance-limiting factor.

For most models of serial recall, including the primacy model, the positional decay model, and SOB, an extension to these other paradigms is far from straightforward. SIMPLE, a model formulated on a more abstract level, can be applied to at least some other paradigms beyond serial recall, including free recall and probed recall (Brown et al., 2007; for a debate on whether serial and free recall should be handled by a single mechanism, see Brown et al., in press; Murdock, in press), but it is not clear how SIMPLE would apply to tasks that integrate short-term retention with information manipulation, such as the memory-updating task. For some of the immediate memory paradigms besides serial recall, formal models exist (for free recall: Davelaar et al., 2005; for memory updating: Oberauer & Kliegl, 2006; for immediate recognition: Oberauer, in

press), but they are not connected yet to existing models of serial recall.

A theory of forgetting over the short term should be applicable across all tasks that can be validated, by experimental and factor analytic research, to reflect immediate memory. Computational models of short-term or working memory should aspire to the same breadth of scope. Having built a first catwalk between the literatures on short-term and working memory by focusing on a single integrative paradigm, we consider it an important next step to broaden the scope of modeling to encompass the full set of tasks employed to test immediate memory in both fields.

## References

- Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55*, 64–88.
- Anderson, J. A. (1991). Why, having so many neurons, do we have so few thoughts? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 477–507). Hillsdale, NJ: Erlbaum.
- Anderson, J. A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.
- Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84*, 413–451.
- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.
- Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limits on retrieval. *Cognitive Psychology, 30*, 221–256.
- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Clarendon Press.
- Baddeley, A. D. (2001). Is working memory still working? *American Psychologist, 56*, 851–864.
- Baddeley, A. D., & Lewis, V. (1984). When does rapid presentation enhance digit span? *Bulletin of the Psychonomic Society, 22*, 403–405.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575–589.
- Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavioral Research Methods, Instruments, & Computers, 28*, 584–589.
- Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*, 83–100.
- Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 570–585.
- Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay? *Journal of Memory and Language, 45*, 1–20.
- Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113*, 201–233.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.
- Brown, G. D. A., Chater, N., & Neath, I. (in press). Serial and free recall: Common effects and common mechanisms—A reply to Murdock (2008). *Psychological Review*.
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*, 539–576.
- Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*, 127–181.
- Bunting, M. F., Cowan, N., & Saults, J. S. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology, 59*, 1691–1700.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551–581.
- Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*, 627–652.
- 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 & Review, 12*, 769–786.
- Cowan, N. (2005). *Working memory capacity*. New York: Psychology Press.
- Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*, 42–100.
- Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language, 46*, 153–177.
- Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science, 25*, 315–353.
- Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigation of recency effects. *Psychological Review, 112*, 3–42.
- Estes, W. K. (1991). On types of item coding and sources of recall in short-term memory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 155–173). Hillsdale, NJ: Erlbaum.
- Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language, 55*, 587–600.
- Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9*, 59–79.
- Farrell, S., & Lewandowsky, S. (2003). Dissimilar items benefit from phonological similarity in serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 838–849.
- Fitzgerald, P., & Broadbent, D. E. (1985). Order of report and the structure of temporary memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 217–228.
- Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language, 51*, 136–158.
- Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition, 26*, 263–276.
- Geiger, S. M., & Lewandowsky, S. (in press). Temporal isolation does not facilitate forward serial recall—or does it? *Memory & Cognition*.
- Glenberg, A. M., Bradley, M. M., Kraus, T. A., & Renzaglia, G. J. (1983). Studies of the long-term recency effect: Support for a contextually guided retrieval hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 231–255.
- Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*, 73–137.
- Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49(A)*, 80–115.
- Hudjetz, A., & Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. *Memory & Cognition, 35*, 1675–1684.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

- Johnson, M. K. (1992). MEM: Mechanisms of recollection. *Journal of Cognitive Neuroscience*, 4, 268–280.
- Jolicoeur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology*, 36, 138–202.
- Jones, D. M., Beaman, P., & Macken, W. J. (1996). The object-oriented episodic record model. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 209–238). Hove, England: Psychology Press.
- Lepine, R., Barrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychonomic Bulletin & Review*, 12, 165–170.
- Lewandowsky, S., Brown, G. D. A., Wright, T., & Nimmo, L. M. (2006). Timeless memory: Evidence against temporal distinctiveness models of short-term memory for serial order. *Journal of Memory and Language*, 54, 20–38.
- Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review*, 11, 771–790.
- Lewandowsky, S., & Farrell, S. (2008). Phonological similarity in serial recall: Constraints on theories of memory. *Journal of Memory and Language*, 58, 429–448.
- Lewandowsky, S., & Farrell, S. (in press). Short-term memory: New data and a model. *Psychology of Learning and Motivation*.
- Lewandowsky, S., Geiger, S. M., & Oberauer, K. (in press). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language*.
- Lewandowsky, S., Nimmo, L. M., & Brown, G. D. A. (2008). When temporal isolation benefits memory for serial recall. *Journal of Memory and Language*, 58, 415–428.
- Lewandowsky, S., & Oberauer, K. (in press). The word-length effect provides no evidence for decay in short-term memory. *Psychonomic Bulletin & Review*.
- Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 53(A), 1–22.
- Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281.
- McCabe, D. (2008). The role of covert retrieval in working memory span tasks: Evidence from delayed recall tests. *Journal of Memory and Language*, 58, 480–494.
- McElree, B., & Doshier, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, 118, 346–373.
- Murdock, B. (in press). Issues with the SIMPLE model: Comments on Brown, Neath, and Chater (2007). *Psychological Review*.
- Naveh-Benjamin, M., & Jonides, J. (1984). Maintenance rehearsal: A two-component analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 369–385.
- Neath, I., & Brown, G. D. A. (2006). SIMPLE: Further applications of a local distinctiveness model of memory. *Psychology of Learning and Motivation*, 46, 201–243.
- Nimmo, L. M., & Lewandowsky, S. (2005). From briefs gaps to very long pauses: Temporal isolation does not benefit serial recall. *Psychonomic Bulletin & Review*, 12, 999–1004.
- Oberauer, K. (2003a). Selective attention to elements in working memory. *Experimental Psychology*, 50, 257–269.
- Oberauer, K. (2003b). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language*, 49, 469–483.
- Oberauer, K. (in press). How to say no: Single- and dual-process theories of short-term recognition tested on negative probes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Oberauer, K., & Kliegl, R. (2001). Beyond resources: Formal models of complexity effects and age differences in working memory. *European Journal of Cognitive Psychology*, 13, 187–215.
- Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55, 601–626.
- Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity: Facets of a cognitive ability construct. *Personality and Individual Differences*, 29, 1017–1045.
- Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105, 761–781.
- Palladino, P., & Jarrold, C. (in press). Do updating tasks involve updating? Evidence from comparisons with immediate serial recall. *Quarterly Journal of Experimental Psychology*.
- Palmer, S. E., & Ornstein, P. A. (1971). Role of rehearsal strategy in serial probed recall. *Journal of Experimental Psychology*, 88, 60–66.
- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, 116, 220–244.
- Pashler, H., Johnston, J. C., & Ruthruff, E. (2000). Attention and performance. *Annual Review of Psychology*, 52, 629–651.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Phaf, R. H., & Wolters, G. (1993). Attentional shifts in maintenance rehearsal. *American Journal of Psychology*, 106, 353–382.
- Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing: A minimal executive function. *Cortex*, 43, 135–145.
- Raye, C. L., Johnson, M. K., Mitchell, K. J., Reeder, J. A., & Greene, E. J. (2002). Neuroimaging a single thought: Dorsolateral PFC activity associated with refreshing just-activated information. *NeuroImage*, 15, 447–453.
- Rickard, T. C., & Bajic, D. (2004). Memory retrieval given two independent cues: Cue selection or parallel access? *Cognitive Psychology*, 48, 243–294.
- Ruthruff, E., Pashler, H. E., & Klaassen, A. (2001). Processing bottlenecks in dual-task performance: Structural limitation or strategic postponement? *Psychonomic Bulletin & Review*, 8, 73–80.
- Saito, S., & Miyake, A. (2004). An evaluation of the task switching account of working memory span scores: Evidence against a temporal decay assumption. *Journal of Memory and Language*, 50, 425–443.
- Sanders, A. F., & Willemsen, E. M. (1978). The course of proactive interference in immediate probed recall. *Acta Psychologica*, 42, 133–144.
- Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 419–425.
- Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 51(A), 283–304.
- Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, 33, 5–42.
- Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 57, 421–457.
- Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition*, 28, 341–348.
- Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, 123, 34–80.
- Unsworth, N., & Engle, R. E. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114, 104–132.
- Vallar, G., & Baddeley, A. D. (1982). Short-term forgetting and the articulatory loop. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 34(A), 53–60.

Received December 19, 2006

Revision received February 12, 2008

Accepted February 16, 2008 ■