Journal of Experimental Psychology: Learning, Memory, and Cognition

On the Existence of Semantic Working Memory: Evidence for Direct Semantic Maintenance

Geeta Shivde and Michael C. Anderson

Online First Publication, August 15, 2011. doi: 10.1037/a0024832

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/11/$12.00 DOI: 10.1037/a0024832

On the Existence of Semantic Working Memory: Evidence for Direct Semantic Maintenance

Geeta Shivde
West Chester University

Michael C. Anderson
Medical Research Council Cognition and Brain Sciences Unit, Cambridge, England

Despite widespread acknowledgment of the importance of online semantic maintenance, there has been astonishingly little work that clearly establishes this construct. We review the extant work relevant to short-term retention of meaning and show that, although consistent with semantic working memory, most data can be accommodated in other ways. Using a new concurrent probe paradigm, we then report experiments that implicate a semantic maintenance capacity that is independent of phonological or visual maintenance that may build on a mechanism of direct semantic maintenance. Experiments 1 through 5 established that while subjects maintain the meaning of a word, a novel delay-period marker of semantic retention, the *semantic relatedness effect,* is observed on a concurrent lexical decision task. The semantic relatedness effect refers to slowed response times when subjects make a lexical decision to a probe that is associatively related to the idea they are maintaining, compared to when the probe is unrelated. The semantic relatedness effect occurred for semantic but not for phonological or visual word-form maintenance, dissipated quickly after maintenance ends, and survived concurrent articulatory suppression. The effect disappeared when subjects performed our immediate memory task with a long-term memory strategy rather than with active maintenance. Experiment 6 demonstrated a parallel phonological relatedness effect that occurs for phonological but not semantic maintenance, establishing a full double dissociation between the effects of semantic and phonological maintenance. These findings support a distinct semantic maintenance capacity and provide a behavioral marker through which semantic working memory can be studied.

Keywords: semantics, working memory, inhibition, attention

Much of human thought relies on the ability to temporarily store and retain information in an active, highly accessible state. Although considerable research in cognitive psychology and cognitive neuroscience has addressed this ability in the context of storing phonological and visual information, much of our mental life concerns the processing of meaning. The ability to actively maintain semantic representations underlies our success and efficiency in nearly all complex cognitive activity, whether we are solving a problem, devising a plan, deciding between options, learning a new fact, comprehending an utterance, or preparing the next thought for translation into written words. In this article, we refer to the ability to maintain semantic representations in a stable, highly accessible state as *semantic working memory*. Our primary aims in this article are to build an empirical case for semantic working memory and to argue that this capacity is supported by a system that is functionally distinct from other known working memory subsystems.

Despite a dramatic expansion of research on working memory over the last two decades, surprisingly little attention has been devoted to how people temporarily store semantic content. On computational grounds, a system that maintains semantics in an active state is seen as a necessary component to theories of higher level cognition and is included in many computational models, and this type of maintenance follows from the frameworks of several authors (e.g., Cowan, 1995; Martin & Saffran, 1997; Ruchkin, Grafman, Cameron, & Berndt, 2003). Theoretically targeted studies examining semantic maintenance are limited, despite its putative importance. Some attention was devoted to the issue in early research on short-term memory, mainly in reaction to the claim that short-term memory was fundamentally acoustic (Raser, 1972; Shulman, 1970, 1972). But research on this topic never took firm root in the classic literature, and later decades have seen little work on the subject. The notable exception comes from studies with neuropsychological patients showing that performance on phonological and semantic maintenance tasks is anatomically dissociable (Martin & Romani, 1994; Martin, Shelton, & Yaffee, 1994). These findings provide the most focused evidence to date for a distinct system supporting semantic maintenance (see also Haarmann & Usher, 2001, for another approach to semantic maintenance).

Geeta Shivde, West Chester University; Michael C. Anderson, Memory Group, Medical Research Council Cognition and Brain Sciences Unit, Cambridge, England.

Portions of this work were completed in fulfillment of the requirements for the dissertation of Geeta Shivde or were supported by a National Institute of Mental Health grant to Michael C. Anderson. We are grateful to Douglas Hintzman for his helpful advice, Timothy Carman and Sarah Parrish for help running the experiments, and Nathan Foster for help in preparation of the manuscript. We thank Douglas Hintzman, Ben Levy, Ulrich Mayr, and Mike Posner for helpful comments on initial drafts of this paper.

Correspondence concerning this article should be addressed to Geeta Shivde, Department of Psychology, West Chester University, West Chester, PA 19383. E-mail: gshivde@wcupa.edu

Although existing research is compatible with semantic working memory, a significant ambiguity runs through all of the evidence for this construct. First, we illustrate this ambiguity by briefly reviewing research relevant to semantic working memory conducted over four decades. Through this review, we show that nearly all findings appearing to support semantic working memory are compatible with the alternative view that evidence for semantic maintenance can be accommodated by (a) the storage and retrieval of items in episodic memory or (b) long-term semantic priming for processed concepts (e.g., Joordens & Becker, 1997; Woltz, 2010), persisting without the need for aid from a maintenance process. Given these alternatives, we argue that no current evidence establishes that people can temporarily store and retain semantics over a delay through semantic maintenance. Without evidence for semantic maintenance, it is argued, there is no reason to posit semantic working memory.

To address this persisting ambiguity, we then introduce a new, theoretically targeted approach for isolating semantic maintenance: the *concurrent probe paradigm*. We use this method to address two issues. First, we examine whether semantic representations can be maintained actively over a delay, rather than merely being retained passively as episodic memories or primed concepts. Second, we examine whether semantic maintenance is achieved by a process focusing directly on semantic representations or instead by mediation through other systems. Can an idea be retained over time, without phonological or visual rehearsal, through direct semantic maintenance? Showing that semantic maintenance does not rely on other systems would bolster the case for a distinct capacity. Supporting these aims, we report six experiments using the concurrent probe paradigm to show that active, direct semantic maintenance exists and can be dissociated from other forms of working memory.

The Case for Direct Semantic Maintenance

In its most basic form, the idea of semantic working memory amounts to the claim that semantic representations can be actively maintained without having to maintain phonological or visuospatial representations.[1] As with other types of working memory, semantic maintenance is thought to keep items accessible; when active maintenance is terminated, access to traces is reduced. Storage and maintenance can be accomplished either by domain-general structures and processes or by a distinct system dedicated to semantics. By the latter view, semantic working memory has its own storage capacity, operating characteristics, and maintenance mechanisms, all of which are supported by structures anatomically distinct from those supporting other varieties of working memory. Accordingly, items can be maintained in semantic working memory while items are maintained in other storage systems, with minimal interference. Theorists addressing the idea of semantic working memory (e.g., Martin & Romani, 1994; Martin et al., 1994) have often endorsed the specialized systems view, though not all work has assumed this.

Although the domain-general and domain-specific views of semantic maintenance differ in the systems that support semantic working memory, both assume *direct semantic maintenance*—that is, they assume that active maintenance can be targeted directly and selectively at semantic representations. Here we review the extant evidence for direct semantic maintenance and argue that even the case for this elementary claim needs to be better built. The core problem has to do with the assumption that immediate memory performance reflects maintenance. It is shown that this assumption is often unclear (and in some cases wrong), and, as a result, the inference of semantic maintenance is not warranted. In most cases, studies of semantic retention do not distinguish the contributions of direct semantic maintenance from those of long-term episodic and semantic retrieval. These alternatives arise in nearly all of the research on semantic maintenance that has been done in the last four decades.

Past research has attempted to isolate semantic maintenance in a variety of ways. Before presenting our approach, we give examples of such work, highlighting useful contributions while also illustrating the theoretical ambiguities that pervade work in this area. We classify findings according to four lines of evidence: the disruptive effects of semantic interference; the beneficial effects of semantic blocking; the dependence of semantic retention on recency; and the beneficial effects of redundantly storing semantics and phonology. We then discuss work that has attempted to dissociate semantic and phonological working memory in neuropsychological patients. We begin by discussing assumptions important to our analysis of prior studies.

Assumptions of the Review

A key premise of the review is that performance in putative semantic working memory tasks may be supported by structures other than semantic working memory. If so, the contributions of those systems must be controlled to establish semantic maintenance. We assume several conventional (though not universally accepted) processes and structures that might support maintenance. First, we assume that performance can often be supported by either episodic or semantic memory, even in rapid tasks. In particular, we assume that presenting an item even very briefly (e.g., 1 s) may lead to its storage in episodic memory and also to priming of its concept in semantic memory. Correct performance on a putative working memory task based on the retrieval of an episodic trace would not constitute evidence for semantic working memory, as a semantic representation would have not have been actively retained throughout the delay. Similarly, enhanced semantic retrieval based on semantic priming may or may not be evidence of semantic maintenance. If the priming of semantic concepts in long-term memory is intentionally sustained over a delay period, it would be

[1] By trying to establish that semantics can be maintained independent of visual working memory, we do not mean to imply that visual representations are not semantic. In fact, generalized knowledge about the appearance of objects might well be considered semantic. Rather, the focus here is on showing that there exists a generalized capacity for maintaining meaning that does not rely on already established visual maintenance systems and that stores nonvisuospatial (abstract) aspects of meaning.

The term *working memory* is used in a variety of ways by different investigators, with some investigators requiring both maintenance and manipulation for something to be considered working memory (Engle, 2002). Here we use the term *semantic working memory* specifically to refer to active maintenance over a delay, based on semantic maintenance processes alone, without the need for manipulation of the contents being maintained. We believe that this maintenance system underpins performance in complex tasks that require manipulation of semantic content.

evidence of active semantic maintenance; if semantic priming was induced at encoding but not maintained over a delay, this would not be evidence of semantic maintenance. We view the capacity to maintain semantic content for sustained online storage to be a key component of direct semantic maintenance.

Second, we assume that performance can be supported by phonological and visuospatial working memory systems. In particular, presenting an item may lead to its storage in the phonological loop or visual working memory, which may in turn indirectly preserve activation of that item's semantic representation. Correct performance in a putative semantic working memory task based on retention in these other systems would not be evidence of semantic working memory, even if such maintenance indirectly sustained activation of the concepts underlying the items. Such activation would not have been generated by a maintenance process acting directly on the semantic representation and thus would constitute mediated semantic maintenance. The strongest inference of direct semantic maintenance would be permitted when episodic memory, primed semantic retrieval and mediated maintenance can be rendered implausible.

Inferring Direct Semantic Maintenance Through Semantic Interference

One way to determine the nature of the representations in working memory is to disrupt retention with interference. By this approach, if semantically similar distractors affect retention more than do phonologically similar ones, working memory must be storing semantics.

Presenting semantically related lists immediately before a working memory trial impairs short-term retention. Evidence for this comes from studies using the classic buildup of proactive interference procedure (Wickens, Born, & Allen, 1963). In this procedure, subjects receive trials in which they encode triads of verbal items (e.g., three letters, numbers, or words) that they then recall in order after varying filled delays (typically 0 to 20 s long). In studies of proactive interference, recall of a given triad is examined as a function of its serial position in a sequence of trials presenting related triads. Typically, recall of the first triad is very good, but performance on the following triads declines substantially with each list. This buildup of proactive interference occurs only when the preceding triads are related to the current one. For instance, Wickens et al. (1963) found that recall declined when the first three triads contained numbers; when the fourth triad was composed of letters, recall dramatically improved. Similar buildup and release effects have been found when the triads presented words from one semantic category (e.g., fruits) and then shifted to another (e.g., to rocks; Loess, 1967, 1968; Turvey, Cremins, & Lombardo, 1969; Wickens & Clark, 1968; for a review, see Wickens, 1970). Thus, encoding semantically similar items before a working memory trial disrupts short-term retention.

Although these studies show that proactive interference from semantically similar lists impairs short-term retention, this is unlikely to be a pure measure of working memory. In the buildup of proactive interference procedure, each triad is tested after a delay filled with a demanding distractor activity (e.g., mental arithmetic) that precludes rehearsal. Under these conditions, subjects are likely to recall the triad by using episodic retrieval rather than working memory. If so, semantic interference from previous sets may impair episodic retrieval rather than working memory (for a clear statement of this position, see Wickens, Moody, & Dow, 1981; Wickens, Moody, & Vidulich, 1985). Consistent with this view, proactive interference in the buildup of procedure is limited to tests given after filled delays, with effects most readily observed after 5–10 s. Proactive interference effects are far smaller at shorter intervals (e.g., Hofer, 1965; Keppel & Underwood, 1962; Loess, 1964) and are absent altogether when retention is tested immediately after a triad is presented, as measured on a speeded item recognition test (Wickens et al., 1981). Taken together, these considerations suggest that proactive interference in classical studies may not reflect semantic interference in working memory. Similar ambiguities affect the interpretation of semantically based retroactive interference effects in short-term retention (e.g., Brown, 1958; Corman & Wickens, 1968; Dale & Gregory, 1966; Weeks & Katz, 1970). Thus, interference effects do not provide clear support for direct semantic maintenance. However, semantic working memory also provides a viable account, even if it is not compelled by the data.

Inferring Direct Semantic Maintenance Through Semantic Blocking at Encoding

Immediate recall is better when a list of words is presented blocked by semantic category than when it is presented with the items randomly interspersed, a phenomenon known as a semantic blocking effect. For instance, Calfee and Peterson (1968) presented eight-item lists at a rate of one word per second. The lists contained four exemplars from each of two categories, presented either contiguously in blocks or randomly interspersed. Subjects recalled an item from one of the eight positions following presentation. Recall was better with blocked presentation, even though the same items had been studied in each case (see also Sanders & Schroots, 1968; Warrington, Kinsbourne, & James, 1966). Huttenlocher and Newcombe (1976) even found semantic blocking effects in a serial recall task when items were presented four per second. These findings support semantic maintenance, insofar as semantic codes would have to be available in working memory for blocking to have an effect.

Here again, this reasoning presumes that immediate memory tasks are pure reflections of working memory—an assumption questioned since the early days of research on short-term memory (Waugh & Norman, 1965). This assumption is suspect in supraspan lists such as those used in all studies of semantic blocking. Blocking effects may simply reflect the influence of semantic organization on long-term retention and not on maintenance. Indeed, blocking effects were discovered in tests of episodic memory (Cofer, Bruce, & Reicher, 1966; Cohen, 1966; Dallett, 1964). Blocking may highlight relatedness among the exemplars at encoding, leading to greater organization in long-term memory. Blocking may also facilitate the initial semantic processing of those items later in a block through priming, increasing the effectiveness with which they are encoded in the limited time given. Independent of these effects, blocking may ensure that subjects use categorical retrieval cues (Calfee & Peterson, 1968), providing a potent retrieval advantage. Finally, enhanced semantic encoding may facilitate the use of long-term memory representations at retrieval to reconstruct degraded traces stored in the phonological loop (Poirier & Saint-Aubin, 1995). Thus, semantic blocking effects do not provide clear evidence for semantic maintenance.

Inferring Direct Semantic Maintenance Through Semantic Recency Effects

The recency effect has often been taken to reflect the contribution of working memory to recall. Several lines of work have tried to use this assumption to build a case for semantic working memory. Haarmann and Usher (2001), for instance, found evidence favorable to semantic maintenance. Subjects rapidly encoded 12-item lists and were given an immediate free-recall test. Each list presented six related word pairs either adjacently or separated by five other words. Prior work had established that adjacent presentation facilitates recall of related items because it improves encoding (Glanzer & Schwartz, 1971). Would this "adjacency effect" occur if the items appeared in the final serial positions, which are more likely to reflect working memory? Some adjacency effect might occur in recent positions due to long-term memory. But would some residual adjacency effect remain even after long-term memory was considered? The contributions of long-term memory were estimated by running another condition in which subjects recalled the list after a filled interval (Craik & Levy, 1970; Levy & Baddeley, 1971). When the recency effect in immediate retention was corrected for this contribution, a residual benefit was observed, consistent with the storage of semantic information in working memory.

A related argument based on recency was published in classical work by Shulman (1971, 1972). Many early studies of recency effects found strong evidence for phonological coding in short-term memory, perhaps because the tasks used did not specifically encourage semantic coding. Shulman (whose procedure we adapt in this article) was the first to address this concern. In Shulman's procedure, subjects encoded 10-item lists and were immediately tested on a single word. Just prior to the test probe, a task signal (a letter) appeared, telling whether they were supposed to judge whether the probe was identical to, rhymed with, or was synonymous with an item on the list. Because the test task varied from list to list and was not revealed until just before the probe, subjects were forced to encode words at all levels of analysis in preparation for all types of cue. Shulman found very similar retention functions for all probe types. Marked recency effects were observed for the synonym task, as should occur if semantic information was maintained in a short-term store (see also Raser, 1972). Thus, these studies indicate that semantic information can be coded in some short-term tasks and that semantic retention shows recency effects consistent with a short-term storage system (see Bregman, 1968, for a similar finding).

Shulman's (1971,1972) findings, like Haarmann and Usher's (2001), provide clear evidence that semantic recency effects exist, consistent with the idea that semantic representations can be in a heightened state of activation that is gradually lost. However, even granting this interpretation, nothing in these findings establishes whether semantic representations can be actively maintained. Without demonstrating a distinct capacity to actively maintain semantics over a delay, these findings can be explained by appeal to semantic priming effects (particularly long-term semantic priming), without requiring a separate maintenance system. Thus, although these findings are quite compatible with semantic maintenance, they do not compel this construct.

Inferring Direct Semantic Maintenance From Apparent Dual Storage Effects

If an independent capacity supports direct semantic maintenance, subjects should have an additional way of achieving immediate retention that improves recall compared to when only phonological maintenance is possible. Consistent with this, several findings demonstrate an immediate recall advantage for items with richer content that could take advantage of both systems. Bourassa and Besner (1994) found that lists composed of content words (e.g., *farm, great*) were recalled reliably better than lists of semantically impoverished function words (*such, while*). This advantage even persisted when items were encoded under articulatory suppression, showing that semantic information did not merely facilitate rehearsal in the phonological loop. In memory span tasks, words are better recalled than nonwords (Hulme, Maughan, & Brown, 1991; Hulme, Roodenrys, Brown, & Mercer, 1995), high-frequency words are recalled better than low-frequency words (Gregg, Freedman, & Smith, 1989; Tehan & Humphreys, 1988; Watkins, 1977), and nonwords that have word "entries" in semantic memory (e.g., *brane*) are better recalled than nonwords that do not, even under articulatory suppression (Besner & Davelaar, 1982).

Here again, however, direct semantic maintenance cannot be inferred. These effects can be attributed to long-term episodic or semantic memory. For example, it is known that "deeper" encoding tasks typically yield better long-term retention than "shallower" ones based on the phonology or appearance (Craik & Tulving, 1975). The semantic content associated with short-term memory items could facilitate deeper encoding. Later, subjects may recall items from both episodic memory and the phonological loop, yielding a recall advantage for meaningful items. Second, a preexisting semantic representation may improve recall by facilitating reconstruction of a degraded trace in the phonological store—a hypothetical process known as *redintegration* (Schweickert, 1993). Thus, semantic information can influence immediate retention without implicating a semantic working memory system. Consistent with these arguments, most investigators have interpreted these findings as evidence for the role of long-term memory during span tasks (Bourassa & Besner, 1994; Hulme et al., 1991, 1995; Poirier & Saint-Aubin, 1995), rather than for semantic maintenance. However, these findings do not compel the long-term memory interpretation either. Dual storage might better explain some of these findings, a possibility that has not generally been considered.

Inferring Direct Semantic Maintenance From Anatomical Dissociations

Neuropsychological research suggests that direct semantic maintenance might be supported by brain systems that are distinct from phonological maintenance. Case studies (Hoffman, Jefferies, Ehsan, Hopper, & Lambon Ralph, 2009; Hoffman, Jefferies, & Lambon Ralph, 2011; Martin & Romani, 1994; Martin et al., 1994; Romani & Martin, 1999; Wong & Law, 2008) have shown qualitative differences in the working memory impairment in patients with different lesions. For instance, Martin and colleagues showed that patient A.B., who has damage to the left prefrontal and adjacent parietal cortex, has difficulty with semantic maintenance but is better at phonological tasks. Patient E.A. shows deficits on phonological working memory tasks but has better semantic maintenance and has damage to the left temporal and parietal cortices. If semantic and phonological working memory can be dissociated in this way, it would strongly support a distinct semantic maintenance capacity.

In an early study, Martin et al. (1994) compared E.A. and A.B. on procedures intended to measure phonological and semantic working memory. For instance, in probe recall tasks designed after Shulman's (Shulman, 1970, 1972), each patient was presented with lists containing one to seven words, one item per second. A novel probe followed 2 s after the end of the list. In the rhyme condition, subjects judged whether the probe rhymed with one of the presented items, whereas in the semantic condition, they judged whether it belonged to the same category as one of the studied items. Estimates of phonological and semantic working memory span were then derived for the two patients and a matched group of controls. In the rhyme judgment condition, A.B.'s span (4.62 items) was impaired relative to that of controls (7.02) but was significantly better than E.A.'s (2.65). In contrast, A.B.'s category span (2.19) was worse than E.A.'s (2.82), and the spans for both patients were worse than those of controls (5.38). A.B. also failed to show effects of the previously discussed semantic variables (e.g., lexicality) in short-term retention tests. In contrast, A.B., showed the normal disruption by phonological similarity on letter and word span tasks, whereas E.A. did not, at least when visual presentation was used. Together, these findings suggest that although both patients were impaired, E.A. is more noticeably impaired in maintaining phonology, whereas A.B. was more noticeably impaired on tasks involving semantic maintenance.

Martin and colleagues have explored how these deficits affected the patients' language comprehension. In a strong demonstration, the patients were asked (auditorily) general knowledge questions, such as "Which is quiet, a concert or a library?" To answer, the subjects had to maintain the attribute and the options in memory. Strikingly, E.A. answered 100% of these questions correctly, whereas A.B. answered only 20%. However, A.B. scored perfectly when asked about this same knowledge in a simplified way (e.g., auditory presentation of "Is a library quiet?") or when given an unpaced visual test with the original questions, showing that he knew the answers. On another test, the patients were read sentences varying in complexity and requesting them to perform an action. For instance, when asked to "touch the large green circle," they were supposed to touch the relevant object in an array in front of them. When the sentences required the maintenance of several adjectives (as in the previous example), A.B. made many more errors than did E.A. In related findings Martin and Romani (1994) showed that A.B.'s ability to detect semantic anomalies was poor when there were several adjectives preceding a noun (e.g., "rusty old red swimsuit") but much better when only one adjective preceded the noun or when the adjectives followed the noun (e.g., "The swimsuit was old, red, and rusty"). They argued that the before/after difference reflected the fact that the adjectives could be integrated with the noun immediately in the after condition and this reduced the need to rely on retention of individual word meanings. These findings thus suggest the importance of semantic maintenance for language.

Although A.B. appears to have disrupted semantic maintenance, other interpretations are also possible. Unfortunately, many of the tasks and manipulations used to establish A.B.'s deficit have not themselves been shown to require semantic working memory. For instance, the fact that A.B.'s retention is not modulated by semantic variables such as lexicality is not diagnostic of a deficit in semantic maintenance; as discussed earlier, these effects typically have been attributed to the role of long-term memory (Hulme et al., 1991, 1995; Poirier & Saint-Aubin, 1995). By this view, A.B. may simply be less able to use episodic or semantic memory during the test to reconstruct phonological traces. Similarly, as discussed earlier, the category probe task adapted from Shulman (1970, 1972) does not require semantic maintenance; residual semantic priming or episodic memory can explain probe performance on this task as well. A similar priming account could be offered for the sentence comprehension findings.

If not a deficit in semantic maintenance, what might A.B.'s deficit be? One possibility is that he has difficulty retrieving word meanings, given their visual or phonological forms. A deficit in retrieving semantics would be consistent with neuroimaging research demonstrating the role of the left inferior prefrontal cortex (damaged in A.B.) in semantic retrieval (e.g., Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Thompson-Schill et al., 1998; Wagner, Pare-Blagoev, Clark, & Poldrack, 2001). This work has shown that simple semantic judgments about words (e.g., concreteness or animacy judgments) activate this region and that the degree of activation predicts later episodic memory for those words (Wagner et al., 1998). If A.B. had a deficit in semantic retrieval, his semantic encoding should be more prone to fail, particularly with rapid presentation. Consistent with this concern, Romani and Martin (1999) later reported a dramatic episodic memory deficit in A.B. that was restricted to his memory for words—a deficit so severe that it rendered A.B. as or more impaired than a group of patients with traditional amnesia. It is crucial that A.B's episodic memory deficit (as measured by recognition memory) disappeared when he had 2 s to encode each study item instead of just 1 s. This strongly supports the speculation that the 1-s presentation rate used in Martin et al.'s (1994) studies of semantic retention may have undermined A.B.'s long-term encoding. By this view, even if semantic working memory did not exist, A.B. should show the immediate memory deficits reported in Martin et al., based purely on deficiencies in initial encoding speed (for a related perspective, see Hoffman et al., 2009, 2011).

More recent evidence from patient M.L., however, replicates many of the findings with A.B. except that M.L. had single-word semantic processing speed that was close to the normal range (Martin & He, 2005; see also Wong & Law, 2008). This suggests that degraded semantic retrieval may not be the whole story and that semantic maintenance may truly be disrupted. More fundamentally, however, if the same neural structures (left inferior prefrontal cortex) support semantic retrieval and semantic maintenance, it may be difficult to separate these abilities based on lesion evidence alone. Deficits in encoding and maintenance will usually co-occur. If so, establishing direct semantic maintenance will require analytic cognitive methods that can separate these components in neurologically normal subjects. If evidence for direct semantic maintenance can be obtained, it would converge with Martin et al.'s (1994) focused evidence for a distinct semantic maintenance capacity, lending strong support to that view. One of the objectives of the current work is to provide such support and specify the mechanisms by which maintenance is accomplished.

Summary and Evaluation

As the preceding review illustrates, various findings are consistent with semantic maintenance. Despite these favorable findings, a persisting ambiguity runs through all the evidence for semantic maintenance. As we have argued, the influence of episodic and semantic memory on immediate recall can accommodate nearly all of the findings supporting semantic maintenance. Given the persistence of this alternative for over four decades, a new approach is needed. Showing that semantic manipulations influence performance is not sufficient to establish direct semantic maintenance; immediate memory tasks are influenced by multiple short- and long-term systems. Showing that short-lived semantic representations underlie immediate memory performance is not sufficient; we already know that long-term semantic priming exists and that such effects last long enough (over a dozen trials) to account for most evidence in immediate memory tasks without active maintenance (e.g., Woltz, 2010). What is needed to demonstrate a full maintenance-storage system is a way to show that not only is performance guided by active semantic representations but also that their activational state has been intentionally maintained over a delay. In the next section, we describe a new paradigm to demonstrate active maintenance and to evaluate whether maintenance is semantic in character.

A New Approach to Establishing Active Maintenance: The Concurrent Probe Method

As should be evident from the preceding review, the case for semantic maintenance has yet to be established clearly. Nearly all of the evidence can be accommodated by the contributions of long-term episodic and semantic memory to immediate retention. If the relative contributions of these systems vary across conditions, it is difficult to conclude that one's manipulations are affecting working memory per se. A new approach to semantic maintenance is clearly needed.

In the current experiments, we address this ambiguity and examine whether a distinct working memory capacity supports semantic maintenance. We developed a new paradigm that is diagnostic of working memory maintenance: the concurrent probe method. The concurrent probe method builds on the idea that the state of maintained items should differ from that of items not being maintained, to develop behavioral markers diagnostic of semantic maintenance. We used the concurrent probe method with the delayed judgment procedure to develop a delay-period marker of direct semantic maintenance. We describe this procedure next.

The Delayed Judgment Procedure

The delayed judgment task (see Figure 1) is a simple procedure we devised to allow subjects to sustain a single item over a delay. Each trial began by presenting a word in red (the maintained item), which subjects kept in mind over a variable length delay (the maintenance interval). Allowing subjects to focus on a single item ensured that memory load was within their capacity, minimizing special strategies they might use to retain multiple items (e.g., chunking). When the maintenance interval ended, a blue item appeared (the target item), and subjects made one of several judgments about its relationship to the maintained item. Thus, subjects never overtly recalled the maintenance item but kept it in mind to make the target judgment.

Trial Type

Event	During		After
1	**ANGRY**	**<--- Maintained item --->**	**ANGRY**
2	shoe		shoe
3	water		water
4	anchor	**Target item --->**	**MAD**
5	lammer		lammer
6	bise		bise
7	gold		gold
8	yell	**<--- Probe word --->**	yell
9	slode		slode
10	gask		gask
11	fog		fog
12	**MAD**	**<----Target item**	anchor
13	dilk		dilk

Figure 1. The sequence of events in a typical *during* (left) and a typical *after* (right) trial in the delayed judgment paradigm. All trials are embedded in a continuous lexical decision task. Each trial begins with the presentation of a maintenance item (presented in red but shown here in gray boldface), which the subject is to keep in mind continuously throughout the delay period while performing the lexical decision task. After a delay, a target item (presented in blue but shown here in black boldface) appears and the subject is to make a judgment about the target's relatedness to the maintenance item, with the nature of this task varying between conditions (in this figure, a synonym judgment target is depicted). Buried in the lexical decision stream is a single probe (see marked item) that is either semantically associated or unrelated to the maintenance item. In the during condition, this probe appears during the maintenance period (in between the presentation of the maintenance item and the target); in the after condition, the probe appears after the maintenance period has ended. Across the during and after conditions, the position of the probe with respect to the maintenance item is held constant.

The delayed judgment task can be used with different instructions and target judgments to encourage distinct kinds of working memory. In Experiments 1 and 2, we encouraged either semantic or phonological maintenance. In the semantic maintenance condition, we asked subjects to sustain attention on the meaning of the maintained item during the delay and to not keep the word in mind by pronouncing it to themselves. At the end of the maintenance interval, a blue target item appeared and subjects judged whether it meant the same thing as the idea they had in mind. This synonym task, modeled after one used by Shulman (1970), ensured that semantic information was relevant to the maintenance task and encouraged semantic working memory.

We contrasted the semantic maintenance condition to one in which subjects maintained the word's phonology. The phonological condition matched the semantic condition but focused subjects' attention on the sound of the red word instead of its meaning. At the end of the delay, to ensure that the word's sound was task relevant, a single-syllable nonsense word appeared as the blue target and subjects judged whether it shared a vowel sound with the maintained item. These measures should have encouraged phonological rehearsal. As discussed in the introduction to Experiment 4, a similar strategy can be used to encourage visual word-form maintenance. The method's power is that precisely the same maintenance items can be presented across conditions and the underlying structures and processes that support maintenance can be manipulated.

We expected high target task performance, given that subjects needed to maintain only a single item. Thus, unlike in most working memory studies, target task performance was not a concern (except in Experiment 5). Rather, interest centered on performance on a secondary task done continuously throughout the experiment. It is through manipulations of this task that we implemented the concurrent probe method crucial in establishing semantic maintenance. Next, we describe the rationale underlying the concurrent probe method.

The Concurrent Probe Method

The concurrent probe method assumes that maintaining an item preserves it in a state of heightened accessibility—a state that changes after terminating maintenance. Given this, the strongest evidence for maintenance would be to establish that an item is in a special state while being maintained, compared to when it is not. To establish this, we must probe the item's state both during and after maintenance through concurrent and post-maintenance probes. If these probe measures differ reliably even when the encoding-probe delay is matched (to control for decay), active maintenance has occurred. Thus, the basic requirement is that maintenance yield an effect distinct from what occurs in its absence. If not, maintenance cannot be inferred, even if overall probe performance is high. This approach contrasts with that of most working memory studies, which measure performance after maintenance ends. As argued previously, measuring performance after maintenance cannot establish that active maintenance has occurred: Correct recall may be due to episodic retrieval or to priming that lingers.

In the present experiments, we adapted the delayed judgment task for use with the concurrent probe method. To do this, we had subjects perform the delayed judgment task while engaged in an apparently unrelated lexical decision task. The lexical decision task presented a stream of words and nonwords, one at a time, continuously throughout the maintenance interval and after it. During each working memory trial, a single critical lexical decision served as a probe into the state of the maintained item (see Figure 1). This probe occurred at various points during maintenance (*during* condition) or after it (*after* condition). We varied whether probes were semantically related or unrelated to the maintenance item. We hypothesized that performance on related and unrelated probes might differ because attention to a concept often influences associated ideas (the predicted direction of this difference is discussed in the next section). We refer to a reliable difference in reaction time to related and unrelated probes as a semantic relatedness effect. It is important that a semantic relatedness effect is an indirect indication of maintenance that circumvents direct reporting of the maintained item. Indeed, indirect tests are the only tractable way to assess an item's state both during and after maintenance, because direct tests after maintenance has putatively ended could induce subjects to continue maintaining.

Of course, a semantic relatedness effect does not itself indicate active maintenance. Semantic processing during encoding of the maintained item could have lingering effects that alter probe performance. Such long-term semantic priming might arise even if no maintenance occurred (e.g., Woltz, 2010). To establish that maintenance caused the semantic relatedness effect, we compared this effect across the during and after conditions because the effect should vary depending on whether maintenance is occurring. To ensure that there were no differences in priming across the during and after conditions, we matched the lag between encoding of the maintenance item and the probe (see Figure 1). Because priming should not differ when lag is matched, any difference between these conditions must arise from sustaining the maintenance item. Thus, this comparison meets the requirement (for inferring maintenance) of the concurrent probe logic that maintenance have an effect different from what would occur in its absence.

Delay-Period Markers of Direct Semantic Maintenance

Even if our semantic maintenance condition produced a semantic relatedness effect that was delay period specific, it might simply reflect the indirect influence of phonologically rehearsing the maintained item. An effect arising under these conditions would not constitute evidence for direct semantic maintenance. Establishing direct semantic maintenance rather requires that the semantic relatedness effect is specific both to the delay period and to semantic maintenance itself.

To establish the semantic relatedness effect as a delay-period maker for semantic maintenance, we contrasted how semantic and phonological maintenance instructions influenced this hypothesized effect in Experiments 1–2, in hopes of observing a dissociation. If semantic but not phonological maintenance induces this effect, phonological maintenance would be an unlikely cause of effects in the semantic condition. We also chose abstract maintenance words to reduce the chance that visual working memory could underlie semantic maintenance. To validate that subjects did not use visual imagery, we administered a postexperimental questionnaire in all experiments. In a later experiment, we also instructed subjects to use visual working memory to perform our task, in hopes of observing a similar dissociation.

The foregoing dissociations would constitute strong evidence for active semantic maintenance and establish the semantic relatedness effect as a marker of this process. Observing this marker would be compelling, given that the various conditions use precisely the same maintenance items, lexical decisions, probes, and the dual task requirements, with the main variation being in subjects' orientation toward different maintenance types. Such effects could not easily be attributed to persisting semantic priming or to episodic retrieval. The question remains, however, about the nature of this hypothetical effect and its mechanisms.

The Semantic Relatedness Effect: Contrasting Hypotheses

Until now, we have not specified the nature of the semantic relatedness effect because its predicted direction depends both on the theoretical framework and the empirical precedents to which one refers. On one hand, attending related meanings may prime probe items; on the other hand, sustained attention to a concept may inhibit related ideas. We discuss these hypotheses next.

Priming Hypothesis

Maintaining the meaning of a word might spread activation (Collins & Loftus, 1975) to related concepts, hastening reaction times for related probes. Although traditional semantic priming studies have not required the sustained prime processing (Neely, 1991), if attending primes facilitates related words at short delays, sustained prime processing might continue to spread activation, affecting related words appearing later in the delay interval. After maintenance, however, minimal priming should occur, because attention has been removed from the maintenance item and the after probe occurs four items (about 6 seconds) after the target.

According to this hypothesis, phonological maintenance should produce less priming than semantic maintenance. Processing an item's sound might spread activation to its meaning (and further to associates, like the probe), but this indirect, mediated route seems likely to produce a smaller priming effect than would attention to semantics.

Inhibition Hypothesis

Maintaining a word's meaning instead might suppress associated concepts. This may arise precisely because sustaining attention to a concept may activate other concepts to which it is associated. If activation accumulates as the maintenance interval progresses, neighboring concepts may grow activated enough to intrude. If enough intrusions occur, the odds of attention shifting away from the maintained item increase, undermining maintenance. Thus, even if attending to the maintenance item initially spreads activation, sustaining attention on that item over a long delay may ultimately require cognitive control to suppress associated concepts. Put simply, concentrating on an idea requires us to stop our mind's tendency to wander from one idea to the next. This view predicts slower reaction times for related probes, especially at the longest delays.

The suppression hypothesis is recommended by research on inhibitory control in memory. For instance, Anderson and Green (2001) found that when subjects were asked to prevent a memory item from entering awareness when confronted with a reminder to it, later recall for the excluded memory was impaired. Anderson and Green established that active inhibition of the distracting memory itself produced this effect. Although the goals differ in the current paradigm, similar mechanisms may apply. Instead of trying to keep a particular memory out of awareness, subjects in the current procedure instead are directed to sustain attention on an idea over a delay. Although this task does not state that particular concepts are to be excluded, this requirement is implicit, insofar as associated concepts activated during maintenance would undermine maintenance goals. Although Anderson and Green's findings concerned episodic memory, similar mechanisms have been demonstrated in semantic memory (see, e.g., Johnson & Anderson, 2004; see also Blaxton & Neely, 1983; Dagenbach, Carr, & Barnhardt, 1990; for a review, see Levy & Anderson, 2002). Indeed, several investigators have suggested that inhibition supports working memory (Conway & Engle, 1994; Hasher & Zacks, 1988; Jonides, Smith, Marshuetz, Koeppe, & Reuter-Lorenz, 1998; Lustig, Hasher, & Tonev, 2001). Thus, semantic maintenance may cause an inhibitory semantic relatedness effect. However, because noninhibitory mechanisms might also underlie response slowing (see the General Discussion), we would consider such a finding only suggestive.

General Predictions

As the preceding discussion illustrates, the semantic relatedness effect may take on several forms. The effect's direction is not central, however, to our main question: the existence of direct semantic maintenance. The concurrent probe logic does not require the effect to be positive or negative, only that two standards be met. First, whatever the effect is, its magnitude should differ during the retention interval or after it (given a constant encoding-probe delay), showing delay- period specificity. The most obvious manifestation of this would be greater priming or inhibition during the delay. However, the main requirement is that a marker effect is altered by maintenance. If this can be established, the mechanisms underlying it can be specified later.

Second, however maintenance affects the semantic relatedness effect, this effect should differ with maintenance type. If so, one has established both delay-period and semantic specificity. We hoped to establish both these features to build a case for semantic working memory. Experiments 1 and 2 apply this logic to dissociate semantic and phonological working memory.

Experiment 1

In Experiment 1, we manipulated whether subjects performed semantic or phonological maintenance. In the semantic condition, subjects received a single word in red and were asked to think of its meaning continuously throughout the delay, without repeating its sound. By explicitly instructing subjects to use semantic maintenance, we hoped to increase the use of this strategy compared to that in prior studies. In the phonological condition, subjects instead were asked to think of the word's sound. After the delay, a target word appeared in blue and subjects made a synonym judgment (semantic condition) or a rhyme judgment (phonological condition). Subjects performed a concurrent lexical decision task, in which a critical probe word was inserted that was either related or unrelated in meaning to the maintained item.

If direct semantic maintenance occurs, we should find a semantic relatedness effect (the difference in reaction times between related and unrelated probes) that is modulated by whether maintenance is occurring or not and also by whether maintenance is semantic or phonological.

Method

Subjects. Thirty-two undergraduate native English speakers took part to fulfill a course requirement.

Design. Maintenance strategy, probe relatedness, and probe position were manipulated in a $2 \times 2 \times 2$ within-subjects design. Subjects were asked to use either a semantic or a phonological maintenance strategy during a block; on each trial, probe words were either semantically unrelated or related to the maintenance item for that trial and appeared either during the maintenance interval or afterward. We recorded subjects' reaction time to make a lexical decision to the probe.

In addition, we manipulated whether the correct answer to the final target judgment in the maintenance task was "yes" or "no." In each condition, half of the targets had a "yes" response and half of the targets had a "no" response. Target judgment accuracy was also recorded.

Materials.

Stimuli for the semantic condition. In the semantic maintenance condition, 80 triplets were designed, each containing a maintenance item (e.g., *ANGRY*), a target judgment item (e.g., *MAD*), and a critical probe (e.g., *HIT*). The maintenance and target items were synonyms that were each 3–10 letters long, with an average concreteness rating of 3.0 (1 being abstract and 7 being concrete; Friendly, Franklin, Hoffman, & Rubin, 1982) and word frequencies ranging from 0 to 399 (Kucera & Francis, 1967). Each critical probe was chosen to be associated but not synonymous with the maintained item. According to the Nelson, McEvoy, and Schreiber (1994) norms, the average forward strength of association from the maintained item to the probe item was .17. The probes were 3–9 letters long and ranged in frequency from 0 to 170 (Kucera & Francis, 1967). In addition, the 80 triplets were inspected to prevent strong intertriplet associations. There were an additional 14 synonym pairs without a corresponding probe word used in filler trials.

Half of the critical trials contained related probes, and half of them contained unrelated probes. The 40 unrelated probe trials were created by swapping the probes from 20 triplets with those in another set of 20, so that the probes for a given trial would no longer be semantically related to the maintenance item. In a similar manner, the trials for which the correct target response was "no" were created by reassigning half of the original list of target items to different triplets, such that there were an equal number of target-yes and target-no responses for related and unrelated trials and for during and after conditions. The stimuli were fully counterbalanced so that every word triplet participated in each of the within-subjects conditions across subjects (see Appendix A).

Stimuli for the phonological condition. To make the stimuli for the phonological condition, we created new target items for the triplets in the semantic condition (see Appendix A). We made target items by creating a single-syllable nonsense word that shared a vowel sound with a syllable from the maintenance word (e.g., *ANGRY ZEE*). The nonsense words were orthographically dissimilar to the maintenance word to discourage visual maintenance strategies. In addition, 14 nonsense words were used as stimuli during filler trials.

Lexical decision stimuli included 370 English words (3–9 letters long) and 370 nonwords created by substituting one or two letters in a real word. The lexical decision words ranged in frequency from 10 to 400 and were not strongly related to the maintenance items, target items, or probes in any trial. The ordering of words and nonwords was determined randomly, with the constraint that no more than three word or nonword trials in a row were permitted.

Procedure.

Semantic maintenance block. On each trial of the semantic maintenance block, subjects first saw the red maintenance word in the center of the screen. After 2 s, the word disappeared and subjects were asked to think about its meaning for several seconds until another word appeared in blue (the target item). Subjects were instructed to focus on the meaning or idea of the red word and not its sound. When the blue word appeared, subjects were asked to decide whether it meant the same thing as the idea they were keeping in mind. Subjects pressed a *yes* key if they thought the red and blue words meant the same thing and pressed a *no* key otherwise. The target word disappeared as soon as the subject responded or after 3 s, whichever came first.

During the interval between the presentation of the maintenance item and the target (the delay interval) and also between each target judgment and presentation of the next maintenance item, subjects performed lexical decisions. Subjects viewed black letter strings one at a time and decided whether each was a word by pressing a *yes* or *no* key. Letter strings disappeared once a judgment was made or after 2 s had elapsed, with the next string appearing 2 s after the last stimulus onset. Subjects were asked to make their judgments as quickly as possible while paying attention to accuracy. As described earlier, the critical probes were embedded in this sequence of lexical decisions, but the instructions did not indicate any relation between the maintenance and lexical decision tasks. Only one third of the trials in a block (taking into account filler trials) contained probe items that were semantically related to the maintenance item. Subjects were told to make these lexical decisions while they were thinking about the meaning of the maintenance item and were warned that doing two tasks at once would be challenging.

Before the semantic maintenance block, subjects were given two miniblocks of practice trials of increasing complexity. This allowed them to grow accustomed to the maintenance and lexical decision tasks. In the first practice block, subjects did six trials of the maintenance task alone. In the second block, the lexical decision task was added and subjects completed three trials. Each experimental block consisted of 47 trials (40 experimental and seven filler), which were presented as a continuous stream without breaks within the block.

Figure 1 shows schematics of the trial structure in the during and after conditions. Each trial began with a maintenance item presented in red. In the during condition, 10 lexical decision events followed, with the probe presented as the seventh item. The target appeared in blue as the 12th event. After the target, one more lexical decision followed before the next maintenance item appeared. In the after condition, the maintenance item was followed by two lexical decisions and then the target. After the target, nine lexical decisions followed before a maintenance item appeared for the next trial. The probe in the after condition appeared as the fourth lexical decision after the target, equating the number of intervening events between the maintenance item and the probe across the during and after conditions. On filler trials, the maintenance and target items were separated by either five or eight lexical decisions to reduce the predictability of when the target item would appear. Fillers contained no related or unrelated probe trials.

Phonological maintenance block. On each trial, subjects again saw a word in red in the center of the screen for 2 s. They were asked to keep the sound of the red word in mind until the blue target appeared. After 10 lexical decision events, a single-syllable blue nonsense word appeared. Subjects decided whether the vowel sound in the target occurred in any part of the word they were keeping in mind by pressing the *yes* or the *no* key. As in the semantic maintenance block, subjects performed a concurrent lexical decision task. The timing for each type of stimulus (i.e., maintained, probe, or target item), the trial structure, and the composition of trials matched that in the semantic block. Subjects completed two practice blocks.

Postexperimental questionnaire. After the working memory tasks, subjects rated how often they used mental imagery, either of the word itself or of a related concept, to keep the meaning of the word in mind. Subjects made this rating on a scale of 1 (*never*) to 5 (*always*).

Results and Discussion

A mixed analysis of variance was performed on probe reaction time, probe accuracy, target reaction time, and target accuracy, using a 2 (maintenance type) × 2 (probe relatedness) × 2 (probe position) design. Presentation order of the tasks (semantic/phonological vs. phonological/semantic) was included as a between-subjects factor. There was no main effect of presentation order, nor were there interactions with the within-subjects variables in any of the following analyses in Experiment 1 ($p > .2$ in all cases). At first, the probe reaction time data were also analyzed with target type (synonym or nonsynonym) included as a within-subjects factor. Although this factor could not have affected probe reaction times in the during condition (because the probe comes before the target), related probes occurring after the target could have been differentially primed. Surprisingly, none of the effects reported in this article interacted with target type (i.e., whether the target was a synonym of the maintenance item). Thus, all analyses exclude this factor. Appendix B summarizes probe reaction times broken out by target response.

Because there were few reliable effects of our manipulations on probe or target accuracy in this or any of the experiments, analyses of these variables are reported only when reliable results were obtained (see Appendices C and D for a summary of probe and target accuracies for all experiments). Subjects' imagery ratings on the questionnaire were analyzed for this and all the following experiments. These data are discussed in the Results of Experiment 5.

Probe reaction time. If subjects actively sustained the meaning of the maintenance item, reaction times to related and unrelated probes should differ in the semantic maintenance condition. Subjects were indeed slower to respond to related ($M = 850$, $SD = 181$) than to unrelated probe items ($M = 764$, $SD = 122$) during the maintenance interval, $F(1, 30) = 17.54$, $p < .01$, $MSE = 6{,}696$ (see Figure 2), showing that there is a semantic relatedness effect that takes the form of a reaction time slowing.

The existence of a semantic relatedness effect does not by itself imply that subjects were actively maintaining the meaning of the maintenance item. It is possible, for example, that encoding the maintenance item might have affected semantically related representations in a way that would have been observed on the probe regardless of whether the item had been maintained. If processing the maintenance item deeply is sufficient to cause the semantic relatedness effect, this effect should arise in the after condition as well—a condition in which critical probe item occurred at the same absolute distance from the maintenance item as it did in the during condition. This did not occur: There was no reliable semantic relatedness effect in the after condition, $F(1, 30) = 2.03$, $p > .1$, $MSE = 4{,}975$, and there was a two-way interaction between probe relatedness and probe position, $F(1, 30) = 5.27$, $p < .05$, $MSE = 5{,}565$. These findings show that the semantic relatedness effect is delay period specific, consistent with active maintenance.

Next, we examined whether the semantic relatedness effect occurred in the phonological maintenance condition. If phonological maintenance caused the inhibitory semantic relatedness effect, this effect should appear when subjects silently repeat the maintenance item. However, phonological maintenance did not cause a semantic relatedness effect in the during condition; in fact, probe reaction times in the related condition were numerically faster ($M = 780$, $SD = 132$) than in the unrelated condition ($M = 807$, $SD = 148$), although this effect did not reach significance, $F(1, 30) = 3.95$, $p = .056$, $MSE = 3{,}006$; see Figure 2). There was no semantic relatedness effect after the maintenance interval ($F < 1$), nor was there an interaction of probe relatedness and probe position ($F < 1$). Thus, phonological maintenance clearly generates a different pattern than do semantic maintenance instructions, an impression confirmed by a reliable three-way interaction of maintenance strategy, probe relatedness, and probe position, $F(1, 30) = 7.76$, $p < .01$, $MSE = 5{,}447$. This difference suggests that subjects in the semantic maintenance condition were unlikely to have been rehearsing phonologically, as the patterns would then be expected to be qualitatively similar. These findings support the notion that semantic and phonological maintenance are dissociable.

Probe and target accuracy. Probe accuracy was slightly higher in the phonological condition ($M = 96\%$, $SD = 6\%$) than in the semantic maintenance condition ($M = 95\%$, $SD = 7\%$), $F(1, 30) = 9.38$, $p < .01$, $MSE = 0.002$. Target judgments were more accurate in the semantic condition ($M = 88\%$, $SD = 13\%$) than in the phonological condition ($M = 78\%$, $SD = 14\%$), $F(1, 30) = 23.87$, $p < .01$, $MSE = 0.028$. Although it is unclear why this difference in target accuracy arose, variability in pronunciation of the target syllable may have increased errors in the rhyme task.

Experiment 1 sought to determine whether semantic maintenance could be dissociated from phonological maintenance in a way that demonstrated maintenance during the retention interval. Using the concurrent probe method, we found that when subjects maintained a word's meaning over a delay, they were slower to make a lexical decision about a related probe. This semantic relatedness effect did not occur when the probe occurred after the maintenance interval, suggesting that it is a true maintenance effect. It never occurred in the phonological rehearsal condition either during or after the maintenance interval, supporting the view that the maintenance effect is semantic. Taken together, these findings show that the semantic relatedness effect reflects subjects sustaining the meaning of the maintenance item during the delay.

The semantic relatedness effect fits well with a model of working memory in which maintenance is accomplished in part through the suppression of interfering information (Anderson & Green, 2001; Dagenbach & Carr, 1994; Jonides et al., 1997; Lustig & Hasher, 2001). One difficulty with this conclusion, however, is that subjects responded on both the lexical decision and target judgment tasks by pressing the same keys labeled "yes" and "no." The overlap in the response mappings might have momentarily confused subjects, leading to longer reaction times for related probes. Experiment 2 addressed this possibility.

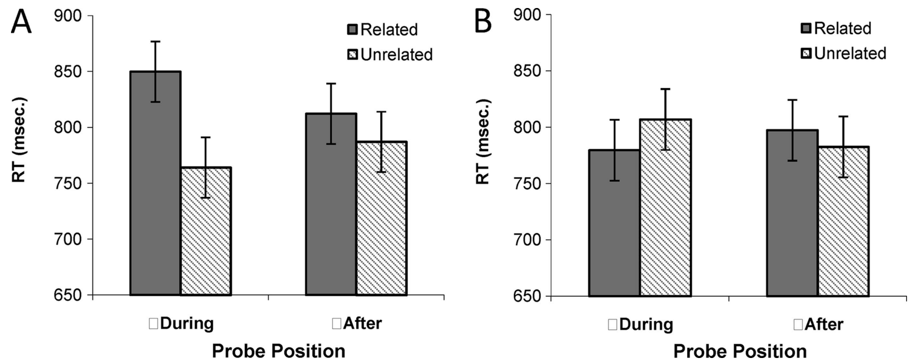

Figure 2. Reaction time to make a lexical decision to the probe item in the semantic (Panel A) and the phonological (Panel B) maintenance conditions, as a function of probe position (during or after the maintenance interval) and the relatedness of the probe to the maintenance item (related or unrelated) in Experiment 1, with 95% confidence intervals. Note that in the semantic maintenance condition, semantic relatedness led to increased reaction time (RT) during but not after the maintenance interval. Phonological maintenance did not produce this pattern.

Experiment 2

In Experiment 2, we sought to replicate the semantic relatedness effect using a procedure that separated the response modalities for the probe and target judgment tasks. In Experiment 1, subjects judged whether the target and the maintenance item were synonymous by pressing either the *yes* or the *no* key—the same keys they used to make responses during lexical decision trials. In the semantic condition, when subjects encountered a related probe, they might have been confused about whether they were pressing "yes" or "no" to make a lexical decision or a target judgment. This response confusion may have slowed reaction times for the related probe words.

To reduce response confusion, we had subjects respond to lexical decision items by pressing *yes* or *no* buttons but respond to target items by saying "same" or "different." Thus, the output modality and the labels for the responses were separated. If slower reaction times arose from the similarity between task responses, the semantic relatedness effect should not be observed.

Method

Subjects. Thirty-two undergraduate native English speakers took part to fulfill a course requirement.

Design, materials, and procedure. The design, stimuli, and procedures were those of Experiment 1, except that subjects responded to blue targets by saying "same" or "different" instead of by pressing a *yes* or a *no* key. Each target remained on the screen for 3 s while the experimenter recorded the response.

Results and Discussion

Probe reaction time. Replicating Experiment 1, subjects were slower to respond to related (M = 811, SD = 191) than to unrelated (M = 731, SD = 111) probes in the during condition, $F(1, 30) = 9.48$, $p < .01$, MSE = 10,746, but not in the after condition ($F < 1$), a difference evident in the interaction of probe relatedness with probe position, $F(1, 30) = 7.35$, $p = .01$, MSE = 7,064 (see Figure 3). Thus, an inhibitory semantic relatedness effect was observed that was delay period specific.

Also replicating Experiment 1, the phonological maintenance condition yielded no evidence for an inhibitory semantic relatedness effect. In the during condition, responses to related probes (M = 792 ms, SD = 140) were not reliably slower than those to unrelated probes (M = 772 ms, SD = 119, $F < 1$), and this pattern did not interact with probe position ($F < 1$). However, the three-way interaction of maintenance type, probe relatedness, and probe position did not reach significance, $F(1, 30) = 1.13$, $p = .29$, MSE = 8,660, as it did in Experiment 1, suggesting that the dissociation between phonological and semantic maintenance was not as robust.

To explore whether the results of Experiment 2 are reliably different from those of Experiment 1, we entered the probe data into an analysis with experiment as a factor. Collapsing across experiments yielded a significant semantic relatedness effect in the during position, $F(1, 60) = 25.15$, $p < .0001$, MSE = 8,709, but not in the after position ($F < 1$) of the semantic maintenance condition, with this interaction being highly significant, $F(1, 60) = 12.63$, $p < .001$, MSE = 6,303. In the phonological condition, no reliable relatedness effect was observed in either the during or the after condition, and there was no interaction between these positions ($F < 1$ in all cases). The overall interaction of probe position, probe relatedness, and maintenance type was significant, $F(1, 60) = 6.88$, $p < .05$, MSE = 6,480, and it did not interact across experiments ($F < 1$). Thus, although the semantic specificity of the semantic relatedness effect was not as robust in Experiment 2, it was not reliably different from that of Experiment 1. Nevertheless, the semantic relatedness effect is clearly present in the semantic condition and is delay period specific.

Probe and target accuracy. Unlike in Experiment 1, probe accuracy was higher in the semantic condition (M = 97%, SD = 6%) than in the phonological condition (M = 94%, SD = 7%), $F(1, 30) = 14.09$, $p < .01$, MSE = 0.002. It is interesting that subjects were less accurate at making lexical decisions for related than for unrelated probes, $F(1, 30) = 5.10$, $p < .05$, MSE = 0.003,

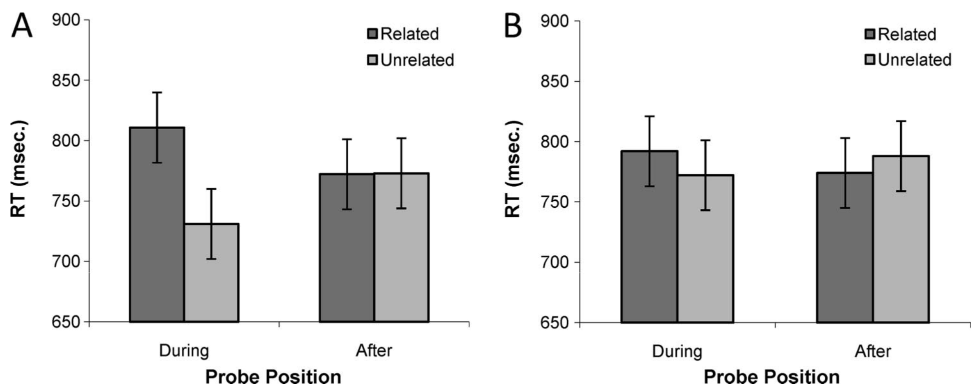

Figure 3. Reaction time to make a lexical decision to the probe item in the semantic (Panel A) and the phonological (Panel B) maintenance conditions, as a function of probe position (during or after the maintenance interval) and the relatedness of the probe to the maintenance item (related or unrelated) in Experiment 2, with 95% confidence intervals. Note that, as in Experiment 1, semantic relatedness led to increased reaction time (RT) during but not after the maintenance interval in the semantic maintenance condition. Phonological maintenance did not produce this pattern.

in the during condition but not in the after condition ($F < 1$), with this interaction being marginally significant, $F(1, 30) = 3.86$, $p = .059$, $MSE = 0.003$. This pattern shows that semantic maintenance can also impair accuracy for related probes.

In the phonological maintenance condition, subjects responded more accurately to related than to unrelated probes after maintenance, $F(1, 30) = 5.29$, $p < .05$, $MSE = 0.004$, but not during maintenance, $F(1, 30) = 1.36$, $p = .25$, $MSE = 0.006$. This produced an interaction of probe position and probe relatedness, $F(1, 30) = 8.47$, $p < .01$, $MSE = 0.003$. Unlike in the semantic condition, there was no evidence for impaired accuracy for related probes in any probe position. For target judgments, subjects were marginally more accurate in the semantic condition ($M = 88\%$, $SD = 15\%$) than in the phonological condition ($M = 83\%$, $SD = 13\%$), $F(1, 30) = 4.73$, $p = .06$, $MSE = 0.053$.

Overall, the results of Experiment 2 confirm those of Experiment 1: When subjects maintain the meaning of a word, they are slower (and also less accurate) to respond to semantically related probes during the maintenance interval but not after it. Phonological maintenance, however, does not cause a semantic relatedness effect either during or after the maintenance interval, suggesting that phonological rehearsal did not cause this effect. These findings further show that confusion between response mappings did not influence the pattern of results.

Experiment 3

Experiments 1 and 2 demonstrate that semantic and phonological maintenance can be functionally dissociated, consistent with a distinct capacity for semantic working memory. If a distinct semantic maintenance capacity exists, occupying phonological rehearsal with irrelevant information should leave semantic maintenance performance intact. In Experiment 3 we tested this hypothesis by asking subjects to perform articulatory suppression during semantic maintenance. In articulatory suppression, subjects repeat a single word out loud continuously while doing some primary task (e.g., memory span) to reduce the contributions of phonological rehearsal to that task. Articulatory suppression greatly reduces the effects on working memory of variables taken as markers of phonological loop involvement, such as effects of word length and phonological similarity with visual presentation (Baddeley et al., 1975; Murray, 1968). If semantic maintenance can be done without the phonological loop, a semantic relatedness effect should still occur in our delayed judgment task even when people perform a concurrent articulatory suppression task.

A second goal was to explore the role of inhibitory control in producing the semantic relatedness effect. As noted in the introduction, we view semantic maintenance as requiring sustained attention to a concept, despite our mind's tendency to wander to related items that grow activated in the process. This view suggests that the involvement of inhibition may grow more apparent as time unfolds during the maintenance interval of the delayed judgment task. In particular, one should expect an initial wave of activation to prime associated concepts, followed by gradual suppression as inhibition grows more necessary to maintain focus. Analogous findings have been observed in other designs sensitive to inhibition. In experiments on long-term memory retrieval, for example, Shivde and Anderson (2001) and Johnson and Anderson (2004) found that a single retrieval practice on the subordinate meaning of an ambiguous word facilitated later recall of the dominant meaning, but that additional retrieval practices gradually suppressed that facilitation. Similar findings have been observed in other paradigms (e.g., Blaxton & Neely, 1983; see also Kuhl & Anderson, 2011). We were interested to see whether such a nonmonotonic inhibition function may be operating during the semantic maintenance interval of our task.

To look for a nonmonotonic inhibitory function, we included probes at two delays during maintenance and also two delays after maintenance. The probe occurred six items after the maintenance item in the short delay trials and 10 items after the maintenance item in the long trials. As in prior experiments, the delays between the maintenance and probe items in the after and during conditions were matched. If there is a nonmonotonic function, a facilitatory relatedness effect may occur in the short delay condition, which turns into an inhibitory relatedness effect at long delays. If this pattern arises from active maintenance, it should occur during but not after the maintenance interval.

Method

Subjects. Forty undergraduate native English speakers were each paid $7.00 for participating.

Design. Probe relatedness (related vs. unrelated) and probe position (during-short, during-long, after-short, after-long) were manipulated within subjects using the semantic condition from the previous experiments. We matched the number of events between the maintenance and probe items for the during-short and after-short conditions and also for the during-long and after-long conditions (see Figure 4). We measured subjects' reaction time to make lexical decisions to probes.

Materials. The stimuli were those from the semantic maintenance condition of Experiments 1 and 2 except that only 688 of the lexical decision filler words and nonwords were necessary.

Procedure. We used the semantic maintenance instructions of Experiment 1, except we added an articulatory suppression task. We asked subjects to repeat the word *blank* out loud continuously while doing the maintenance and lexical decision tasks. The experimenter monitored subjects to ensure they complied with the articulation instructions. The procedure matched Experiment 1, except that when subjects made lexical decisions, the next stimulus appeared automatically instead of after a fixed delay; if subjects did not respond, the next stimulus appeared after 2 s. Practice with articulation was given in the maintenance task.

Results and Discussion

We analyzed all data in a 2 (probe relatedness) × 4 (probe position) mixed analysis of variance that included presentation order as a between-subjects factor.

Probe reaction time. The pattern of semantic relatedness effects replicates those observed in Experiments 1 and 2, despite longer overall probe reaction times due to articulatory suppression. The delays of Experiments 1–2 are most closely approximated by the long-lag condition (in fact, the long-delay condition is 4–5 s shorter than previous delays). In this long-lag condition, subjects responded more slowly to related (M = 881, SD = 245) than to unrelated (M = 819, SD = 176) probes during maintenance, $F(1, 39) = 4.05$, $p = .05$, MSE = 19,168, but not after it ($F < 1$; see Figure 5). This pattern suggests that the semantic relatedness effect was maintenance specific. However, this difference across the during-long and after-long conditions (the interaction of probe relatedness with during-long vs. after-long) did not reach significance, $F(1, 39) = 1.83$, $p = .18$, MSE = 16,303. Nevertheless, the similarity of these findings to those in prior experiments and the absence of relatedness effects in all previous (and forthcoming) "after" conditions suggest that these relatedness effects are caused by maintenance. If so, the fact that the relatedness effect occurred during the delay suggests that semantic maintenance can occur during articulatory suppression.

Our second aim was to explore the development of the semantic relatedness effect across the maintenance interval. If associated items are primed initially and then gradually suppressed, more inhibition should occur at the long than at the short delay, with the possibility of positive priming at the short delay. Although at the short delay there was a numerical difference suggesting positive priming, this was not significant ($F < 1$). Nevertheless the semantic relatedness effect significantly grew over the during-short and during-long conditions, $F(1, 39) = 4.33$, $p < .05$, MSE = 111,955, indicating that these effects build up with time, consistent with a gradually emerging inhibition. Inhibition may build up because activation spreading from the maintenance item ultimately interferes with maintenance by heightening the activation of related competitors. Alternately, the unrelated words in the lexical decision task may add interference that would build with the number of intervening trials. Because these two possibilities are confounded here, further research is necessary to disentangle them.

In Experiment 3, we took a different approach to dissociating semantic and phonological working memory. If semantic working memory has its own capacity, it should be possible to maintain semantic information while the phonological loop is occupied. Consistent with this prediction, subjects exhibited the semantic relatedness effect during semantic maintenance even though they were simultaneously performing articulatory suppression. The relatedness effect does not occur after the maintenance interval has ended, suggesting it arises from active maintenance. Although this effect was numerically smaller (62 ms) than in Experiments 1 and 2 (86 and 80 ms, respectively), this slight reduction likely arose from the demanding triple task, the coordination of which may have diminished attention to the maintenance task. Alternatively, it may have arisen because the current "long delay" probes appeared earlier in the maintenance interval than they did in previous experiments, reducing the time for inhibition to build. Regardless of this slight reduction, the observation of the semantic relatedness effect during articulatory suppression converges with the results of Experiments 1 and 2 in suggesting that semantic maintenance is not accomplished by the phonological loop.

	During		After	
Event	Short	Long	Short	Long
1	**ANGRY**	**ANGRY**	**ANGRY**	**ANGRY**
2	shoe	shoe	shoe	shoe
3	water	water	water	water
4	anchor	anchor	**MAD**	anchor
5	lammer	lammer	anchor	lammer
6	bise	bise	lammer	**MAD**
7	yell	gold	yell	bise
8	gold	rinse	bise	gold
9	rinse	gask	gold	rinse
10	**MAD**	rust	rinse	gask
11	gask	yell	gask	yell
12	rust	plim	rust	rust
13	plim	cup	plim	plim
14	cup	**MAD**	cup	cup
15	bool	bool	bool	bool

Figure 4. Modification of the delayed judgment paradigm to study the effects of probe delay (short or long) on the semantic relatedness effect. The event sequence is similar to that used in other experiments, except that the probe (in this example, the word *yell*) appeared either at a shorter delay after the maintenance item (the left half within the during and after conditions) or at a longer delay (the right half within the during and after conditions). Note that while subjects performed this task, they also performed articulatory suppression continuously.

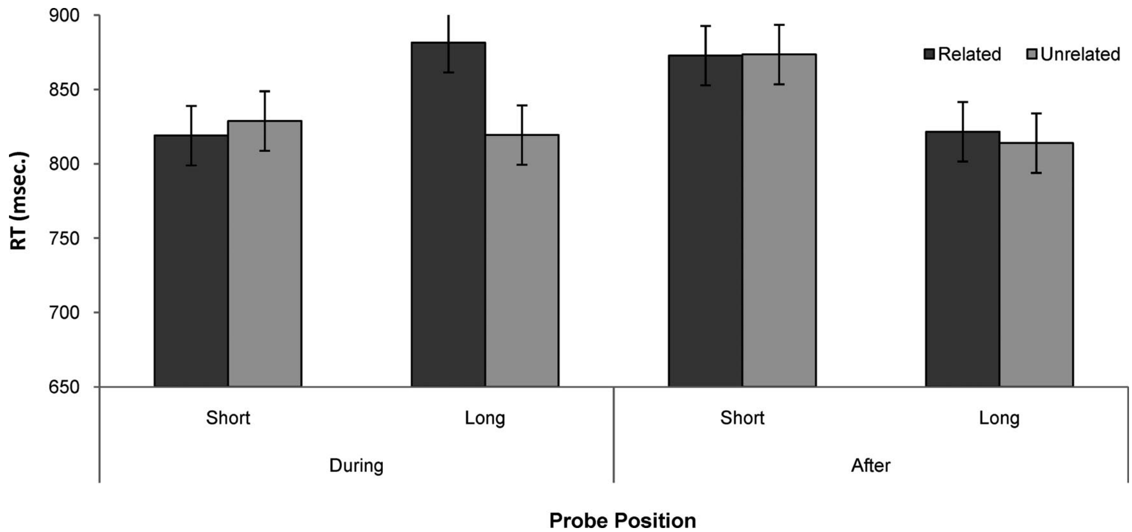

Figure 5. Reaction time (RT) to make a lexical decision to the probe item during semantic maintenance as a function of probe position (during or after the maintenance interval) and the probe delay (short or long, relative to the presentation of the maintenance item) in Experiment 3, with 95% confidence intervals. Note that the semantic relatedness effect (increased reaction time to respond to related vs. unrelated probe items) takes time to develop during the maintenance interval, with a greater relatedness effect in the long than in the short probe condition. After the maintenance interval has ended, no relatedness effects are observed, as in prior experiments.

Experiment 4

Experiments 1–3 indicate that semantic and phonological maintenance can be dissociated. These findings are compatible with a separable semantic working memory capacity that may be recruited independently of other working memory subsystems. However, we need to consider whether the visuospatial sketchpad was used to keep the semantic item in mind. Although we designed maintenance items to be abstract and difficult to image, subjects may have visualized an associated concept to maintain the target's "meaning" (e.g., imagining scales for "justice"). Alternatively, subjects may have kept the visual word form in mind (e.g., imagining the letters *B-U-Y* for the word *buy*). Either of these types of maintenance could have been done in parallel with articulatory suppression, and they may have produced the semantic relatedness effect.

In Experiment 4, we asked subjects to perform visual word-form maintenance to see whether it produces a semantic relatedness effect. Subjects performed semantic maintenance in one block of trials and visual word-form maintenance in another. For the target task in the visual condition, synonym judgment was replaced with a font judgment task in which subjects had to judge whether the target word was presented in the same font as the maintenance word. Using font judgment as the target task emphasized the importance of maintaining the exact visual appearance of the item and provided behavioral evidence of subjects' memory for this information. If visual word-form maintenance contributes to the semantic relatedness effect, we should observe the effect in both the semantic and visual maintenance conditions. The potential role of imagery for associated concepts is addressed in the Discussion of Experiment 5.

Method

Subjects. Thirty-two undergraduates native English speakers took part to fulfill a course requirement.

Design. The design of Experiment 3 matched that of Experiments 1 and 2 except that the phonological maintenance condition was replaced by visual maintenance.

Materials. The stimuli from Experiments 1–2 were used in the semantic and visual maintenance conditions. In the visual maintenance condition, each item appeared in one of eight fonts. The fonts were chosen to be easily readable and not distinguishable by a single visual feature. To assure that the semantic and visual maintenance presentations were closely matched, we also randomly assigned the maintained and target items in the semantic condition to appear in these fonts. Four additional fonts were used for filler trials. All the lexical decision items appeared in a single font that was different from the maintenance and target item fonts.

In the visual condition, target items were the same words as the maintenance items but appeared in either the same or a different font. Probe words were those used in Experiment 1 and were either semantically related or unrelated to the maintenance item.

Procedure. The procedure for the semantic maintenance condition was that of Experiment 2. In the visual maintenance condition, subjects were asked to keep an exact image of the red maintenance word in mind throughout the delay. For example, if the red word had been *angry* they were asked to picture the letters *A-N-G-R-Y* just as these appeared on the screen. They were told that they would see the same word in blue and that they would have to decide whether it was printed in the same font. Instructions for the lexical decision task matched those of Experiments 1 and 2. Subjects made lexical decisions by pressing *yes* and *no* keys and

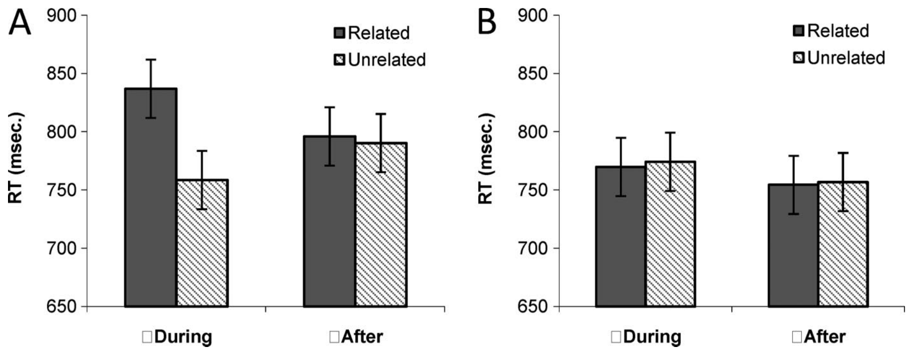

Figure 6. Reaction time (RT) to make a lexical decision to the probe item in the semantic (Panel A) and the visual (Panel B) maintenance conditions, as a function of probe position (during or after the maintenance interval) and the relatedness of the probe to the maintenance item (related or unrelated) in Experiment 4, with 95% confidence intervals. Note that in the semantic maintenance condition, semantic relatedness led to increased reaction time during but not after the maintenance interval. Visual maintenance did not produce this pattern.

responded to the target by saying "same" or "different" aloud. The pace of the events was fixed, as in Experiments 1–2.

Results and Discussion

Experiment 4 was analyzed with the same general design as were Experiments 1–2. On the target task, subjects were more accurate in the semantic condition ($M = 89\%$, $SD = 13\%$) than in the visual condition ($M = 84\%$, $SD = 14\%$), $F(1, 30) = 4.21$, $p = .05$, $MSE = 0.032$. This showed that visual maintenance was difficult.

Probe reaction time. The semantic condition replicated Experiments 1–3: Subjects responded more slowly to related probes ($M = 837$, $SD = 142$) than to unrelated probes ($M = 759$, $SD = 116$) when they occurred during maintenance, $F(1, 30) = 26.7$, $p < .001$, $MSE = 3{,}694$, but not after maintenance ($F < 1$), and this interaction was significant, $F(1, 30) = 10.29$, $p < .01$, $MSE = 4{,}121$ (see Figure 6). These findings again demonstrate a delay-period-specific semantic relatedness effect.

During visual word-form maintenance, however, there were no reliable semantic relatedness effects in either the during or the after conditions ($F < 1$ in both cases), and there was no interaction of probe relatedness and probe position ($F < 1$). The semantic relatedness effect was greater in the semantic than in the visual maintenance condition during the delay, $F(1, 30) = 19.5$, $p < .001$, $MSE = 2{,}823$, and the interaction of maintenance strategy, probe relatedness, and probe position was significant, $F(1, 30) = 8.28$, $p < .01$, $MSE = 2{,}699$. This shows that the semantic relatedness effect is specific not only to the semantic condition but also to the period of active maintenance.

Experiments 1–3 demonstrated that semantic relatedness effects arising during maintenance do not occur when phonological rehearsal is used. Experiment 4 provides an important control by further showing that subjects are not using their visuospatial sketchpad to keep an image of the visual word form in mind during the semantic maintenance interval—or at least, if they are, it is not contributing to the semantic relatedness effect. These findings also clearly replicate the maintenance-specific semantic relatedness effect observed in Experiments 1–3.

Experiment 5

In Experiments 1–4, we examined whether semantic representations could be maintained in a state of heightened accessibility. We have argued that this ability is supported by a controllable attention process. If semantic maintenance is controllable, subjects should be able to "turn off" this mechanism and rely instead on long-term episodic memory to perform target judgments. Indeed, the capacity to take either a working memory or a long-term memory approach to semantic retention is at the heart of the theoretical ambiguity that our concurrent probe method addresses.

In Experiment 5, we tested whether semantic maintenance is controllable. To do this, we introduced a long-term memory condition that was structurally identical to semantic maintenance but with new instructions. We asked subjects to encode the maintenance item so that they could remember it when the blue target appeared. Subjects were then asked to focus their full attention on the lexical decision task during the delay. Later, when the blue word appeared, they were asked to think back to the red word and retrieve it to make their synonym judgment. Performance on this task was contrasted to that in the semantic working memory task. If the semantic relatedness effect reflects controlled maintenance, subjects should be able to turn off maintenance in the long-term memory condition; if so, the relatedness effect should only occur in the working memory condition. If semantic working memory affords a functional advantage, we should find better performance on the target judgment in the working memory condition.

Method

Subjects. Thirty-two undergraduate native English speakers took part to fulfill a course requirement.

Design. The design replaced the phonological condition of Experiments 1–2 with a long-term memory condition. Both conditions required subjects to make a synonym judgment on the target.

Materials. The stimuli were the same as those used in Experiment 1. All 80 synonym pairs were used and were fully rotated through the working memory and long-term memory conditions.

Procedure. We told subjects that we were interested in the effects of doing two things at once versus doing them one after the other. The working memory and long-term memory tasks were identical except for instructions on what to do with the red word. In the semantic maintenance condition, subjects received the standard semantic maintenance instructions asking them to actively maintain the idea of the word throughout the delay. In the long-term memory condition, however, subjects were asked to encode the red word so that they could remember it after the delay. After the delay, they would have to recall the red word and decide whether it was synonymous with the blue target. As in prior experiments they were given practice doing just this memory task alone. They were then told about the lexical decision task. Subjects in the long-term memory condition were asked to focus completely on the lexical decision task after the red word disappeared and before the blue word appeared. They received practice doing the memory task with lexical decision. The order of presentation of the working memory and long-term memory conditions was counterbalanced across subjects.

Results and Discussion

The same analysis approach was taken in the current study as in prior ones. Unlike in previous experiments, presentation order influenced the semantic relatedness effect in the working memory condition. Thus, we discuss the data broken out by presentation order.

Probe reaction time. The pattern of probe reaction times differs according to whether subjects did the working memory task first (see Figure 7). When the working memory task came first, we observed the predicted interaction of probe relatedness and probe position within the working memory data, $F(1, 30) = 14.5$, $p < .001$, $MSE = 6{,}190$, with a semantic relatedness effect in the during condition, $F(1, 30) = 16.95$, $p < .001$, $MSE = 7{,}334$, but not in the after condition ($F < 1$). We observed no such interaction in the long-term memory condition ($F < 1$), because there was no semantic relatedness effect in either the during or the after condition ($F < 1$ in each case). Indeed, the three-way interaction of maintenance strategy, probe position, and probe relatedness was significant, $F(1, 30) = 4.9$, $p < .05$, $MSE = 8{,}903$. Thus, the specificity of the semantic relatedness effect to the maintenance interval was stronger in the working memory condition than in the long-term memory condition. In the during condition, the effect of probe relatedness varied by maintenance type, $F(1, 30) = 11.5$, $p < .01$, $MSE = 6{,}834$, showing a greater relatedness effect during active maintenance than when subjects simply encoded the item for later memory retrieval.

When the long-term memory condition came first, however, there was no interaction of maintenance type, probe relatedness, and probe position, $F(1, 30) = 1.09$, $p = .31$, $MSE = 9{,}677$. Indeed, this critical three-way interaction varied across presentation orders, as seen in the four-way interaction of order with these factors, $F(1, 30) = 5.3$, $p < .05$, $MSE = 8{,}903$. This interaction reflects a difference in the relatedness effect during maintenance for the working memory condition across the orders, $F(1, 30) = 7.16$, $p < .05$, $MSE = 7{,}334$, with an effect when working memory came first (see previous paragraph) but none when it came second ($F < 1$). Thus, doing the long-term memory condition first eliminated the relatedness effect in the later working memory condition.

The reduced semantic relatedness effect when the long-term memory task came first may reflect carryover of the long-term memory strategy to working memory trials. Unlike in the prior studies, subjects did the same synonym judgment target task in both conditions, with the only difference being whether subjects were asked to maintain the meaning of the red word continuously or to remember it later only when needed. Because doing the maintenance task and lexical decision concurrently is not easy, some subjects may have adopted a long-term memory strategy during working memory trials. This may have been exacerbated when the long-term memory task came first, emphasizing the strategy's effectiveness. Carrying over this long-term memory strategy might have reduced the relatedness effect for the later working memory task.

Because of this presentation order interaction, collapsing over this factor yields weaker differences in relatedness effects between working memory and long-term memory than might be expected. The predicted interaction of probe relatedness with probe position in the working memory condition, for example, was only marginally significant, $F(1, 30) = 3.99$, $p = .055$, $MSE = 6{,}190$, although the relatedness effect in the during condition was highly significant, $F(1, 30) = 9.91$, $p < .01$, $MSE = 7{,}334$, whereas it was not in the after condition ($F < 1$). The relatedness effect in the long-term memory condition was significant neither for the during condition nor for the after condition ($F < 1$ in both cases; see Figure 7). Although the three-way interaction of maintenance strategy, probe relatedness, and probe position was not significant in the overall data ($F < 1$), the crucial semantic relatedness effect during the delay interval was greater in the working memory condition than in the long-term memory condition, $F(1, 30) = 4.62$, $p < .05$, $MSE = 6{,}834$. Thus, even if the diluting effects of presentation order is considered, the relatedness effect was greater during the maintenance interval in the working memory condition than in the long-term memory condition, as predicted.

Probe and target accuracy. There were no reliable effects on probe accuracy. However, subjects made more accurate target judgments in the working memory condition ($M = 84\%$, $SD = 16\%$) than in the long-term memory condition ($M = 76\%$, $SD = 21\%$), $F(1, 30) = 5.78$, $p < .05$, $MSE = 0.062$. Because the judgments made were the same in both cases, this difference suggests that working memory was effective at keeping the maintenance item available despite interference from the lexical decision task. In the working memory condition, target judgment accuracy was unaffected by whether the preceding probe was related ($F < 1$). In the long-term memory condition, however, there was a trend toward lower target accuracy when a related probe appeared during the delay ($M = 72\%$, $SD = 22\%$) than when an unrelated probe appeared ($M = 76\%$, $SD = 20\%$), $F(1, 30) = 2.54$, $p = .12$, $MSE = 0.01$. The trend showed that related probes interfere with target judgments. This supports the idea that maintaining an item reduces interference from related items that might disrupt its accessibility.

The findings of Experiment 5 add confidence that the semantic relatedness effect is due to active maintenance. In previous experiments, this conclusion rested on the finding that the relatedness effect occurred during the maintenance interval but not after it. However, these conditions differed in that subjects responded to a target before seeing the probe in the after but not in the during condition. Experiment 5 avoided any problems this creates by comparing two during conditions, one in which subjects maintain an idea over a delay and another in which they use long-term memory. The results show that even when subjects do precisely the same target task and lexical decisions, the relatedness effect occurs only when they use working memory. Thus, the semantic relatedness effect can serve as a delay-period marker of semantic maintenance. In the next section, we consider subjects' retrospective reports of imagery in the semantic maintenance condition for all five experiments and how these reports relate to the semantic relatedness effect.

Combined Analysis of Experiments 1–5 by Self-Rated Imagery

In Experiments 1–5, we used abstract words to minimize involvement of visuospatial working memory to semantic maintenance. To verify that this control reduced imagery-based maintenance, we asked subjects to rate how often they formed an image of the word's referent or an associated concept to do semantic maintenance. Subjects rated their use of imagery on a 5-point scale (1 = *never,* 3 = *sometimes,* 5 = *always*). Aggregating across all five experiments (N = 168), this scale revealed that some subjects did use imagery to achieve semantic maintenance, although infre-

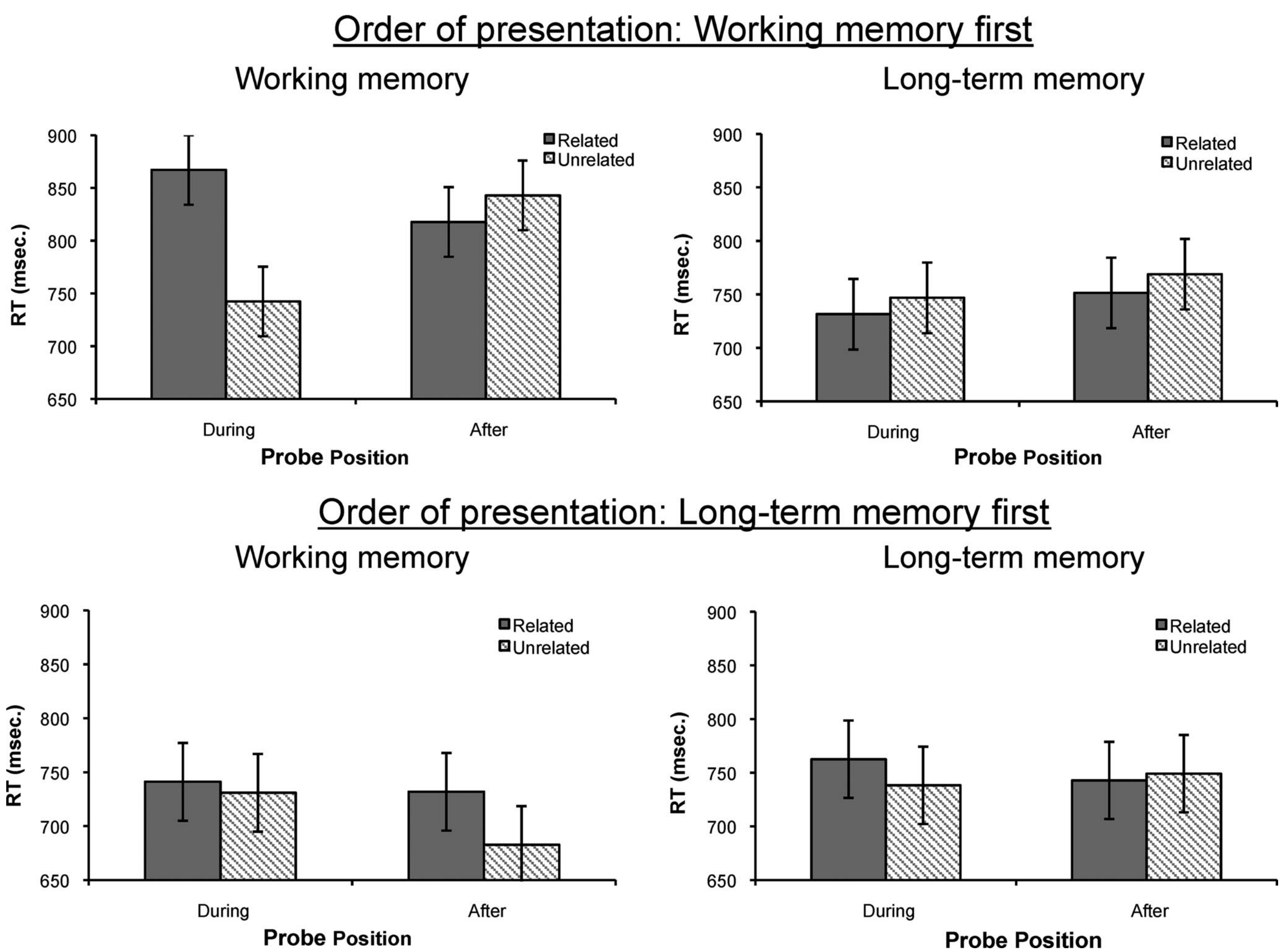

Figure 7. Reaction time (RT) to make a lexical decision to the probe item in the semantic working memory condition (Panel A) and the long-term memory condition (Panel B), as a function of probe position (during or after the maintenance interval) and the relatedness of the probe to the maintenance item (related or unrelated) in Experiment 5, with 95% confidence intervals. The data are broken out by whether subjects performed the semantic working memory condition followed by the long-term memory condition (top half of figure) or performed them in the reverse order (bottom half of the figure). Note that when subjects did the semantic working memory condition first, semantic relatedness led to increased reaction time during the maintenance interval but not after it in the semantic maintenance condition, replicating the prior experiments. Encoding into long-term memory, however, did not produce this pattern. When the long-term memory task was performed first, however, no semantic relatedness effect was observed in either condition, suggesting that subjects did not use semantic working memory when it was clear that long-term memory would suffice for the task.

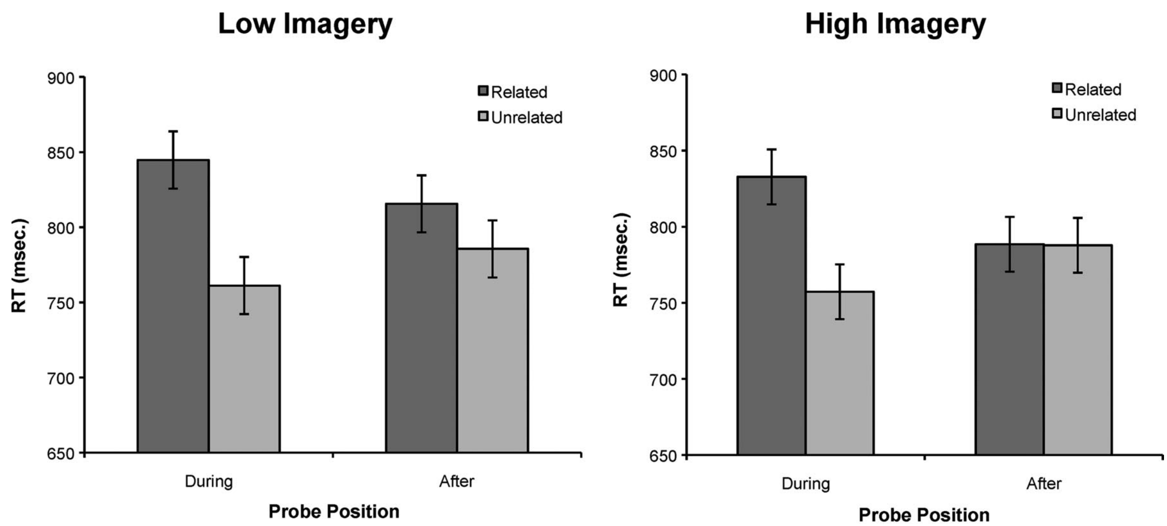

Figure 8. A meta-analysis of the semantic maintenance conditions across Experiments 1–5, with subjects divided (by median split) into those who reported a higher degree of visual imagery during the maintenance trial (right panel) and those who reported a lower amount (left panel). Each panel depicts the reaction time (RT) to make a lexical decision to the probe item as a function of probe position (during or after the maintenance interval) and the relatedness of the probe to the maintenance item (related or unrelated), with 95% confidence intervals. As expected, the semantic relatedness led to increased reaction time during the maintenance interval but not after it. The size of this semantic relatedness effect did not vary across the low- and high-imagery groups.

quently ($M = 2.46$, $SD = 1.3$). This rate was consistent across experiments (2.23 to 2.85). Thus, we succeeded in limiting imagery, though this strategy was not eliminated entirely.

Although subjects did not use imagery frequently, the semantic relatedness effect could nevertheless be related to it. To explore this possibility, we examined whether imagery predicts the size of the semantic relatedness effect by combining the probe reaction time data for the semantic conditions from all five experiments. Within each experiment, we performed a median split of subjects based on their imagery ratings, matching for item counterbalancing. Then, we aggregated the low-imagery groups (84 subjects, mean rating = 1.62, $SD = 0.1$) and the high-imagery groups (84 subjects, mean rating = 3.31, $SD = 0.1$) from all experiments, allowing for powerful statistical tests.

There was no evidence that the semantic relatedness effect was modulated by the reported use of imagery (see Figure 8). The relatedness effect during the maintenance interval was not reliably different across the groups ($F < 1$), nor was the three-way interaction of probe relatedness, probe position, and strategy use ($F < 1$). Furthermore, the relatedness effect in the during condition was highly reliable in both the low-imagery group, $F(1, 166) = 24.65$, $p < .0001$, $MSE = 11{,}902$, and the high-imagery group, $F(1, 166) = 20.13$, $p < .0001$, $MSE = 11{,}902$. If anything, the relatedness effect appears to be numerically smaller in the high-imagery group.

One might be concerned that the relatedness effect in the low-imagery group may be caused by the minor amount of imagery reported. To address this possibility, we focused on only those subjects who reported never using imagery (rating of 1, $N = 58$). These subjects showed a substantial semantic relatedness effect (89 ms). These results argue that the semantic relatedness effect is not due to subjects' infrequent use of imagery. Thus, these findings suggest that semantic maintenance was not performed by visuospatial working memory.

Experiment 6

Experiments 1–5 established a semantic relatedness effect that is specifically induced by semantic and not phonological or visual maintenance, consistent with the direct semantic maintenance hypothesis. One might argue, however, that subjects were not directly maintaining the semantics of a word but rather were engaging in mediated maintenance. Mediated maintenance might be accomplished by repeating the maintenance word phonologically but secondarily attending to conceptual information elicited by this process. By this account, phonological maintenance would not induce this effect because the task does not require attention to meaning. Because it is easier to focus on one dimension of the maintenance word than two, subjects opt to simply maintain the word's phonology.

Experiment 3 provided evidence against mediated maintenance by showing that the semantic relatedness effect survives articulatory suppression. However, although articulatory suppression ought to disrupt maintenance in the phonological loop, this disruption may be partial, particularly because only a single item must be maintained. Thus, converging evidence against mediated maintenance is desirable. Another approach is to seek a distinctive delay-period marker of phonological maintenance, which could indirectly indicate whether people are indeed attending to phonology when asked to rehearse a word's sound. If such a marker could be established in the phonological maintenance condition, it could be used to examine whether hidden phonological maintenance is taking place in our semantic maintenance condition.

In Experiment 6 we sought evidence for behavioral markers of phonological maintenance using the concurrent probe method. As in our earlier experiments, we inserted critical probes in the lexical decision stream either during or after maintenance. Unlike in prior studies, however, we manipulated whether these probes were related to the maintenance item on the basis of phonology (rhyming or nonrhyming) instead of semantics. If subjects actively maintained the phonological maintenance item, attention to it might influence phonologically similar probes during the delay. As such, we sought evidence for a phonological relatedness effect, measured as a reaction time difference to rhyming (as compared to nonrhyming) probes. A phonological relatedness effect during but not after phonological maintenance would constitute evidence for a delay-period marker of active maintenance. Given this marker, we can examine whether phonological relatedness effects occur during semantic maintenance. If so, it would support the mediated maintenance hypothesis. If not, it would establish (when combined with Experiments 1 and 2) a double dissociation of semantic and phonological maintenance with respect to semantic and phonological relatedness effects, providing strong support for direct semantic maintenance.

Method

Subjects. Thirty-two undergraduate native English speakers took part to fulfill a course requirement.

Design. The design was the same as that in Experiments 1 and 2, except that the semantically related probe words were replaced by phonologically related probe words as described below.

Materials. All stimuli were the same as those used in Experiments 1 and 2, with the exception of the probes. We chose real words that rhymed with the 40 original maintenance words as phonologically related probes. Care was taken to use rhyming probes that were orthographically dissimilar to the maintenance item (e.g., maintenance item *TALE,* probe *SNAIL*). These probes were used in both the semantic and phonological maintenance conditions. Unrelated probes were created by re-pairing related probes among half of the trials for each subject.

Procedure. Procedures matched those in Experiments 1 and 2.

Results and Discussion

The present analysis was structured similarly to previous ones. There was no main effect of presentation order, $F(1, 30) = 0.69$, $p = .41$, $MSE = 6{,}4178$, and presentation order did not interact with our manipulations ($p > .1$ in all cases), with one exception discussed below.

Probe reaction time. Subjects responded more slowly to phonologically related probes ($M = 816$, $SD = 150$) than to unrelated probes ($M = 749$, $SD = 136$) during phonological maintenance, $F(1, 30) = 15.09$, $p < .01$, $MSE = 4{,}780$, but not after it, $F(1, 30) = 0.01$, $p > .5$, $MSE = 4{,}863$. The interaction between probe relatedness and probe position was significant, $F(1, 30) = 9.48$, $p < .01$, $MSE = 3{,}640$, establishing a phonological relatedness effect specific to active maintenance.

Because the phonological relatedness effect occurs during phonological maintenance, it can serve as a marker for whether people were maintaining the maintenance item phonologically during our semantic condition. During semantic maintenance, there was no significant difference in phonologically related ($M = 732$, $SD = 94$) and unrelated ($M = 740$, $SD = 110$) probe reaction times, $F(1, 30) = 0.23$, $p > .5$, $MSE = 3{,}805$. After semantic maintenance, there was no significant difference in these probe reaction times, $F(1, 30) = 1.46$, $p = .24$, $MSE = 4{,}002$, and there was no interaction between probe relatedness and probe position, $F(1, 30) = 1.31$, $p = .26$, $MSE = 5{,}317$. Thus, the phonological relatedness effect occurs only during active maintenance of phonological information and not during semantic maintenance (see Figure 9).

Probe and target accuracies. Probe accuracies differed between the semantic condition ($M = 87\%$, $SD = 12\%$) and pho-

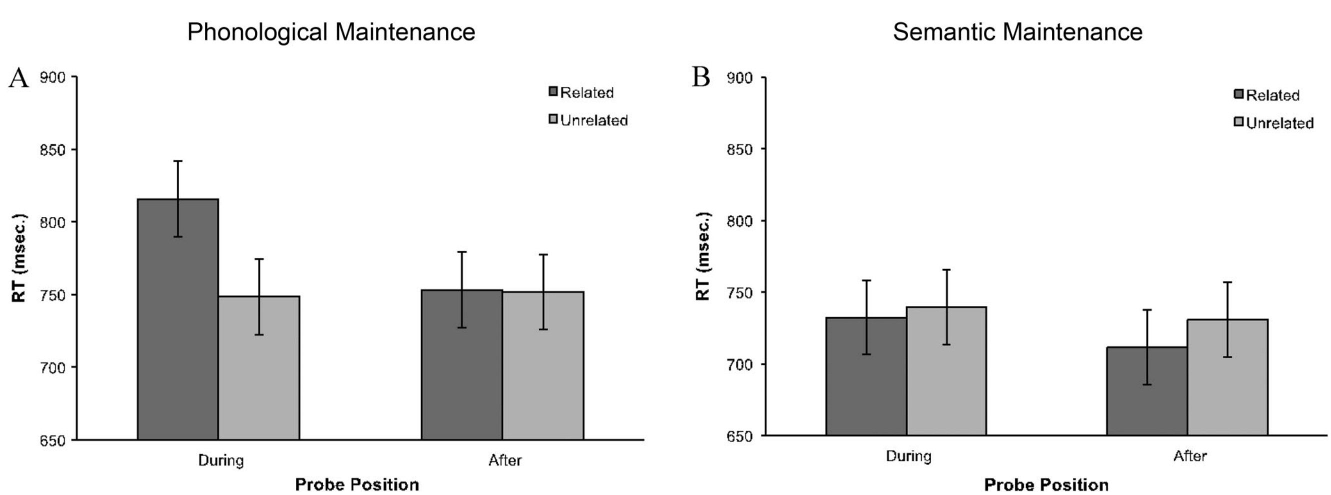

Figure 9. Reaction time (RT) to make a lexical decision to the phonologically related probe item in the phonological (Panel A) and the semantic (Panel B) maintenance conditions, as a function of probe position (during or after the maintenance interval) and the relatedness of the probe to the maintenance item (related or unrelated) in Experiment 6, with 95% confidence intervals.

nological condition (M = 90%, SD = 11%), $F(1, 30) = 4.40$, p = .04, MSE = 0.007. There were no main effects of probe relatedness, probe position, or presentation order ($p > .49$ in all cases). Further analyses showed that probe accuracy varied with maintenance type (semantic/phonological) for related probes, $F(1, 30) = 4.87$, $p = .04$, $MSE = 0.005$, but not for unrelated probes, $F(1, 30) = 1.35$, $p = .25$, $MSE = 0.007$. Probe relatedness did not interact with probe position within either the phonological condition, $F(1, 31) = 1.89$, $p = .20$, or the semantic condition, $F(1, 31) = 1.44$, $p = .24$ (see Appendix E).

Target accuracies were higher in the semantic condition (M = 89%, SD = 12%) than in the phonological condition (M = 80%, SD = 15%), $F(1, 30) = 18.59$, $p < .01$, $MSE = .03$. There were no significant effects of probe position and probe relatedness ($p >$ .5 in both cases).

With Experiments 1 and 2, the present findings demonstrate a full double dissociation between semantic and phonological maintenance. Whereas semantic maintenance causes a semantic relatedness effect, phonological rehearsal does not; phonological maintenance, by contrast, causes a phonological relatedness effect, whereas semantic maintenance does not. This double dissociation strongly supports the existence of a distinct direct semantic maintenance as a process. This experiment further demonstrates that the concurrent probe method can be useful in finding behavioral markers of active maintenance beyond semantic working memory.

General Discussion

The present experiments demonstrate that retaining the meaning of an item over a prolonged delay causes measurable changes in cognitive state that appear to be both delay period specific and semantically specific. The evidence for a change in cognitive state took the form of altered response times to lexical decision probes during the delay period. In five experiments, sustaining a word's meaning over a prolonged delay altered lexical decision times to an associated probe appearing during that delay compared to unrelated probe items. This semantic relatedness effect confirms that subjects semantically processed the maintenance item and that this semantic processing may be detected on an apparently unrelated lexical decision task. This effect was observed in reaction time, although a deficit in accuracy was found in Experiment 2.

It is important that this semantic relatedness effect occurred only during the maintenance period and thus demonstrated delay-period specificity. When the same lexical decision probes appeared after maintenance, related and unrelated probe reaction times did not differ. This delay-period specificity occurred in all five experiments with semantic probes, even though we matched the lag between encoding the maintenance item and the probe across the during and after conditions. The lack of relatedness effects after maintenance also occurred regardless of whether the target synonym judgment (which came before the probe in the after condition) was related or unrelated to the maintained item. The delay-period specificity of the semantic relatedness effect indicates that this effect is related to cognitive operations occurring during the delay interval. Whatever underlies the semantic relatedness effect dissipates rapidly after subjects judge the target, indicating that this effect can serve as a delay-period marker of active maintenance.

The semantic relatedness effect was specific not only to the delay period but also to the requirement to maintain semantics over that delay. Neither phonological maintenance (Experiments 1–2) nor visual word-form maintenance (Experiment 4) produced this effect, even though the very same subjects showed it robustly during semantic maintenance. The lack of semantic relatedness effects in these conditions is striking, given that they differed mainly in subjects' endogenous orientation to semantic, phonological, or visual information during maintenance. Moreover, the effects are unrelated to visual imagery, insofar as our abstract words discouraged this; indeed, the relatedness effect was not modulated by the (infrequently) reported use of imagery and occurred even in subjects who claimed never to use this strategy.

Evidence for this semantic specificity is particularly striking in the case of the dissociation from phonological maintenance. One might have easily imagined that instructions to retain the meaning of a word would have required some rehearsal of its phonological form. Yet, phonological rehearsal proved neither necessary nor sufficient to induce a semantic relatedness effect. When we occupied phonological loop with articulatory suppression in Experiment 3, a semantic relatedness effect occurred, indicating that phonological rehearsal is not necessary for this effect. Even when phonological rehearsal was available and was adopted as the primary maintenance strategy (Experiments 1–2), no semantic relatedness effect occurred, indicating that it is not sufficient to induce this effect. Indeed, if phonological rehearsal supported semantic maintenance, we should have found a phonological relatedness effect during semantic retention in Experiment 6, and we did not. The independence of the semantic relatedness effect from phonological rehearsal was strongly supported by the full double dissociation between semantic and phonological instructions with respect to semantic and phonological relatedness effects across Experiments 1, 2, and 6.

The process engaged by instructions to maintain meaning appears to be an optional, controllable mechanism that functions to enhance retention of the sustained item, increasing its resistance to semantic interference during the delay. Beyond the fact that people can voluntarily switch maintenance method across blocks (e.g., from semantic to phonological to visual), Experiment 5 showed that these mechanisms can be "turned off." When subjects semantically encoded the maintenance item but did not maintain it during the delay, the semantic relatedness effect disappeared, even though the task was otherwise matched to the semantic maintenance condition. The lack of active maintenance came at a cost: Subjects made less accurate target judgments, especially when a related probe occurred during the delay. Thus, semantic maintenance improved target performance and reduced interference from the probe. Even without this process, however, performance on the target task was high. This vividly illustrates the ambiguity pervading all prior measures of semantic maintenance that motivated this work: Performance on an immediate retention task can be high and sensitive to semantic variables (e.g., interference from related items) but not at all reflect maintenance.

The retention benefits of actively sustaining semantics also come at a cost, however. During the delay interval of our semantic maintenance tasks, subjects found it significantly more difficult to semantically process related probes than unrelated ones. Thus, the semantic relatedness effect in all five experiments took the form of deficit in reaction time (and in some cases, accuracy) for semantically related probes. Indeed, Experiment 3 showed that the cost of semantic maintenance depended on how long the maintenance item had been semantically maintained, building up with increasing delay: There was nonsignificant priming at short delays, with significant slowing at longer ones. Taken together, these studies demonstrate a behavioral effect that is directly tied to maintenance, related to the intention to maintain semantics, under voluntary control, and related to enhanced performance on semantic tasks at a delay. These characteristics all suggest that the semantic relatedness effect constitutes an informative delay-period marker for semantic maintenance. Next, we consider the implications of these findings for the status of semantic working memory.

Implications

The status of direct semantic maintenance. For over four decades, a fundamental ambiguity has pervaded research on semantic maintenance: Is semantic retention supported by direct semantic maintenance, or does apparent evidence reflect episodic memory or semantic priming? Although many theories presuppose semantic maintenance and many findings are consistent with it, the empirical case has never been established convincingly: If test performance after maintenance can be jointly determined by working- and long-term memory, one cannot address this question. In the present studies, we addressed this persisting ambiguity by developing a delay-period marker diagnostic of semantic maintenance. Next, we consider alternative accounts of the semantic relatedness effect and the extent to which it favors semantic working memory. We argue that the present findings clearly favor the conclusion that direct semantic maintenance occurs, can occur without involvement of other maintenance systems, and is used to retain semantics over short delays, as proposed by Martin (Martin & Romani, 1994; Martin et al., 1994).

Theoretical accounts. In complex cognitive activities, keeping semantic information highly accessible can surely be accomplished in many ways, some of which do not require direct semantic maintenance. In this section, we consider whether other systems or strategies can explain the semantic relatedness effect and the patterns we have observed in it. We begin with a consideration of whether maintenance needs to be assumed at all. We then consider nonsemantic and semantic forms of maintenance and how they might produce the semantic relatedness effect.

Nonmaintenance accounts. Perhaps active maintenance need not be assumed at all, in the present studies. One might, for instance, explain the semantic relatedness effect if the probe functions as a cue that activates related traces in long-term memory. The probe may trigger retrieval of either an episodic or a semantic representation of the maintenance item, which appeared only 8–15 s earlier and was likely very accessible. This retrieval event may have distracted subjects, delaying their lexical decisions to related probes. This delay should not arise when probes are unrelated to the maintenance items. By this view, the more activated or accessible the maintenance item is, the more likely it should be retrieved, increasing distraction and the semantic relatedness effect. For instance, if phonological or visual word-form rehearsal encourages shallow processing of the maintenance item, its semantic representation may not be activated. If so, semantically related probes may not trigger their retrieval, eliminating the semantic relatedness effect.

This simple version of the nonmaintenance hypothesis cannot explain the conditions under which the relatedness effect occurs. For instance, this hypothesis predicts semantic relatedness effects in both the during and after conditions. Because we matched the lag between the maintenance item and the probe in these conditions, the maintenance item should have been equally accessible, and subjects should have been reminded of it in both. Indeed, relatedness effects should be larger in the after condition because the probe appears after the target—a synonym to the maintenance item for half the trials—and because the synonym is closer in time to the after probe. Thus, this hypothesis fails to explain the delay-period specificity of the semantic relatedness effect and predicts the wrong ordering of this effect across the during and after conditions. For similar reasons, the lack of relatedness effects in the long-term memory condition of Experiment 5 poses difficulties. In both the semantic and the long-term memory conditions, subjects deeply processed the maintenance item, which should have yielded similarly accessible items and similar relatedness effects. Finally, the growth of semantic relatedness effects with time in Experiment 3 runs contrary to the nonmaintenance account. Because the maintenance item's accessibility should decline over the interval (or at most remain unchanged), the relatedness effect should be larger early in the delay or at least not be different. The opposite occurred. Thus, variations in accessibility predicted by the nonmaintenance view are not tied to patterns in the relatedness effect, suggesting other factors dictate when they occur. Active maintenance, however, presents a consistent account of these patterns.

Although the simple nonmaintenance model does not fare well, additional assumptions may improve matters. For instance, while waiting for the target, subjects may keep the maintenance item accessible by quickly re-retrieving it from episodic or semantic memory in between lexical decision trials. This would explain why the semantic relatedness effect does not emerge in the long-term memory or after conditions (in which subjects believe they do not have to maintain the item). This modification is reasonable, although it does not clearly predict a buildup in this effect over time. Unfortunately, however, it may be impossible to distinguish this view from a working memory approach based on behavioral evidence. Working memory models do not require maintenance operations to be continuously applied, especially when multiple things are maintained. If so, then the many short-lag (2-s or less) retrievals needed during a delay may resemble reactivations by a maintenance process. This ambiguity is not unique to our semantic maintenance proposal and applies to phonological and visual working memory domains.

On this issue, neuroimaging data provide useful converging evidence for a continuously active semantic maintenance process. Shivde and Thompson-Schill (2004) used the delayed judgment procedure developed here. Subjects received trials presenting a to-be-remembered word for 2 s (the words used here), which they were then to maintain for 10 s. Subjects were asked either to continuously maintain the word's meaning and not its sound (semantic maintenance) or the word's sound by subvocal repetition and not its meaning. Unlike the current studies, no lexical decision task was performed, so that semantic maintenance activations could be disentangled from lexical decision processes. After the delay, subjects judged a target item's relationship to the maintenance item (synonym or rhyme judgments, as in the present studies). Thus, judgment tasks together with the maintenance instructions enabled us to manipulate maintenance strategy while holding the memory items constant across conditions.

Shivde and Thompson-Schill (2004) observed clear evidence for the involvement of a frontotemporal circuit in semantic maintenance: There was more activation in anterior left inferior prefrontal cortex (hereinafter LIPFC) and right inferior prefrontal cortex (Brodmann's area; BA 47/45) and in left middle temporal gyrus (BA 21) during semantic than phonological maintenance. This fits well with findings implicating the LIPFC in controlled semantic processing and left temporal cortex in semantic representation. Moreover, because these activations occurred during a delay period (which excluded several seconds after encoding), they indicate that anterior LIPFC can be engaged continuously to sustain tonic input to the temporal cortex supporting maintenance. In contrast, phonological maintenance activated left parietal cortex (BA 7), consistent with studies implicating BA40 and BA7 in phonological storage. Given that we used similar procedures and materials, the present functional dissociations most likely reflects these different underlying networks. Although concurrent lexical decision may have required intermittent maintenance, this function seems simply explained by the reengagement of the network implicated by Shivde and Thompson-Schill's findings (see also Fiebach, Friederici, Smith, & Swinney, 2007). Nevertheless, short-lag episodic retrievals also contribute to immediate memory performance in addition to activation-based maintenance (e.g., Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005), as we showed clearly in Experiment 5.

Mediated maintenance accounts. Although our results clearly favor the notion that semantic representations were activated over the delay, this activation may have been mediated through some other maintenance system, such as phonological or visual working memory. Under this view, the semantic relatedness effect emerges because phonological (or visual) rehearsal either (a) continuously activates a word's semantics, distracting subjects as described previously, or (b) suppresses concepts associated to the word. This approach can explain the semantic relatedness effect in the semantic condition and its delay-period specificity: Once phonological rehearsal is terminated (after the target judgment), the priming or inhibition might dissipate, reducing the effect. The lack of relatedness effects in the long-term memory condition of Experiment 5 emerges insofar as phonological rehearsal should be absent. Finally, the failure of articulatory suppression to disrupt the semantic relatedness effect in Experiment 3 follows if articulatory suppression leads subjects to abandon phonological rehearsal and engage instead in visual word-form rehearsal.

We have argued throughout, however, that this mediated maintenance hypothesis suffers from conspicuous difficulties. If visual word-form or phonological rehearsal produce semantic relatedness effects, they should have arisen in the phonological or visual word-form rehearsal conditions of Experiments 1, 2, and 4. They did not. In addition, a mediated maintenance account predicts that semantic maintenance should induce phonological relatedness effects. In fact, Experiment 6 shows that although phonological maintenance induces phonological relatedness effects, semantic maintenance does not. Nevertheless, we do not wish to argue that mediated maintenance never occurs, nor that it is not useful as one means of augmenting the retention of semantics over a delay. Rather, we argue that mediated maintenance is not necessary to retain semantics and that direct semantic maintenance exists.

Direct semantic maintenance. We have used the term *direct semantic maintenance* to refer to the idea that semantics can be maintained directly, without mediation through other working memory subsystems. This elementary assumption is made by theories of working memory that posit a general, domain-independent capacity, as well as those positing a specialized semantic maintenance system (e.g., Martin et al., 1994; Wong & Law, 2008). Semantic maintenance in the present one-item maintenance procedure theoretically could be accomplished by sustaining a domain-general attention mechanism on the meaning of the maintenance item, even if that attention mechanism is not part of any working memory system (McElree, 2006). The present results are consistent with each of these varieties of theory, although there is reason to suspect the involvement of a domain-specific semantic maintenance process (Shivde & Thompson-Schill, 2004).

Although the present behavioral data do not distinguish the above possibilities, there is reason to believe that semantic maintenance reflects the operation of a distinct storage/maintenance system. Over the last 15 years, evidence has mounted favoring the view that controlled semantic processing recruits a frontal-posterior network distinct from the one involved in phonological maintenance. A large number of studies have revealed lateralized activation in the LIPFC during semantic processing together with activations in (primarily) left temporal cortex. Phonological processing also activates regions within the LIPFC, though the regions activated are anatomically distinct. Moreover, phonological processing generally activates distinct posterior cortical structures, including posterior parietal cortex. These observations suggest functionally specialized (though surely interacting) systems for processing semantic and phonological information. We have hypothesized that semantic maintenance involves a sustained signal originating from anterior LIPFC modulating representational areas in left temporal cortex, a speculation supported by work with this paradigm (Shivde & Thompson-Schill, 2004).

Possible Mechanisms Underlying the Semantic Relatedness Effect

In five experiments we found that semantic maintenance hindered reaction times to semantically related probe words. Here we consider two accounts of this effect: the inhibition account and the contingent capture hypothesis. Both accounts assume that semantic maintenance is accomplished by a maintenance process that sustains the target in the face of interference.

The inhibition hypothesis. Sustaining attention to a word's meaning activates associated concepts, and if this activation accumulates over the delay, those associates may intrude, triggering inhibition. The need for inhibition may be especially potent in our paradigm, which requires subjects to focus their attention on a word's meaning while making a lexical decision about each of a quick-moving stream of letter strings, potentially increasing the background distraction in semantic memory. Thus, sustaining attention may require suppression of associated concepts to prevent or overcome reflexive orienting to those items triggered during maintenance.

The inhibition hypothesis accounts for the current data well. Inhibition predicts that associates of the maintained concept should be suppressed before the probe appears, as a result of sustained attention. As such, lexical decisions about those probes should be hindered. Ideas underlying unrelated probes, in contrast, should not be activated by maintenance and so should not be inhibited. Inhibition would only be needed during maintenance, and so removing attention from the maintained item after the delay should allow inhibition to dissipate and probe reaction times should be unaffected. This view predicts a nonmonotonic relation between maintenance duration and inhibition: At short intervals, attending the maintenance item should prime associates as activation spreads at subthreshold levels; at longer intervals, associates will exceed threshold and capture attention, requiring inhibition. Consistent with this, larger semantic relatedness effects were observed at long delays. Visual or phonological maintenance, by contrast, should not prime related concepts as much, and, correspondingly, inhibition should be unnecessary. Such maintenance might inhibit visual or phonological associates, as suggested by the phonological relatedness effect in Experiment 6. Thus, inhibition explains the present effects. This view builds on a broader framework indicating a role for inhibition in confronting competition in memory (Anderson & Green, 2001; Conway & Engle, 1994; Hasher & Zacks, 1988; Jonides et al., 1998; Lustig & Hasher, 2001), and there is precedent to suppose that inhibition affects semantic memory (see, e.g., Johnson & Anderson, 2004; see also Blaxton & Neely, 1983; Dagenbach et al., 1990; for a review, see Levy & Anderson, 2002). Indeed, patient M.L., who has a deficit in semantic working memory, has difficulties with inhibition in verbal tasks involving attention and working memory (Hamilton & Martin, 2005).

Contingent capture hypothesis. The present findings are also compatible with a contingent capture mechanism. Research on perceptual attention has established that abrupt onsets can trigger reflexive orienting of attention to those onsets. For instance, a visual stimulus that abruptly appears at an unattended location can "capture" attention, causing automatic reorienting to that region (Jonides & Irwin, 1981; Yantis & Jonides, 1990; for a review, see Yantis, 1996; for examples in audition, see Green & McKeown, 2001; McDonald, Teder-Sälejärvi, & Hillyard, 2000). "Capture" occurs quickly and involuntarily (Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989), even when subjects know that a target will never appear in the onset's location (Remington, Johnston, & Yantis, 1992) and are motivated to ignore it. Susceptibility to capture depends on the relevance of the stimulus to subjects' goals. According to this *contingent involuntary orienting hypothesis*, a stimulus property captures attention if it matches the current top-down attentional control settings (Folk & Remington, 1998). Thus, when one is seeking an object with a characteristic in one location, objects with that characteristic in irrelevant locations capture attention.

The semantic relatedness effect may reflect contingent capture by goal-relevant stimuli. By this view, subjects establish a top-down attentional control set for sustaining attention to the maintained concept and for detecting the target trial. During the delay, maintenance of that set ought to lead concepts related to the maintenance item to capture attention and improperly trigger the task set. If so, subjects may need to disengage attention from the maintenance item and return to the probe task. This added time to reengage may underlie the semantic relatedness effect. Thus, a probe's tendency to trigger capture, not inhibition, may hinder response times. Because unrelated probes are not goal relevant, they would not capture attention. This hypothesis resembles the nonmaintenance memory account but crucially posits that the maintenance item is sustained continuously (i.e., it assumes direct semantic maintenance).

The contingent capture hypothesis explains many features of the data. A key prediction is that semantic relatedness effects should disappear when the attentional control omits the semantically related maintenance item. For instance, the effect's disappearance after the maintenance interval may reflect subjects abandoning their attentional set after the target judgment (at least the maintenance portion of it), reducing capture. Similarly, when subjects encode the maintenance item (Experiment 5) and do not think about it until the target appears, no relatedness effect should occur because the control set is inactive during the delay. Moreover, the semantic relatedness effect should never arise in the phonological and visual working memory conditions, because probes are not visually or phonologically related to the maintenance item and so are not goal relevant. However, when the probe is phonologically related (Experiment 6), a phonological relatedness effect should (and does) occur when subjects retain phonology. The main weakness of this hypothesis is explaining why semantic relatedness effects grow with increasing delays. The most obvious implication is that because the attentional set should be most accessible after it is formed, short delays should induce more capture, contrary to what we found.

Thus, the contingent capture hypothesis does not explain all the data as neatly as does inhibition. These experiments do not strongly differentiate these hypotheses, however, and further research is required to clarify the mechanisms at work. However, even the capture hypothesis hinges on semantic information being retained. Moreover, it must assume that semantic retention does not occur during phonological or visual word-form rehearsal. Thus, whether slowing reflects inhibition or capture, the semantic relatedness effect remains a marker of semantic maintenance.

Concluding Remarks

The ability to sustain meanings in a highly accessible state is fundamental to any theory seeking to explain the mechanisms governing thought. Indeed, this capacity is presumed to exist by computational theories of cognition and theories of individual differences in intellectual capacity. Despite these facts, surprisingly little work has examined semantic working memory, and what has been done, with few exceptions, does not require acceptance of this construct. Over four decades, nearly all studies have not adequately separated semantic maintenance from other varieties of working memory, semantic priming, and episodic memory.

The present experiments indicate that semantic representations can be maintained over a delay by a process acting directly on the semantics and that is not mediated by other forms of working memory or long-term memory. Thus, a separate semantic working memory capacity may exist. Semantic maintenance leaves a behavioral footprint during maintenance that is absent afterward; that occurs during semantic but not phonological or othographic maintenance; that is not a by-product of semantic encoding but arises only with sustained semantic processing; and that may be supported by an inhibition process that suppresses the semantic space surrounding the maintained item. The present paradigm and results take a useful step in elucidating the basic mechanisms that allow focused attention to conceptual information in working memory.

References

Anderson, M. C., & Green, C. (2001, March 15). Suppressing unwanted memories by executive control. *Nature, 410,* 366–369. doi:10.1038/35066572

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.

Besner, D., & Davelaar, E. (1982). Basic processes in reading: Two phonological codes. *Canadian Journal of Psychology/Revue canadienne de psychologie, 36,* 701–711. doi:10.1037/h0080665

Blaxton, T. A., & Neely, J. H. (1983). Inhibition from semantically related primes: Evidence of a category-specific inhibition. *Memory & Cognition, 11,* 500–510. doi:10.3758/BF03196987

Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review, 1,* 122–125. doi:10.3758/BF03200768

Bregman, A. S. (1968). Forgetting curves with semantic, phonetic, graphic and contiguity cues. *Journal of Experimental Psychology, 78,* 539–546.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12–21. doi:10.1080/17470215808416249

Calfee, R. C., & Peterson, R. E. (1968). Effect of list organization on short-term probe recall. *Journal of Experimental Psychology, 78,* 468–474. doi:10.1037/h0026464

Cofer, C. N., Bruce, D. R., & Reicher, G. M. (1966). Clustering in free recall as a function of certain methodological variations. *Journal of Experimental Psychology, 71,* 858–866. doi:10.1037/h0023217

Cohen, B. H. (1966). Some-or-none characteristics of coding behavior. *Journal of Verbal Learning and Verbal Behavior, 5,* 182–187. doi:10.1016/S0022-5371(66)80014-2

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407–428. doi:10.1037/0033-295X.82.6.407

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123,* 354–373. doi:10.1037/0096-3445.123.4.354

Corman, C. D., & Wickens, D. D. (1968). Retroactive inhibition in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 7,* 16–19. doi:10.1016/S0022-5371(68)80157-4

Cowan, N. (1995). *Attention and memory: An integrated framework.* New York, NY: Oxford University Press.

Craik, F. I., & Levy, B. A. (1970). Semantic and acoustic information in primary memory. *Journal of Experimental Psychology, 86,* 77–82. doi:10.1037/h0029799

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294. doi:10.1037/0096-3445.104.3.268

Dagenbach, D., & Carr, T. H. (1994). Inhibitory processes in perceptual recognition: Evidence for a center-surround attentional mechanism. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 327–357). San Diego, CA: Academic Press.

Dagenbach, D., Carr, T. H., & Barnhardt, T. M. (1990). Inhibitory semantic priming of lexical decisions due to failure to retrieve weakly activated codes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 328–340. doi:10.1037/0278-7393.16.2.328

Dale, H. C., & Gregory, M. (1966). Evidence of semantic coding in short-term memory. *Psychonomic Science, 5,* 153–154.

Dallett, K. M. (1964). Number of categories and category information in free recall. *Journal of Experimental Psychology, 68,* 1–12. doi:10.1037/h0049092

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review, 112,* 3–42. doi:10.1037/0033-295X.112.1.3

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19–23. doi:10.1111/1467-8721.00160

Fiebach, C. J., Friederici, A. D., Smith, E. E., & Swinney, D. (2007). Lateral inferotemporal cortex maintains conceptual–semantic representations in verbal working memory. *Journal of Cognitive Neuroscience, 19,* 2035–2049. doi:10.1162/jocn.2007.19.12.2035

Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 847–858. doi:10.1037/0096-1523.24.3.847

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation, 14,* 375–399. doi:10.3758/BF03203275

Glanzer, M., & Schwartz, A. (1971). Mnemonic structure in free recall: Differential effects on STS and LTS. *Journal of Verbal Learning and Verbal Behavior, 10,* 194–198. doi:10.1016/S0022-5371(71)80013-0

Green, T. J., & McKeown, J. D. (2001). Capture of attention in selective frequency listening. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 1197–1210. doi:10.1037/0096-1523.27.5.1197

Gregg, V. H., Freedman, C. M., & Smith, D. K. (1989). Word frequency, articulatory suppression and memory span. *British Journal of Psychology, 80,* 363–374. doi:10.1111/j.2044-8295.1989.tb02326.x

Haarmann, H., & Usher, M. (2001). Maintenance of semantic information in capacity limited item short-term memory. *Psychonomic Bulletin & Review, 8,* 568–578. doi:10.3758/BF03196193

Hamilton, A. C., & Martin, R. C. (2005). Dissociations among tasks involving inhibition: A single-case study. *Cognitive, Affective, & Behavioral Neuroscience, 5,* 1–13. doi:10.3758/CABN.5.1.1

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.

Hofer, R. (1965). Intertrial proactive inhibition in short-term memory. *Psychological Reports, 17,* 755–760.

Hoffman, P., Jefferies, E., Ehsan, S., Hopper, S., & Lambon Ralph, M. A. (2009). Selective short-term memory deficits arise from impaired domain-general semantic control mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 137–156. doi:10.1037/a0013985

Hoffman, P., Jefferies, E., & Lambon Ralph, M. A. (2011). Explaining semantic short-term memory deficits: Evidence for the critical role of semantic control. *Neuropsychologia, 49,* 368–381. doi:10.1016/j.neuropsychologia.2010.12.034

Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30,* 685–701. doi:10.1016/0749-596X(91)90032-F

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology, 86,* 527–536. doi:10.1111/j.2044-8295.1995.tb02570.x

Huttenlocher, J., & Newcombe, N. (1976). Semantic effects on ordered recall. *Journal of Verbal Learning and Verbal Behavior, 15,* 387–399. doi:10.1016/S0022-5371(76)90034-7

Johnson, S. K., & Anderson, M. C. (2004). The role of inhibitory control in forgetting semantic knowledge. *Psychological Science, 15,* 448–453. doi:10.1111/j.0956-7976.2004.00700.x

Jonides, J., & Irwin, D. E. (1981). Capturing attention. *Cognition, 10,* 145–150. doi:10.1016/0010-0277(81)90038-X

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience, 9,* 462–475. doi:10.1162/jocn.1997.9.4.462

Jonides, J., Smith, E. E., Marshuetz, C., Koeppe, R. A., & Reuter-Lorenz, P. A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proceedings of the National Academy of Sciences, USA, 95,* 8410–8413. doi:10.1073/pnas.95.14.8410

Joordens, S., & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1083–1105. doi:10.1037/0278-7393.23.5.1083

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1,* 153–161. doi:10.1016/S0022-5371(62)80023-1

Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Kuhl, B. A., & Anderson, M. C. (2011). More is not always better: Paradoxical effects of repetition on semantic accessibility. *Psychonomic Bulletin & Review.* Advance online publication. doi:10.3758/s13423-011-0110-0

Levy, B. A., & Baddeley, A. (1971). Recall of semantic clusters in primary memory. *Quarterly Journal of Experimental Psychology, 23,* 8–13. doi:10.1080/00335557143000013

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences, 6,* 299–305. doi:10.1016/S1364-6613(02)01923-X

Loess, H. (1964). Proactive inhibition in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 3,* 362–368. doi:10.1016/S0022-5371(64)80003-7

Loess, H. (1967). Short-term memory, word class, and sequence of items. *Journal of Experimental Psychology, 74,* 556–561. doi:10.1037/h0024787

Loess, H. (1968). Short-term memory and item similarity. *Journal of Verbal Learning and Verbal Behavior, 7,* 87–92. doi:10.1016/S0022-5371(68)80169-0

Lustig, C., Hasher, L., & Tonev, S. T. (2001). Inhibitory control over the present and the past. *European Journal of Cognitive Psychology, 13,* 107–122.

Martin, N., & Saffran, E. M. (1997). Language and auditory–verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14,* 641–682. doi:10.1080/026432997381402

Martin, R. C., & He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language, 89,* 76–82.

Martin, R. C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multiple-components view. *Neuropsychology, 8,* 506–523. doi:10.1037/0894-4105.8.4.506

Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language, 33,* 83–111. doi:10.1006/jmla.1994.1005

McDonald, J. J., Teder-Sälejärvi, W., & Hillyard, H. D. (2000, October 19). Involuntary orienting to sound improves visual perception. *Nature, 407,* 906–908. doi:10.1038/35038085

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46, pp. 155–200). San Diego, CA: Academic Press.

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 315–330. doi:10.1037/0096-1523.15.2.315

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78,* 679–684.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research, 29,* 1631–1647. doi:10.1016/0042-6989(89)90144-2

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1994). The University of South Florida word association, rhyme and word fragment norms. *Unpublished manuscript.*

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*(A), 384–404.

Raser, G. A. (1972). Recoding of semantic and acoustic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 11,* 692–697. doi:10.1016/S0022-5371(72)80003-3

Remington, R. W., Johnston, J. C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception & Psychophysics, 51,* 279–290. doi:10.3758/BF03212254

Romani, C., & Martin, R. A. (1999). A deficit in the short-term retention of lexical–semantic information: Forgetting words but remembering a story. *Journal of Experimental Psychology: General, 128,* 56–77. doi:10.1037/0096-3445.128.1.56

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences, 26,* 709–728. doi:10.1017/S0140525X03000165

Sanders, A. G., & Schroots, J. J. F. (1968). Cognitive categories and memory span: II. The effect of temporal vs. categorical recall. *Quarterly Journal of Experimental Psychology, 20,* 373–379. doi:10.1080/14640746808400177

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21,* 168–175. doi:10.3758/BF03202729

Shivde, G. S., & Anderson, M. C. (2001). The role of inhibition in meaning selection: Insights from retrieval-induced forgetting. In D. Gorfein (Ed.), *On the consequences of meaning selection* (pp. 175–190). Washington, DC: American Psychological Association.

Shivde, G., & Thompson-Schill, S. L. (2004). Dissociating semantic and phonological maintenance using fMRI. *Cognitive, Affective, & Behavioral Neuroscience, 4,* 10–19. doi:10.3758/CABN.4.1.10

Shulman, H. G. (1970). Encoding and retention of semantic and phonemic information in short term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 499–508. doi:10.1016/S0022-5371(70)80093-7

Schulman, H. G. (1971). Similarity effects in short-term memory. *Psychological Bulletin, 75,* 399–415.

Shulman, H. G. (1972). Semantic confusion errors in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 11,* 221–227. doi:10.1016/S0022-5371(72)80080-X

Tehan, G., & Humphreys, M. S. (1988). Articulatory loop explanations of memory span and pronunciation rate correspondences: A cautionary note. *Bulletin of the Psychonomic Society, 26,* 293–296.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, USA, 94,* 14792–14797. doi:10.1073/pnas.94.26.14792

Thompson-Schill, S. L., Swick, D., Farah, M. J., D'Esposito, M., Kan, I. P., & Knight, R. T. (1998). Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings. *Proceed-*

ings of the National Academy of Sciences, USA, 95, 15855–15860. doi:10.1073/pnas.95.26.15855

Turvey, M. T., Cremins, J. J., & Lombardo, T. (1969). Taxonomic categories and proactive interference in short-term memory. *Psychonomic Science, 15,* 307–308.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31,* 329–338. doi:10.1016/S0896-6273(01)00359-2

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., . . . Buckner, R. L. (1998, August 21). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science, 281,* 1188–1191. doi:10.1126/science.281.5380.1188

Warrington, E. K., Kinsbourne, M., & James, M. (1966). Uncertainty and transitional probability in the span of apprehension. *British Journal of Psychology, 57,* 7–16. doi:10.1111/j.2044-8295.1966.tb00998.x

Watkins, M. J. (1977). The intricacy of memory span. *Memory & Cognition, 5,* 529–534. doi:10.3758/BF03197396

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72,* 89–104. doi:10.1037/h0021797

Weeks, R., & Katz, L. (1970). Similarity in short-term recall: A comparison of Type II operating characteristics and percent recalled. *Psychonomic Science, 20,* 232–233.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77,* 1–15. doi:10.1037/h0028569

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 2,* 440–445. doi:10.1016/S0022-5371(63)80045-6

Wickens, D. D., & Clark, S. (1968). Osgood dimensions as an encoding class in short-term memory. *Journal of Experimental Psychology, 78,* 580–584. doi:10.1037/h0026643

Wickens, D. D., Moody, M. J., & Dow, R. (1981). The nature and timing of the retrieval process and of interference effects. *Journal of Experimental Psychology: General, 110,* 1–20. doi:10.1037/0096-3445.110.1.1

Wickens, D. D., Moody, M. J., & Vidulich, M. A. (1985). Retrieval time as a function of memory set size, type of probes, and interference in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 154–164. doi:10.1037/0278-7393.11.1.154

Woltz, D. J. (2010). Long-term semantic priming of word meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1510–1528. doi:10.1037/a0021039

Wong, W., & Law, S.-P. (2008). The relationship between semantic short-term memory and immediate serial recall of known and unknown words and nonwords: Data from two Chinese individuals with aphasia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 900–917. doi:10.1037/0278-7393.34.4.900

Yantis, S. (1996). Attentional capture in vision. In A. F. Kramer, M. G. H. Coles, & G. D. Logan (Eds.), *Converging operations in the study of visual selective attention* (pp. 45–76). Washington, DC: American Psychological Association.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 121–134. doi:10.1037/0096-1523.16.1.121

Appendix A

Materials Used in the Current Experiments

Maintained (item)	Semantic (target item)	Phonological (target item)	Semantically related (probe word)	Phonologically related (probe word)
afraid	scared	kade	ghost	braid
angry	mad	lang	yell	sea
bad	evil	zak	crime	lad
brag	boast	yag	pride	hag
build	construction	pilg	grow	filled
busy	occupied	pizz	bored	fizzy
cautious	careful	baw	danger	raw
center	middle	keff	circle	enter
change	alter	tane	another	arrange
cold	chilly	fole	weather	fold
courage	bravery	der	strength	sir
dead	lifeless	vek	cemetery	bread
different	contrary	zin	similar	sent
disgrace	shame	fiss	failure	place
display	exhibit	fape	show	gray
easy	simple	nee	quick	breezy
embrace	hug	tase	arms	grace
emotion	feeling	po	cry	commotion
endure	survive	byoo	problem	sure
enjoy	like	leb	fun	toy
equal	even	deet	rights	sequel
exit	leave	vip	enter	wit
fever	temperature	teeb	sick	believer

(*Appendices continue*)

Appendix A (*continued*)

Maintained (item)	Semantic (target item)	Phonological (target item)	Semantically related (probe word)	Phonologically related (probe word)
find	locate	wibe	lose	grind
first	primary	terr	last	burst
fraud	deceit	kaw	fake	broad
freedom	liberty	klebe	slavery	gum
genuine	authentic	pell	real	bin
glad	happy	kaz	satisfied	clad
go	proceed	loe	race	toe
grasp	clutch	tass	release	gasp
happen	occur	dat	event	ten
hate	disdain	paze	love	fate
help	aid	pell	trouble	yelp
hike	trek	yibe	road	spike
honest	truthful	dest	liar	fawn
honor	glory	dahn	dignity	fir
hurt	pain	ler	harm	curt
idea	thought	gee	opinion	sigh
imposter	pretender	hin	disguise	foster
long	lengthy	ponk	short	song
mankind	humanity	jine	world	bind
mistake	error	riss	sorry	break
nothing	zero	pring	empty	wing
old	aged	yole	new	mold
outcome	result	nowt	win	grout
part	piece	lah	portion	tart
permit	allow	rit	deny	emit
pity	sympathy	zee	mercy	bit
prank	joke	nank	fool	thank
purchase	buy	wess	price	fuss
quarrel	spat	taw	dispute	moral
rascal	scoundrel	jad	mischief	cackle
reject	renounce	bek	accept	dissect
rejoice	celebration	poy	pleasure	voice
rise	ascend	nipe	fall	size
same	alike	habe	exact	blame
shape	form	wabe	square	grape
shiny	glossy	ree	silver	tiny
skill	talent	vin	clever	will
sleep	rest	tees	dream	steep
sob	weep	kob	tears	knob
sorrow	sadness	voze	misery	borrow
source	origin	vore	beginning	horse
speak	talk	veek	tongue	seek
stamina	energy	jad	sport	gram
stop	halt	opp	sign	drop
strong	muscular	quong	big	belong
succeed	achieve	wuck	famous	feed
tale	story	dail	fairy	nail
test	exam	pess	score	best
think	ponder	grink	brain	sink
tired	weary	sile	nap	wired
tour	expedition	loog	visit	lure
triumph	victory	bly	overcome	fry
walk	stroll	baw	run	sock
wander	roam	ver	aimless	ponder
want	desire	mont	crave	font
weak	frail	feeb	feeble	leak
work	labor	nerg	job	irk

Note. Phonologically related probe words were used only in the final experiment.

(*Appendices continue*)

Appendix B

Average Reaction Time (in MS) to Probe Items Broken Down by Target Response

	Semantic								Other strategy							
	During				After				During				After			
	Related		Unrelated		Related		Unrelated		Related		Unrelated		Related		Unrelated	
Experiment	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
1	858	845	753	772	808	814	790	782	784	775	809	803	808	788	799	766
2	801	818	737	726	777	767	769	776	800	783	792	756	769	777	798	775
3	867	830	855	796	896	799	863	818	NA	NA	NA	NA	NA	NA	NA	NA
4	816	863	764	756	796	792	818	761	756	782	778	769	741	766	757	760
5	799	809	750	725	779	771	774	751	761	735	731	753	743	753	784	734
M	828	833	772	755	811	789	803	778	776	769	777	770	765	771	784	759

Note. NA = not applicable.

Appendix C

Percent Accuracy of Probe Item Responses Across Experiments

	Semantic				Other strategy			
	During		After		During		After	
Experiment	Related	Unrelated	Related	Unrelated	Related	Unrelated	Related	Unrelated
1	93.7	95	93.7	95.9	97.2	95.3	96.9	95.9
2	95.6	98.7	96.9	96.2	92.8	94.7	97.2	93.4
3	94.4	95.6	95.3	95.6	NA	NA	NA	NA
4	94.4	95	91.9	93.4	93.7	94.1	96.6	94.4
5	95.9	94.7	96.6	97.5	95	95.6	96.6	97.2
M	94.8	95.8	94.9	95.7	94.7	94.9	96.8	95.2

Note. NA = not applicable.

(*Appendices continue*)

Appendix D

Percent Accuracy of Target Item Responses Across Experiments

	Semantic				Other strategy			
	During		After		During		After	
Experiment	Related	Unrelated	Related	Unrelated	Related	Unrelated	Related	Unrelated
1	87.5	88.1	89.7	88.7	77.5	79.7	76.3	80
2	87.8	86.6	90.6	87.2	84.7	84.7	80.6	80.6
3	77.5	76.5	81	79	NA	NA	NA	NA
4	89.1	87.8	89.7	89.1	83.4	84.7	84.1	85
5	85	87.5	87.2	75	72.2	76.2	77.8	78.4
M	85.4	85.3	87.6	83.8	79.5	81.3	79.7	81.0

Note. NA = not applicable.

Appendix E

Percent Accuracy of Probe and Target Item Responses in Experiment 6

	Semantic				Other strategy			
	During		After		During		After	
Variable	Related	Unrelated	Related	Unrelated	Related	Unrelated	Related	Unrelated
Probe accuracy	98.2	88.4	88.4	90	84.7	97.3	89.5	87.5
Target accuracy	79.7	80.7	79.5	80.3	89.8	89	89.5	88.4

Received November 6, 2010
Revision received May 26, 2011
Accepted June 5, 2011 ■