Control of Interference During Working Memory Updating

Arnaud Szmalec, Frederick Verbruggen,
and André Vandierendonck
Ghent University

Eva Kemps
Flinders University

The current study examined the nature of the processes underlying working memory updating. In 4 experiments using the *n*-back paradigm, the authors demonstrate that continuous updating of items in working memory prevents strong binding of those items to their contexts in working memory, and hence leads to an increased susceptibility to proactive interference. Results of Experiments 1 and 2 show that this interference reflects a competition between a process that reveals the degree of familiarity of an item and a context-sensitive recollection process that depends on the strength of bindings in working memory. Experiment 3 further clarifies the origins of interference during updating by demonstrating that even items that are semantically related to the updated working memory contents but that have not been maintained in working memory before cause proactive interference. Finally, the results of Experiment 4 indicate that the occurrence of interference leads to top-down behavioral adjustments that prioritize recollection over familiarity assessment. The implications of these findings for the construct validity of the *n*-back task, for the control processes involved in working memory updating, and for the concept of executive control more generally are discussed.

Keywords: working memory, updating, proactive interference, executive control, *n*-back

Working memory (WM) is a theoretical construct that refers to the structures and processes underlying the temporary retention and manipulation of information in support of higher cognitive tasks (Baddeley, 1986; Miyake & Shah, 1999). One typical feature of WM is that its capacity is limited in the sense that only a reduced amount of information can be kept in an active, directly accessible state at any time (e.g., Cowan, 2005). Therefore, when confronted with a stream of information that exceeds the capacity limits, a person is restricted to temporarily store subsets of the information and to successively update those representations as new information becomes available. Very often, however, it then becomes hard to distinguish older from the more recent information, a behavioral phenomenon referred to as *proactive interference* (e.g., Jonides & Nee, 2006). In what follows, we first introduce the theoretical concept of updating, and then focus on the involvement of cognitive interference in updating.

Working Memory Updating, Interference, and the *n*–Back Task

WM can be conceptualized as the activated part of long-term memory (LTM) with a region of direct access (DA) where infor-

mation is temporarily maintained in a directly accessible state (Oberauer, 2009; see also Cowan, 1988, 1995). The DA region is assumed to be responsible for establishing and holding temporary bindings between contents (e.g., objects or word stimuli) and contexts (i.e., positions in a cognitive coordinate system, such as serial position). During updating, old contents in the DA region of WM are replaced by new ones. Therefore, updating requires *flexible binding and unbinding* of items to their contexts in the DA region of WM (Oberauer, 2009). This flexibility consists of finding an optimal balance between maintaining a representation in WM and replacing it by a new one. As a result, it is assumed that continuous updating prevents strong binding of items to their context in WM, and that, therefore, updating tasks are particularly prone to proactive interference compared with pure WM maintenance tasks.

A number of recent studies examined shielding of WM contents against interference during updating. Kessler and Meiran (2008; see also Kessler & Meiran, 2006), for example, argued for the existence of two dissociable component processes of WM updating. The first one is a local process, which is responsible for modifying the relevant representations in memory. The second one is a global process, which is responsible for stabilizing the representations in memory after the modification took place. According to Kessler and Meiran, this global process protects the contents of WM against interference. Similarly, Bunting, Cowan, and Saults (2006) proposed that subjects try to rehearse during updating to reduce interference. Support for the role of interference control in updating was also found in a series of experiments that used a choice reaction time task that involves updating of response representations (Szmalec, Demanet, Vandierendonck, & Verbruggen, 2009; Szmalec & Vandierendonck, 2007). Combined, these studies suggest that the frequent updating of response representations stored in WM elicits interference among those representations.

Despite the fact that an increasing number of theorists assume that interference occurs during WM updating, the precise nature of this interference, and particularly how the cognitive system deals with it, remains unclear. Therefore, our aim in this study was to clarify how updating produces interference and how this interference is dealt with. To this end, we used the *n*-back paradigm, which offers a measure of WM updating (e.g., Collette & Van der Linden, 2002; Oberauer, 2005). The classical *n*-back task (e.g., Smith & Jonides, 1997) is a recognition task, which requires subjects to determine whether each item (e.g., letter) in a list matches the item that was presented *n* positions before (an example of a 2-back match is F-*B*-L-*B*; an example of a 2-back mismatch is F-*G*-L-*B*). Therefore, subjects are required to remember a specified number (*n*) of the most recently presented items in serial order (*n*-back). While the task evolves and new items are presented, the subjects have to update the memorized string of *n* most recent items: They need to unbind the oldest item and bind the most recent one to a position in WM. Interference occurs in the *n*-back task on so-called *lure* trials. An example of an $n + 1$ lure trial in a 2-back task is **B**-*F*-L-*B* (i.e., a no-longer relevant item—in bold—corresponds to the currently presented item), an $n - 1$ lure in a 2-back task is L-*F*-**B**-*B* (i.e., a not-yet relevant item—in bold—corresponds to the currently presented item). Such lure trials show that when a newly presented item does not match the *n*-back item but one of its neighbors (i.e., in the $n+$ or $n-$ position), subjects are slower and make more errors (e.g., Gray, Chabris, & Braver, 2003; Jonides & Nee, 2006; Kane, Conway, Miura, & Colflesh, 2007; McElree, 2001; Oberauer, 2005). This nicely illustrates the idea that continuously updating items in WM makes it hard to distinguish between relevant and irrelevant items, resulting in cognitive interference or conflict.

The Present Study

The experiments reported here were designed to test the hypothesis that interference during updating reflects a competition between the assessment of *familiarity* of an item in memory and the explicit *recollection* of that item, and that this interference is counteracted by prioritizing recollection over familiarity. Familiarity matching and recollection were initially presented as two processes involved in recognition from LTM (see Yonelinas, 2002, for a review). According to a dual-process model of recognition memory, the assessment of familiarity of a stimulus and the recollection of detailed information about previous events are two dissociable processes that operate in parallel during item recognition. Assessment of familiarity is believed to be a process that merely relies on the identity and activation of a representation in memory. By contrast, recollection is assumed to be an analytic search process that involves the context in which an item was previously encountered.

Several authors have proposed that familiarity matching and recollection also allow access to information in WM (e.g., Goethe & Oberauer, 2008; Oztekin & McElree, 2007). In short-term recognition, familiarity matching is driven by the degree of activation of items in LTM, whereas recollection is the retrieval of items from the DA region of WM (Oberauer, 2009). The quality of recollection is assumed to directly depend on the strength of the stimulus bindings in the DA region. Proactive interference in WM arises when there is mismatch between the familiarity assessment

and the recollection process (Jonides & Nee, 2006; Oberauer, 2005). Applied to the *n*-back task, this means that an item that was presented shortly before will elicit a familiarity signal, but this familiarity signal will not allow differentiating whether this item is in the target *n*-back position or whether it is just a neighboring item of the target. On mismatch trials, it is possible to react accurately on the basis of the familiarity signal only. However, on lure trials, the familiarity signal will fire ("the item has been encountered previously") but the recollection process is needed to override the misleading activation from the familiarity process by providing contextual evidence ("but it is not in the *n*-back position"). As we explained above, continuous updating hinders strong binding of items in the DA region of WM and hence also hinders recollection. Because a weak recollection provides poor contextual information about a previous encounter with the stimulus, it will be much more difficult to counteract the misleading input from the familiarity signal, leading to impaired performance on *n*-back lure trials. Thus, updating tasks would be particularly sensitive to proactive interference.

Four experiments are reported here. Experiments 1 and 2 tested the hypothesis that *n*-back lure interference reflects a competition between the assessment of the familiarity of an item and the explicit recollection of that item. Experiment 3 further clarified the origins of the *n*-back lure interference effects by demonstrating that even memory contents that have never been presented before also cause proactive interference, merely through semantic relatedness to the items that were encountered before. Finally, Experiment 4 tested the hypothesis that lure interference is controlled by a top-down mechanism prioritizing recollection over familiarity processing.

Experiment 1

Starting from the hypothesis that lure interference reflects a competition between familiarity assessment and recollection, we investigated the effect of different kinds of *n*-back lure trials in Experiment 1. We compared the detrimental effects of lures in the $n+$ and $n-$ positions. An $n+$ lure trial (e.g., **B**-*F*-L-*B* in a 2-back task) reflects interference from an old item that is not relevant anymore to the task and that is therefore being discarded (or unbound) from WM. An $n-$ lure trial (e.g., L-*F*-**B**-*B*) reflects interference from an item that is still relevant to the task. Although both $n+$ and $n-$ lures are assumed to cause proactive interference, they differ in that an $n-$ lure item must still be kept activated in the DA region of WM for future processing, whereas an $n+$ lure item is allowed to degrade.[1] The different kinds of lure interference that were used, expressed as a function of the set size ($n = 1, 2,$ or 3), are displayed in Figure 1. Because our theoretical account does not distinguish between the kind of material that is being updated, we

---

[1] It is still an issue of debate whether forgetting is caused by decay or interference (see Oberauer & Lewandowsky, 2008). Recent findings in fact show that interference has a much larger effect on the degradation of stimuli in short-term memory, whereas the effects of temporal decay appear to be much smaller (Berman, Jonides, & Lewis, 2009). Because the present study did not allow strong conclusions in favor of one of both views, we prefer to use the descriptive terms *stimulus degradation* and *forgetting* instead of *decay* throughout this article.

1-back	Match	...G-F-*B*-B
	Mismatch	...G-F-*T*-B
	N+1 Lure	...G-*B*-*T*-B
	N+2 Lure	...**B**-G-*T*-B

2-back	Match	...G-*B*-F-B
	Mismatch	...G-*F*-T-B
	N+1 Lure	...**B**-*T*-G-B
	N−1 Lure	...G-*T*-**B**-B

3-back	Match	...*B*-F-G-B
	Mismatch	...*G*-F-T-B
	N−1 Lure	...G-**B**-T-B
	N−2 Lure	...*G*-T-**B**-B

Figure 1. The different kinds of *n*-back trials used in the verbal condition of Experiment 1.

tested our hypotheses using both verbal contents (i.e., consonants) and visuospatial contents (i.e., dots presented in a 5×5 grid).

In the 1-back variant of the task, two *n*+ lures were used, namely one in the 2-back (i.e., $n + 1$) and one in the 3-back position (i.e., $n + 2$); by definition, *n*– lures do not occur in a 1-back task. The comparison of $n + 1$ and $n + 2$ lures enabled us to estimate how proactive interference in the *n*-back task evolves over time. We hypothesized that the familiarity of an item that has become irrelevant to the task would decrease, reducing the conflict between familiarity and recollection. We therefore predicted that $n + 2$ lures would cause less interference than $n + 1$ lures. In the 2-back variant of the task, one *n*– lure (i.e., in the 1-back position) and one *n*+ lure (i.e., in the 3-back position) were used. The comparison of $n − 1$ and $n + 1$ lures enabled us to estimate interference caused by items that are no longer relevant (i.e., discarded from the DA region) and items that are relevant for future processing (i.e., maintained in the DA region). We hypothesized that the familiarity signal associated with the still-relevant item would be stronger and hence its competition with recollection would be greater. We therefore predicted that in a 2-back task, a lure in the $n − 1$ (i.e., 1-back) position would produce more interference than a lure in the $n + 1$ position (i.e., 3-back). Finally, the 3-back variant of the *n*-back task includes two lures of the type *n*–, namely one in the $n − 1$ (i.e., 2-back) position and one in the $n − 2$ (i.e., 1-back) position. Items in both of these positions are relevant for future processing, so they need to be temporarily maintained in the DA region of WM before arriving in the target *n*-back position. To maintain items readily accessible, subjects typically rehearse those items (e.g., Page & Norris, 1998; but see also Oberauer & Lewandowsky, 2008, for a balanced view on the usefulness of rehearsal in the context of forgetting). Because items in the 2-back position are temporarily older than items in the 1-back position, it can be assumed that 2-back items have been rehearsed more often. Maintenance rehearsal is known to increase the familiarity of an item (e.g., Gardiner, Gawlik, & Richardson-Klavehn, 1994); therefore, we predicted that a lure in the 2-back position would be more familiar and would therefore cause more interference than a lure in the 1-back position.

Method

Subjects and design. Sixty university students (34 women; mean age = 21 years, range = 17–29 years) volunteered for participation. The parameter *n* was manipulated between subjects, which means that there were 20 subjects per variant of the *n*-back task (1-, 2-, and 3-back). Each subject performed the task in both the verbal and the visuospatial modality in a counterbalanced order.

Materials and procedure.

Verbal condition. The materials and procedure were held as close as possible to those in Jonides et al. (1997). For each condition of *n*, four different *n*-back lists were constructed. The order of the lists was counterbalanced across participants. Each list consisted of $45 + n$ letters (i.e., 46, 47, and 48 in the 1-, 2-, and 3-back variants, respectively), which were randomly chosen from the 20 consonants of the alphabet. The consonants were randomly presented in upper or lower case to ensure that subjects matched verbal contents and not visual features. The 45 trials in each *n*-back list contained 15 match trials (i.e., consonant matched the consonant presented *n* positions before) and 30 mismatch trials (i.e., consonant did not match the consonant presented *n* positions before). For each variant of *n*, six of the 30 mismatches were lure trials. In the 1-back variant, three lures were in the $n + 1$ position (e.g., G-*B*-*T*-B; see Figure 1) and three lures were in the $n + 2$ position (e.g., **B**-G-*T*-B). For the 2-back variant, three lures were in the $n + 1$ position (e.g., **B**-*T*-G-B) and three lures were in the $n − 1$ position (e.g., G-*T*-**B**-B). In the 3-back condition, three lures were in the $n − 1$ position (e.g., *G*-**B**-T-B) and three lures were in the $n − 2$ position (e.g., *G*-T-**B**-B). There were no other lures in the list within a 5-back range from the target item. Furthermore, an item could not be more than one type of lure at the same time (e.g., in the 3-back sequence B-*L*-F-B-*B*, the consonant B is both an $n − 2$ and an $n + 1$ lure). No lures were included on the *n*-back matching trials.

The consonants were presented one at a time centered on a 15-in. computer monitor, in 48-point bold Times New Roman font. Subjects were seated approximately 40 cm from the screen. Each consonant remained on the screen for 500 ms, followed by a fixed interstimulus interval (ISI) of 2,500 ms. Subjects were required to indicate as fast and accurately as possible whether a presented item matched the one presented *n* positions before by pressing the *left* (i.e., mismatch) or *right* key (i.e., match) on a response box. They were not informed about the occurrence of lures. Before starting an experimental condition, each participant practiced the condition of *n* that he or she had been assigned to. Practice blocks of 20 trials, which did not include lures, proceeded until an accuracy level of at least 80% was reached.

Visuospatial condition. In the visuospatial modality condition, black dots were presented sequentially in a 5×5 black grid (see Figure 2) and subjects had to indicate whether the dot appeared in the same location as the dot presented *n* positions before. The 5.9×5.9-in. grid was centered on the screen, with the 0.79-in. diameter black dot occurring in one of the 25 locations in the grid. In all other respects, the procedure was the same as in the verbal modality condition.
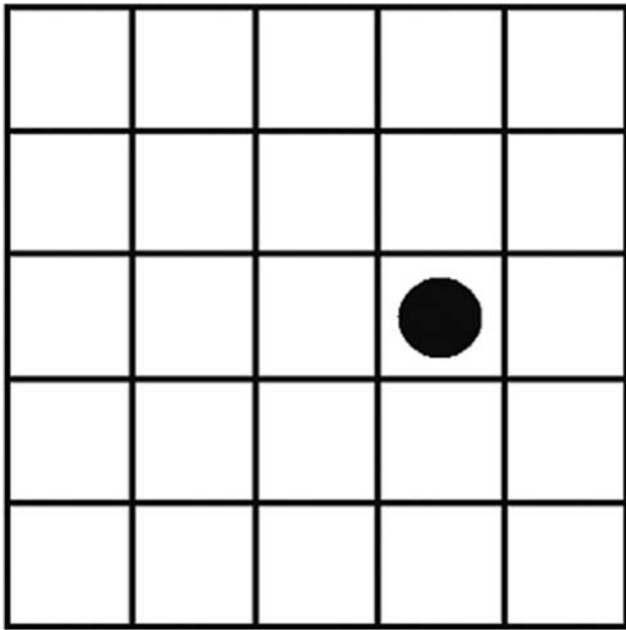
Figure 2. Illustration of the stimulus and display characteristics of the visuospatial *n*-back condition in Experiment 1.

Results

To increase the readability of this section, we present only the planned comparisons, whereas the statistics for the omnibus analyses of variance (ANOVAs) are summarized in Appendix A. Planned comparisons were performed only when the relevant omnibus ANOVA was significant.

The mean reaction times (RTs; correct trials only) for the three variants of the *n*-back task are displayed in Figure 3 as a function of the modality (verbal and visuospatial) and the different trial types (match, mismatch, and lure trials). Table 1 represents the accuracy data for the same conditions. Note that whenever we use the term *mismatch trials*, we only refer to *neutral* mismatches, and not to lure trials, despite the fact that lure trials are essentially also mismatch trials as they also require a "no" response.

1-back. Averaged across the verbal and the visuospatial modalities, the 1-back matches were performed faster than the mismatches, $F(1, 19) = 9.74$, $\eta_p^2 = .34$, $p < .01$. $n + 1$ lures yielded slower RTs compared with mismatches, $F(1, 19) = 31.70$, $\eta_p^2 = .62$, $p < .001$, whereas $n + 2$ lures did not, $F < 1$. The difference between both lure trials was also reliable, $F(1, 19) = 16.69$, $\eta_p^2 = .47$, $p < .01$. These differences did not depend on modality (see Appendix A).

The accuracy data showed the same effects. Averaged across the verbal and the visuospatial modalities, we found that mismatches were more accurate than matches, $F(1, 19) = 14.31$, $\eta_p^2 = .43$, $p < .01$. The $n + 1$ lures produced lower accuracies than the mismatches, $F(1, 19) = 6.36$, $\eta_p^2 = .25$, $p < .05$, whereas the $n + 2$ lures did not, $F < 1$. The difference between both types of lures was reliable, $F(1, 19) = 10.57$, $\eta_p^2 = .36$, $p < .01$. Again, these differences did not depend on modality (see Appendix A).

2-back. Averaged across the verbal and the visuospatial modalities, the 2-back matches were performed faster than the mis-

matches, $F(1, 19) = 55.31$, $\eta_p^2 = .74$, $p < .001$, and both the $n + 1$ and the $n - 1$ lures slowed performance compared with the mismatches, $F(1, 19) = 14.41$, $\eta_p^2 = .43$, $p < .01$, and, $F(1, 19) = 92.31$, $\eta_p^2 = .83$, $p < .001$, respectively. The difference between both types of lure trials was also significant, $F(1, 19) = 4.40$, $\eta_p^2 = .19$, $p < .05$. Similar to the 1-back variant, the lure RT effects were not influenced by modality.

Based on the average across the verbal and visuospatial modalities, mismatches were more accurate than matches, $F(1, 19) = 23.03$, $\eta_p^2 = .55$, $p < .001$. Both the $n + 1$ lures and the $n - 1$ lures

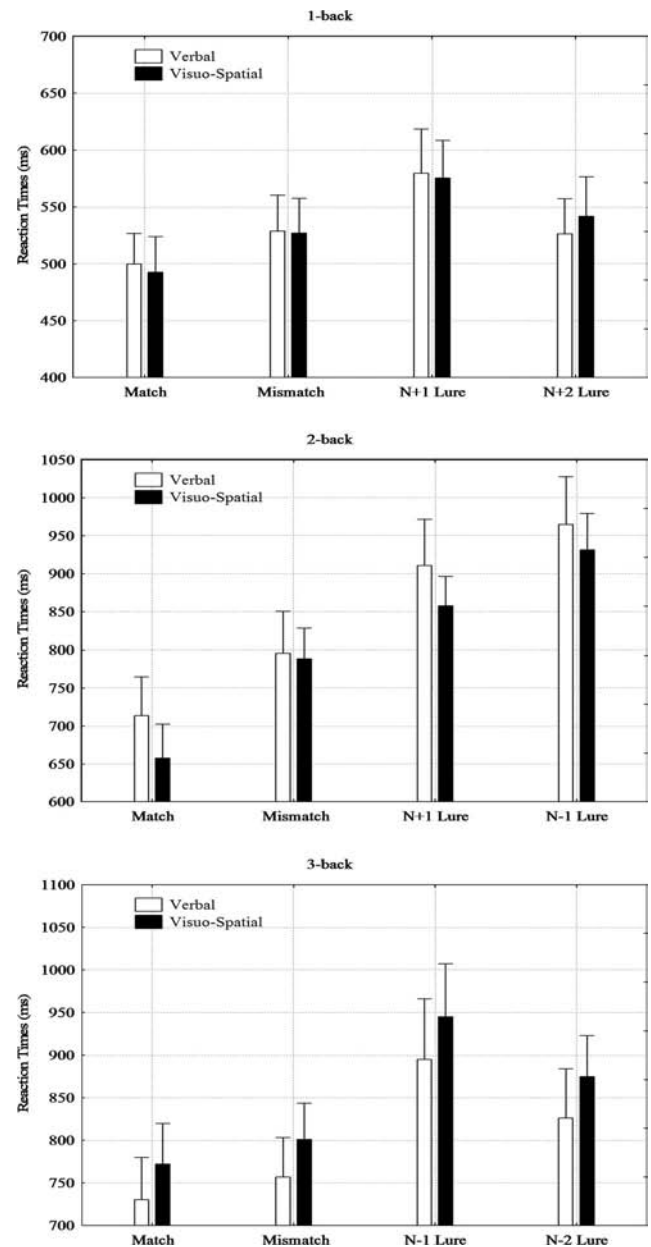*Figure 3.* Mean reaction times for the three variants of the *n*-back task, as a function of modality (verbal and visuospatial) and trial type (match, mismatch, and lure trials) in Experiment 1. Vertical bars denote standard errors.

Table 1

Mean Accuracy and Standard Deviations for the Match, Mismatch, and Lure Trials in the 1-, 2-, and 3-Back Variants of the Verbal and Visuospatial n-*Back Task, Experiment 1*

Task	Trial type			
Verbal				
1-back	Match	Mismatch	$n + 1$ lure	$n + 1$ lure
	90 (12)	98 (02)	94 (06)	98 (04)
2-back	Match	Mismatch	$n + 1$ lure	$n - 1$ lure
	89 (09)	98 (03)	88 (11)	74 (15)
3-back	Match	Mismatch	$n - 1$ lure	$n - 2$ lure
	84 (10)	93 (08)	87 (12)	91 (11)
Visuospatial				
1-back	Match	Mismatch	$n + 1$ lure	$n + 2$ lure
	87 (12)	96 (02)	95 (07)	98 (05)
2-back	Match	Mismatch	$n + 1$ lure	$n - 1$ lure
	88 (09)	94 (03)	92 (11)	68 (19)
3-back	Match	Mismatch	$n - 1$ lure	$n - 2$ lure
	74 (19)	89 (08)	77 (14)	91 (08)

Note. Accuracy scores are percentages; standard deviations appear within parentheses.

produced lower accuracies compared with the mismatch trials, $F(1, 19) = 11.51$, $\eta_p^2 = .38$, $p < .01$, and, $F(1, 19) = 62.19$, $\eta_p^2 = .77$, $p < .001$, respectively. The difference between both types of lures was also reliable, $F(1, 19) = 44.69$, $\eta_p^2 = .70$, $p < .001$.

3-back. Averaged across the verbal and the visuospatial modalities, there was no reliable difference between 3-back matches and mismatches, $F(1, 19) = 1.99$, $\eta_p^2 = .09$, $p > .10$. Both the $n - 1$ and the $n - 2$ lure trials significantly delayed performance compared with the mismatches, $F(1, 19) = 26.42$, $\eta_p^2 = .58$, $p < .001$, and $F(1, 19) = 13.55$, $\eta_p^2 = .42$, $p < .01$, respectively. The difference between both kinds of lure trials was significant, $F(1, 19) = 5.03$, $\eta_p^2 = .21$, $p < .05$. Again, these lure effects were not influenced by modality.

Planned comparisons on the accuracy data showed that there was a main effect of trial type in the verbal domain, $F(3, 57) = 6.55$, $\eta_p^2 = .26$, $p < .001$, and in the visuospatial domain, $F(3, 57) = 13.64$, $\eta_p^2 = .42$, $p < .001$. Matches were less accurate than mismatches with verbal materials, $F(1, 19) = 17.91$, $\eta_p^2 = .48$, $p < .001$, but also with visuospatial materials, $F(1, 19) = 20.29$, $\eta_p^2 = .52$, $p < .001$. $n - 1$ lures affected accuracy in the verbal domain, $F(1, 19) = 5.89$, $\eta_p^2 = .24$, $p < .05$, as well as in the visuospatial domain, $F(1, 19) = 44.12$, $\eta_p^2 = .70$, $p < .001$. Finally, lures in the $n - 2$ position did not differ reliably from mismatch trials, neither in the verbal domain nor in the visuospatial domain, Fs < 1.

1-, 2-, and 3-back comparison. Finally, we analyzed the lure interference effects (in the RT data) for comparable conditions across the different variants of the set size n. First, we conducted a 2 (modality: verbal and visuospatial) $\times$ 2 (set size: 1- and 2-back) $\times$ 2 (trial type: mismatch and $n + 1$ lure) ANOVA, and we found that the $n + 1$ lure effect was comparable across the 1-back and 2-back variants of the n-back task, $F(1, 19) = 2.54$, $\eta_p^2 = .12$, $p > .10$. This finding was not influenced by stimulus modality, $F(1, 19) = 1.94$, $\eta_p^2 = .09$, $p > .10$. Second, we performed a 2 (modality: verbal and visuospatial) $\times$ 2 (set size: 2- and 3-back) $\times$ 2 (trial type: mismatch and $n - 1$ lure) ANOVA, which indicated that also the $n - 1$ lure effect was similar for the 2- and 3-back variant of the task, $F < 1$, independent of stimulus

modality, $F < 1$. Overall, these analyses show that the lure interference effects are relatively comparable across the different n-back conditions. A similar pattern of results was observed in the accuracy data.

Discussion

The aim of Experiment 1 was to provide a detailed description of the different kinds of lure interference effects that occur in the n-back updating task. In the 1-back variant of the task, we compared the proactive interference of lures in the $n + 1$ and $n + 2$ positions. The results showed that only the lure in the $n + 1$ position affected performance. An item that has become irrelevant to the task can be discarded from the DA region of WM, so its level of activation, or familiarity, decreases. If the item has just been discarded ($n + 1$ position), its familiarity is still high enough to cause interference. After a longer period ($n + 2$ position), however, the item has become too unfamiliar to produce interference, at least to a measurable extent. The 2-back version of the task allowed us to directly compare interference caused by discarded items that are no longer relevant ($n + 1$ lure) and recent items that are relevant for future processing ($n - 1$ lure). We observed that the $n - 1$ lure interference was larger than the $n + 1$ lure interference, again in line with the idea that the amount of interference is a function of the familiarity (and relevance) of an item. In the 3-back variant of the task, we compared interference caused by lures in the $n - 1$ and $n - 2$ positions, and observed a reliably larger lure effect in the $n - 1$ position. It is interesting to see that the oldest item, in the $n - 1$ position, actually produces more interference than the more recent item in the $n - 2$ position. One possible explanation for this finding is that an item in the $n - 1$ position is even more familiar because it has been rehearsed more often than an item in the $n - 2$ position. The relation between rehearsal and familiarity in the n-back task was further investigated in Experiment 2.

The same patterns of lure interference were seen in the accuracy data. The finding that people make more errors on a lure trial is consistent with the view that there is competition between familiarity and recollection, and that people still fail to rely on recollection on a minority of the lure trials.

Finally, we also showed that the patterns of lure interference that were found in the different variants of the n-back task were comparable for verbal and visuospatial stimulus materials. This means that we are not merely looking at modality-specific effects, which is an important prerequisite for drawing conclusions about the operation of a central interference control mechanism involved in WM updating.

Experiment 2

The results from the 3-back task used in Experiment 1 showed that an older item (i.e., in the $n - 1$ position) causes more lure interference than a more recent item (i.e., in the $n - 2$ position). Knowing that familiarity decreases over time (see also results of the 1-back task in Experiment 1), stronger interference from an older item is at first sight somewhat at odds with the hypothesis that lure interference is familiarity-based. However, lure items in the $n-$ position need to be maintained in the DA region because these items still need to be recognized once they have arrived in the n-back position. It is therefore quite unlikely that subjects in

an *n*-back task just allow the not-yet-relevant items in memory to degrade, and hence let the familiarity of the items decrease. It seems more plausible that subjects alter the level of activation of an item in memory, for instance, through rehearsal (Page & Norris, 1998), by which the degree of familiarity becomes affected (Gardiner et al., 1994).[2] To test this hypothesis directly, we varied the interval between the presentation of successive items in the *n*-back task given the knowledge that with a longer ISI subjects typically rehearse memorized materials more often (cf. Bunting et al., 2006; Glanzer & Cunitz, 1966; Hockey, 1973) to protect the stored contents against forgetting. Consequently, we predicted that familiarity of items would be higher for the long ISI condition than the short ISI condition because of rehearsal (Gardiner et al., 1994). If interference from lure trials reflects a familiarity-based erroneous response tendency, then lure interference effects will be larger for longer ISIs than for shorter ISIs. We tested this hypothesis in Experiment 2.

Method

Subjects and design. Twenty-six university students (12 women; mean age = 18 years, range = 18–23 years) participated for course requirements. None of them had participated in Experiment 1. We used a 2 (ISI: short and long) × 4 (trial type: match, mismatch, *n* + 1 lure, and *n* − 1 lure) within-subjects design. The order of the ISI conditions was counterbalanced across subjects.

Materials and procedure. The same verbal contents as in Experiment 1 were used. We administered only 2-back lists in this experiment because this version of the *n*-back task enabled us to investigate *n*+ as well as *n*− lures. The ISI was 2,000 ms in the short ISI condition and 4,000 ms in the long ISI condition. The two different ISI conditions were blocked. In all other respects, the materials and procedure were identical to those in Experiment 1.

Results

The mean RTs (correct trials only) for the different trial types in the short and long ISI conditions are illustrated in Figure 4. The
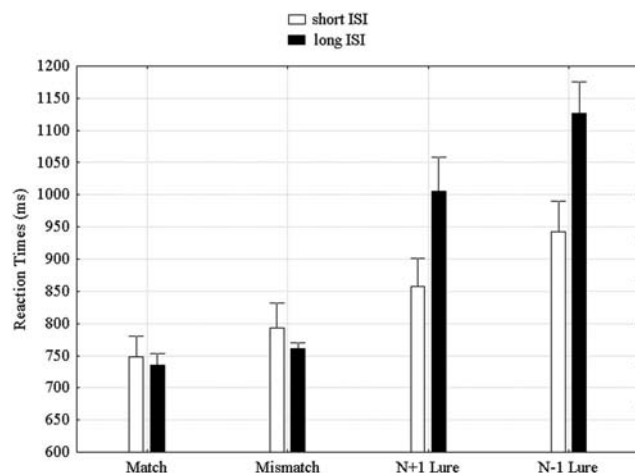

Figure 4. Mean reaction times for the different trial types (match, mismatch, and lure trials) in the short and long interstimulus interval (ISI) conditions in Experiment 2. Vertical bars denote standard errors.

statistics for the omnibus analyses of variance are summarized in Appendix B; only the relevant planned comparisons are summarized below. Matches were performed faster than mismatches, $F(1, 25) = 3.96$, $\eta_p^2 = .14$, $p = .06$, and this difference was similar for both ISI conditions, $F < 1$. We observed a reliable $n + 1$ lure effect (i.e., difference between mismatch and $n + 1$ lure trials) in both the short ISI, $F(1, 25) = 4.70$, $\eta_p^2 = .16$, $p < .05$, and the long ISI conditions, $F(1, 25) = 56.35$, $\eta_p^2 = .69$, $p < .001$, although this effect appeared to be stronger in the long ISI condition, $F(1, 25) = 14.45$, $\eta_p^2 = .37$, $p < .001$. Analogously, we found a reliable $n - 1$ lure effect, both in the short, $F(1, 25) = 24.64$, $\eta_p^2 = .50$, $p < .001$, and the long ISI conditions, $F(1, 25) = 98.56$, $\eta_p^2 = .80$, $p < .001$, and this effect was again stronger in the long ISI condition, $F(1, 25) = 33.35$, $\eta_p^2 = .57$, $p < .001$. Finally, the RT data showed that $n + 1$ lure trials were performed faster than $n - 1$ lure trials, $F(1, 25) = 11.01$, $\eta_p^2 = .31$, $p < .01$, and this difference was similar for both ISI conditions, $F < 1$.

The mean accuracy per trial type and per ISI condition is displayed in Table 2 (see Appendix B for omnibus ANOVA results). Averaged across both ISI conditions, the data showed that 2-back matches and mismatches yielded a comparable level of accuracy, $F(1, 25) = 2.34$, $\eta_p^2 = .08$, $p > .10$. Compared with the mismatch trials, the $n + 1$ lures, $F(1, 25) = 7.39$, $\eta_p^2 = .23$, $p < .05$, as well as the $n - 1$ lures, $F(1, 25) = 7.84$, $\eta_p^2 = .43$, $p < .01$, produced a reliable decrease in accuracy.

Discussion

Consistent with the idea that rehearsal increases familiarity, we found larger interference effects in the long ISI condition. It thus seems that rehearsal plays a somewhat ambiguous role in updating tasks. On the one hand, it protects the contents of memory against forgetting over longer periods of time. On the other hand, it increases the familiarity of the updated items, by which their potential to cause interference (e.g., when they appear in the *n*-back lure position) increases. Oberauer (2009) aptly described this paradox by putting that, in updating tasks, WM has to fight against conflicting tendencies that in fact originate from WM itself. Rehearsal in the *n*-back updating task may thus counteract forgetting, but it does not seem to be a good strategy to counteract lure interference; quite the contrary.

There is an ongoing debate on the role of rehearsal in WM, for example, as a strategy to counteract forgetting (e.g., Oberauer & Lewandowsky, 2008). It can be questioned whether rehearsal is indeed necessary to explain the current results and whether the higher familiarity for longer ISIs could be merely a consequence of longer and thus better encoding. Our data do not allow us to differentiate between both accounts, but our conclusions about the role of stimulus familiarity in lure interference stand relatively independent of the question of whether it is quality of encoding or rehearsal that drives familiarity.

---

[2] Despite the fact that the term *familiarity* belongs to LTM in the first place, we would like to re-emphasize that we refer to *short-term* familiarity here. From that view, we could also have chosen to replace *familiarity* by its short-term counterpart *recency*. However, we decided not to do so because this study was based on the idea that long-term concepts like familiarity and recollection are also applicable to WM (e.g., Goethe & Oberauer, 2008; Oztekin & McElree, 2007).

Table 2

Mean Accuracy and Standard Deviations for the Match, Mismatch, and Lure Trials in the Short and Long Interstimulus Interval (ISI) Conditions of the 2-Back Task, Experiment 2

ISI	Match	Mismatch	$n + 1$ lure	$n - 1$ lure
Short	84 (14)	87 (09)	81 (15)	79 (22)
Long	89 (11)	93 (07)	76 (19)	72 (27)

Note. Accuracy scores are percentages; standard deviations appear within parentheses.

The RT differences between the short and long ISIs were visible only on lure trials, but not on the mismatch and match trials. Because familiarity and recollection provide congruent information on match and mismatch trials, one might also have expected that higher familiarity leads toward faster RTs on match and mismatch trials. However, the current data thus show a large impact of increased familiarity on lure trials, but not on match and mismatch trials. Such a pattern of results implies that the costs of incongruent information on RT are greater than the benefits from congruent information, a phenomenon that is typically found in other interference paradigms as well (see, e.g., Verbruggen & Logan, 2009, for a more detailed account).

Finally, we discuss the role of spatial versus temporal context in the binding of items to a position in WM. The observation that the lure effects are larger with longer ISIs suggests that the interference does not originate from a decreased temporal discriminability between item n and its neighbors; otherwise, one would rather expect the interference to decrease as time elapses. Also the earlier observation (see Experiment 1) that in the 3-back task, lure items in position $n - 1$ cause more interference than lure items that are more recent but stand one position farther away from n, that is, $n - 2$, is more in line with a spatial- than with a temporal-positional view. Lack of spatial–positional discriminability typically leads to confusion of item positions, as reflected in so-called transposition errors (Lewandowsky, Brown, Wright, & Nimmo, 2006; Ng & Maybery, 2002). Such transposition errors are usually seen in classic serial recall tasks (e.g., Henson, 1999), and this brings us to another important theoretical issue in clarifying the nature and origins of interference during WM updating. Although we acknowledge that transposition errors may occur in the n-back updating task, just like they do in serial recall tasks, we believe that proactive interference in the n-back updating task does not exclusively originate from transposition errors. This point was more directly addressed in Experiment 3.

Experiment 3

We propose that the n-back lure interference effects are driven by the familiarity of the lures and are counteracted by recollection. Updating tasks are particularly prone to interference because weak bindings between items and their positional context hamper recollection and hence fail to efficiently counteract familiarity-based processing. However, weak bindings can also lead to transposition errors, irrespective of any misleading familiarity signal. This is a theoretically important distinction in understanding the nature and origins of interference during updating. The reason is that items that are in the DA region of WM are known to automatically activate items in LTM that match them or that are associated with them, and these items in turn compete for being retrieved in WM (Oberauer, 2009). Hence, if n-back lure interference is driven by item familiarity, then any item that is activated (at least to a retrieval threshold) in LTM would have the potential to cause proactive interference, even if that item has never been presented before. If, by contrast, lure interference is a consequence of transposition errors among items that are weakly bound in the DA region ($n - 1$ lures) or that are being unbound from the DA region ($n + 1$ lures), then only those items would have the potential to cause proactive interference.

To address this issue, we conducted an experiment using a 2-back version of the n-back task with Dutch nouns as stimuli (e.g., bike-*knife*-house-*knife*), including lure trials (e.g., **knife**-*neck*-house-*knife*). Most important, we also included semantic lures in our design, which are lure trials that are semantically related to the target word (e.g., **fork**-*neck*-house-*knife*). The rationale behind this manipulation is the following: If n-back lure interference is a consequence of positional errors among items that are continuously bound and unbound from the DA region of WM, then we predict interference from the identical lures, but not from the semantic lures, because the latter have never been positioned in the DA region. If, by contrast, the lure effects originate (at least partly) from item familiarity, then we also expect interference from items that have never been bound in the DA region but that are active merely through semantic relatedness with items in the DA region. Hence, we predict interference from both the identical and semantic lures. In this experiment, we focused on $n + 1$ lures only because these lures are in the process of being unbound from the DA region. Measuring interference in the DA region merely through semantic relatedness with those items would therefore provide strong support for our familiarity account of lure interference.

Method

Subjects and design. Twenty-four university students (18 women; mean age = 18 years, range = 17–31 years) participated for course requirements. None of them had participated in any of the previous experiments. There was one within-subject variable, trial type (match, mismatch, identical $n + 1$ lure, and semantic $n + 1$ lure).

Materials and procedure. Four different 2-back lists were administered to the subjects, each consisting of 45 + 2 trials. The stimuli were Dutch 1- or 2-syllable nouns taken from De Deyne and Storms (2008), who obtained semantic association norms for 1,424 Dutch words. Each 2-back list contained 20 mismatch trials, 15 match trials, five identical $n + 1$ lures, and five semantic $n + 1$ lures. An example of a semantic lure trial is . . . **pear**-*mouth*-hat-*apple* . . . . The study of De Deyne and Storms showed asymmetric priming effects (e.g., stronger semantic priming from *peer* to *apple* than vice versa). We placed the stimulus with the strongest priming properties in lure position (*pear* in the example) and the most strongly semantically associated stimulus (*apple* in the example) in target position. The length (number of syllables) of the two word stimuli between lure and target were matched across the different trial types. The stimuli were presented visually (48-point bold Times New Roman font) for 1,500 ms, after which a fixation cross remained centered on the screen for another 1,500

ms (thus, the ISI was 3,000 ms). The experiment was preceded by two practice lists. The other procedural characteristics were identical to those in the previous experiments. The entire procedure lasted approximately 40 min.

Results

The mean RTs (correct trials only) for the different trial types are illustrated in Figure 5. In the ANOVA, we observed a significant main effect of trial type, $F(3, 69) = 15.96$, $\eta_p^2 = .41$, $p < .001$. Planned comparisons showed that matches and mismatches were performed equally fast, $F < 1$. Both the identical lures and the semantic lures led to slower RTs compared with the mismatch trials, $F(1, 23) = 23.72$, $\eta_p^2 = .51$, $p < .001$, and $F(1, 23) = 22.82$, $\eta_p^2 = .50$, $p < .001$, respectively, but the interference effect of the semantic lures was smaller than that of the identical lures, $F(1, 23) = 8.82$, $\eta_p^2 = .28$, $p < .01$.

The mean accuracy was .87 ($SD = .07$) for the match trials, .98 ($SD = .02$) for the mismatch trials, .77 ($SD = .17$) for the identical lures, and .94 ($SD = .06$) for the semantic lures. The ANOVA revealed a significant main effect of trial type, $F(3, 69) = 21.46$, $\eta_p^2 = .48$, $p < .001$. Match trials yielded a lower level of accuracy than mismatch trials, $F(1, 23) = 60.30$, $\eta_p^2 = .72$, $p < .001$. Compared with the mismatch trials, the identical lures, $F(1, 23) = 35.36$, $\eta_p^2 = .61$, $p < .001$, as well as the semantic lures, $F(1, 23) = 9.12$, $\eta_p^2 = .28$, $p < .01$, produced a reliable decrease in accuracy, but the mean accuracy for the semantic lures was reliably higher than that for the identical lures, $F(1, 23) = 20.46$, $\eta_p^2 = .47$, $p < .001$.

Discussion

The results of Experiment 3 demonstrate that not only identical lures but also lures that are merely semantically related to the items that are updated in WM impair performance in the n-back task. This finding indicates that lure interference in the n-back updating task does not (only) reflect item transposition errors. Our semantic lure effects in fact show that at least part of the interference is achieved through the spreading of activation from WM to LTM and back again, a finding that offers nice empirical support for Oberauer's (2009; see also Cowan, 1988, 1995) design for WM.
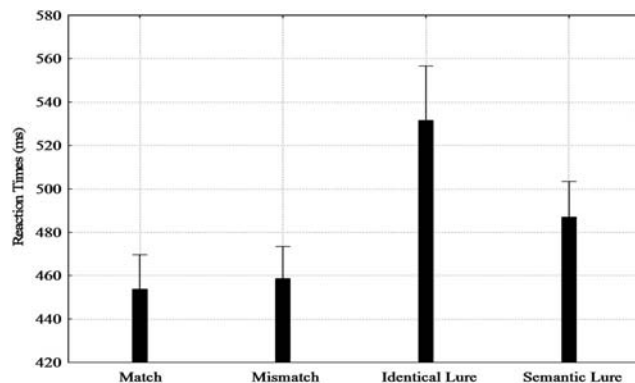

Figure 5. Mean reaction times for the different trial types (match, mismatch, identical $n + 1$ lure, and semantic $n + 1$ lure) in Experiment 3. Vertical bars denote standard errors.

It is particularly important to note that semantic lure interference effects have also been investigated before in a modified version of the Sternberg task, called the Recent Negatives Task (RNT; e.g., Bartha, Martin, & Jensen, 1998; Hamilton & Martin, 2007), which does not involve continuous updating. In the RNT, subjects are asked to memorize a number of items, such as letters or words. Subsequently, a probe item is presented and subjects have to decide whether or not the probe item occurred in the original stimulus set. On each consecutive trial, a new stimulus set is presented, so memory can be completely cleared between trials. Occasionally, lure items are presented that are not part of the current stimulus set n but were part of the set presented on the previous trial $n - 1$. This proactive interference effect is very similar to the $n + 1$ lure interference effect observed in the n-back task. It is interesting that Hamilton and Martin (2007) found that semantic lures did not affect performance in the RNT in healthy individuals (see also Bartha et al., 1998). Hamilton and Martin found semantic lure effects only in a brain-damaged patient who had problems with binding information in short-term memory. Combined with the results of Experiment 3, these findings nicely accommodate the theoretical position that lure interference depends on the strength of bindings in WM. Updating tasks are more sensible to this kind of interference because they require more flexible binding than pure WM maintenance tasks. This high sensibility to interference is evidenced by the observation that virtually any item that is active in LTM (at least to retrieval level) gains access to the DA region, from where it interferes with ongoing cognitive processing. In other words, updating can be held responsible for an important leakage of activation from LTM into the DA region of WM (i.e., proactive interference).

Experiment 4

The data from the previous experiments show that although responses are slower and less accurate on n-back lure trials, subjects still succeed in counteracting the lure interference and produce a correct response on a majority of trials. We propose that the competition between the error prone familiarity-matching process and the analytic recollection process is diminished by prioritizing recollection over familiarity matching through postconflict, top-down behavioral adjustments. The conflict monitoring hypothesis (CMH; Botvinick, Braver, Barch, Carter, & Cohen, 2001) is an influential theoretical framework that provides a detailed and formal description of how a controlled process can be prioritized over an automated process. The CMH assumes that the cognitive system holds a cognitive control loop, which monitors human information processing for suboptimal performance or interference. When interference is detected, it can be counteracted by top-down biasing of the task-relevant processing areas.

We propose that a functionally similar mechanism is responsible for controlling interference when updating WM. Familiarity matching is an automatic process that does not require binding of items to WM, and that is therefore also more error-prone, especially on lure interference trials. Recollection is a controlled process that relies on strong item-to-position bindings and that is therefore more accurate. To test this control-loop hypothesis, we used a well-established effect that is a behavioral manifestation of the operation of the cognitive control loop put forward in the

CMH. This effect is known as the *Gratton effect* (Gratton, Coles, & Donchin, 1992).

Gratton et al. (1992) observed that in the Eriksen flanker task, which is an interference task, the amount of interference on a trial is smaller when there was also interference on the previous trial compared with when there was no interference on the previous trial. This effect has been replicated in other conflict tasks, such as the Stroop task (Kerns et al., 2004) and the Simon task (Stürmer, Leuthold, Soetens, Schröter, & Sommer, 2002). The leading theoretical account of the Gratton effect is that—after careful control for priming effects—it reflects top-down behavioral adjustments following cognitive interference (e.g., Botvinick et al., 2001; Notebaert, Gevers, Verbruggen, & Liefooghe, 2006; Verbruggen, Notebaert, Liefooghe, & Vandierendonck, 2006; Verguts & Notebaert, 2008, 2009). According to the CMH, interference is controlled by biasing the controlled processing route so that it can override the erroneous response tendencies elicited by the automatic route. On incongruent trials, there is interference and the controlled process becomes prioritized; consequently, a subsequent incongruent trial will produce less interference (i.e., the Gratton effect). By contrast, on a congruent trial, there is no interference and there is no need for prioritization of the controlled process; consequently, a subsequent incongruent trial will produce more interference.

To test the hypothesis that interference during updating is regulated in a functionally similar way as described by the CMH, that is, by biasing recollection over familiarity matching, Experiment 4 investigated whether lure interference in the *n*-back updating task is also subject to a Gratton-like conflict adaptation effect.[3]

We tested whether the previous trial type influenced the difference between lure trials and mismatch trials. More specifically, we focused on four trial types in a 2-back task: MM trials (i.e., a mismatch following a mismatch), ML trials (a lure following a mismatch), LM trials (a mismatch following a lure), and LL trials (a lure following a lure).[4] On the basis of the hypothesis that interference during updating is dealt with by a top-down adaptation mechanism, we predicted that the interference of both the $n + 1$ and the $n - 1$ lures would be smaller following another lure trial (LM vs. LL) than following a mismatch trial (MM vs. ML).

Method

Subjects, materials, and procedure. Twenty-eight university students (10 men; mean age = 20 years, range = 17–27 years) were paid €10 for participation. None of them had participated in any of the previous experiments. The materials and procedure were kept identical to those in Experiment 1, unless mentioned otherwise. We only used the 2-back variant of the *n*-back task with lures in the $n + 1$ and $n - 1$ positions. In total, 10 lists of 45 trials each were constructed. Of these 450 trials, 125 (28%) were 2-back match trials (e.g., F-*B*-L-*B*). The remaining 325 (72%) trials were relevant for the conflict adaptation analyses; they are detailed in Table 3. Before the 10 experimental lists, subjects practiced forty-five 2-back trials (i.e., equivalent to one list). The task took approximately 1 hr to complete.

Table 3

Composition of the 325 Trials (2-Back Only) Used for the Conflict Adaptation Analyses, Experiment 4

Type	$n + 1/n - 1$ lure	Example	Amount
MM		F−R−*L*−B−*T*	187 (42%)
ML	$n + 1$	F−R−*L*−B−*R*	28 (6%)
	$n - 1$	F−R−*L*−T−*T*	28 (6%)
LM	$n + 1$	B−R−*L*−B−*T*	23 (5%)
	$n - 1$	F−R−*B*−B−*T*	19 (4%)
LL[a]	$n + 1\ n + 1$	B−T−*L*−B−*T*	13 (3%)
	$n + 1\ n - 1$	B−T−*F*−B−*B*	12 (3%)
	$n - 1\ n + 1$	F−T−*B*−B−*T*	14 (3%)

Note. MM = mismatch following mismatch; ML = lure following mismatch; LM = mismatch following lure; LL = lure following lure.
[a] We did not use LL trials with two subsequent $n - 1$ lures because these (e.g., F−T−*B*−B−*B*) require a 2-back match response.

Results

Reaction times. RTs were calculated for the correct trials only. In a first ANOVA, we tested the effect of trial type (match, mismatch, $n + 1$ lure, and $n - 1$ lure) to see whether we could replicate our previous findings with the 2-back task. For this analysis, we looked only at isolated $n + 1$ and $n - 1$ lure trials (i.e., not preceded by another lure trial) for reasons of comparability with the previous experiments. We observed a main effect of trial type, $F(3, 81) = 22.85$, $\eta_p^2 = .46$, $p < .001$. The difference between 2-back match ($M = 615$, $SD = 193$) and mismatch trials ($M = 650$, $SD = 222$) was not reliable, $F(1, 27) = 2.15$, $\eta_p^2 = .07$, $p > .10$. The $n + 1$ lure trials ($M = 733$, $SD = 254$) and $n - 1$ lures ($M = 786$, $SD = 249$) produced slower RTs compared with mismatch trials, $F(1, 27) = 18.55$, $\eta_p^2 = .41$, $p < .001$, and $F(1, 27) = 28.49$, $\eta_p^2 = .51$, $p < .001$, respectively. The difference between lure types was also significant, $F(1, 27) = 9.96$, $\eta_p^2 = .27$, $p < .01$.

Next, we analyzed the conflict adaptation results. On theoretical grounds, we did not expect a difference in conflict adaptation between $n + 1$ and $n - 1$ lures because they are both assumed to reflect a competition between familiarity and recollection. This was confirmed by a preliminary inspection of the data. Therefore, the data from the different lure types were collapsed in the following analysis. Specifically, this means that the ML trials and the LM trials contained the data from the $n + 1$ and $n - 1$ lure trials. The LL data comprised trials with two consecutive $n + 1$ lures, trials with an $n + 1$ lure preceding an $n - 1$ lure, and trials with an $n - 1$ lure preceding an $n + 1$ lure. As indicated above, trials with two consecutive $n - 1$ lures were not included in the analysis because these require a "yes" response. Finally, the MM trials contained no lures at all.

---

[3] We prefer to use the term *Gratton-like* because the classical Gratton effect has been described in a context of conflict tasks that do not involve exactly the same types of trials as those involved in the *n*-back task.

[4] Because *n*-back match trials require a "yes" response, whereas mismatch and lure trials require a "no" response, we chose not to include *n*-back match trials in the conflict adaptation analysis to prevent the effects of interest from becoming confounded with baseline differences in RTs between yes and no responses.

We conducted an ANOVA with two within-subject factors (lure interference on the current trial and lure interference on the previous trial) on the mean RTs. The results of this analysis are displayed in the upper half of Figure 6. We observed a significant overall lure effect, $F(1, 27) = 18.19$, $\eta_p^2 = .40$, $p < .001$, but most important, the data show that this lure interference effect depended on whether or not interference occurred on the previous trial, $F(1, 27) = 34.85$, $\eta_p^2 = .48$, $p < .001$. We found that there was a reliable lure interference effect after a mismatch trial, $F(1, 27) = 51.22$, $\eta_p^2 = .65$, $p < .001$, but this effect completely disappeared when the preceding trial was also a lure trial, $F < 1$. The disappearance of the lure interference effect when the previous trial was also a lure trial is explained by a strong tendency to become faster on interference trials, $F(1, 27) = 3.79$, $\eta_p^2 = .12$, $p = .06$, and to become slower on mismatch trials, $F(1, 27) = 44.84$, $\eta_p^2 = .62$, $p < .001$.

Accuracy data.   We first tested the effect of trial type (match, mismatch, $n + 1$ lure, and $n - 1$ lure) on accuracy in the 2-back task by looking only at isolated $n + 1$ and $n - 1$ lure trials (i.e., the types ML and LM) for reasons of comparability with the previous experiments. We observed a main effect of trial type, $F(3, 81) = 14.45$, $\eta_p^2 = .35$, $p < .001$. The difference between 2-back match ($M = .86$, $SD = .12$) and mismatch trials ($M = .95$, $SD = .07$) was also reliable, $F(1, 27) = 14.10$, $\eta_p^2 = .34$, $p < .001$. The $n + 1$ lure trials ($M = .88$, $SD = .08$) yielded lower accuracy scores com-pared with mismatch trials, $F(1, 27) = 30.52$, $\eta_p^2 = .53$, $p < .001$. The $n - 1$ lures ($M = .82$, $SD = .13$) also caused a reliable decrease in accuracy compared with the mismatch trials, $F(1, 27) = 34.01$, $\eta_p^2 = .56$, $p < .001$. The difference between both types of lures was also reliable, $F(1, 27) = 9.63$, $\eta_p^2 = .26$, $p < .01$.

We then conducted an ANOVA with two within-subject factors (lure interference on the current trial and lure interference on the previous trial) on the mean accuracy rates. The aim of this analysis was to investigate whether the top-down conflict adaptation effects were also reflected in the accuracy data. The results of this analysis are represented in the lower half of Figure 6. We observed a significant overall lure effect, $F(1, 27) = 21.99$, $\eta_p^2 = .45$, $p < .001$, but most important, the data show that this lure interference effect depended on whether or not interference occurred on the previous trial, $F(1, 27) = 28.97$, $\eta_p^2 = .52$, $p < .001$. Planned comparisons further indicate that although there was a reliable lure interference effect after a mismatch trial, $F(1, 27) = 34.93$, $\eta_p^2 = .56$, $p < .001$, this effect completely disappeared when the preceding trial was also a lure trial, $F < 1$. The disappearance of the lure interference effect when the previous trial was also a lure trial is explained by an increase in accuracy on lure interference trials, $F(1, 27) = 8.23$, $\eta_p^2 = .23$, $p < .01$, and a drop in accuracy on mismatch trials, $F(1, 27) = 37.37$, $\eta_p^2 = .58$, $p < .001$.

Discussion

The results of Experiment 4 show a clear reduction in the lure interference effects when the previous trial was also a lure trial compared with when the previous trial was a mismatch trial. In fact, the lure interference effect disappeared completely. As can be seen in Figure 6, this is the result of both a postinterference acceleration on lure trials, as well as a postinterference slowing on mismatch trials. This finding replicates previous findings in the conflict adaptation literature (see, e.g., Gratton et al., 1992) and fits well with the view that after a lure trial, the controlled recollection process is prioritized over the automatic familiarity-matching process. This will slow performance on the subsequent mismatch trials, on which the fast familiarity-matching process normally activates the correct response, but will speed performance on the subsequent lure trials, on which the fast familiarity-matching process normally activates the incorrect response.

The accuracy data also support the conflict adaptation hypothesis. With respect to the lure trials, we observed that *n*-back lures were performed more accurately after another lure trial than after a mismatch trial. The data for the mismatch trials show that, whereas accuracy is high for mismatch trials that are preceded by another mismatch trial, it is lower for mismatch trials that are preceded by a lure trial. Although this finding might appear to be somewhat counterintuitive (a decrease in accuracy after conflict adaptation), it has been found before in studies examining the Gratton effect (e.g., Notebaert & Verguts, 2008). This reduction in accuracy essentially shows that recollection is not necessarily more effective and less error-prone than familiarity matching; it is only more effective when conflict actually occurs (i.e., accuracy was higher for LL trials than for ML trials). In the absence of conflict (i.e., mismatch trials), familiarity-based matching is the most efficient way to access memory (i.e., accuracy was higher on MM trials than on LM trials). Thus, even if explicit recollection from memory allows resistance to lure interference, it is not as
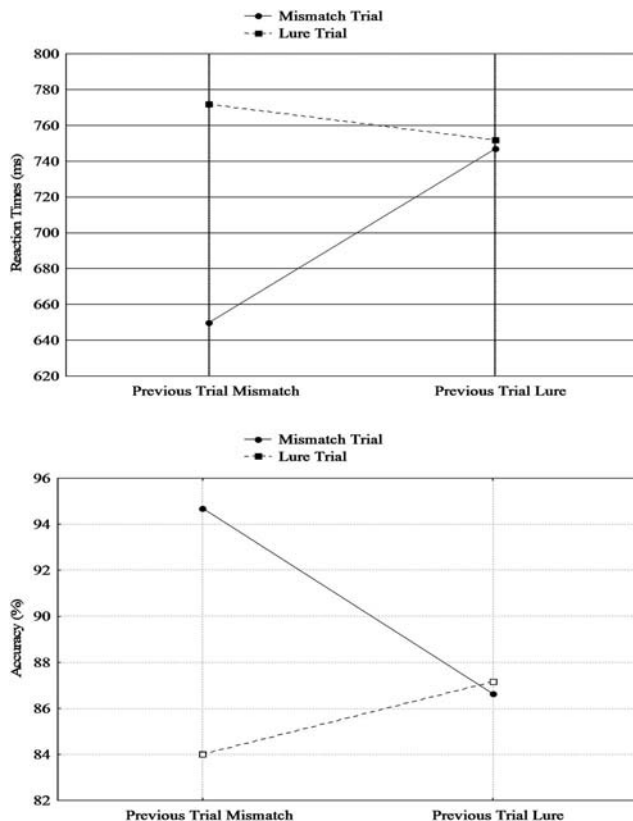

Figure 6.   Mean reaction times (upper panel) and accuracies (lower panel) for mismatch and lure trials as a function of lure interference on the previous trial (Experiment 4).

efficient as familiarity-based matching in the majority of trials that do not involve interference.

General Discussion

The aim of the present study was to investigate the nature of the cognitive processes underlying WM updating. In a series of experiments using the *n*-back task, we demonstrated that continuously updating the contents of WM causes interference between relevant and irrelevant items in WM, as evidenced by impaired performance on *n*+ and *n*– lure trials. We clarified the nature of this interference by arguing that WM updating prevents strong binding of items to WM (Oberauer, 2009) and therefore augments the sensitivity to proactive interference.

In the *n*-back task, subjects receive input from a familiarity signal, which reveals whether or not a newly presented item is familiar or not, and from the explicit recollection of an item in the DA region of WM, which tells whether the newly presented item is in target position *n*. Performance is challenged when a conflict arises between familiarity and recollection, that is, when the familiarity signal recognizes the stimulus as being familiar, but when this stimulus is in fact *not* in target position (i.e., called *n*-back lure trials). In order not to be misled on these occasions, a strong recollection process is required to overrule the conflicting information from the familiarity signal. This recollection process relies on strong bindings of items to their positional context in the DA region of WM. However, items cannot be strongly bound in WM when they are continuously updated. Therefore, updating demands impair the effectiveness of the binding process by which updating tasks such as the *n*-back task become particularly sensitive to interference. This was demonstrated in the patterns of lure interference that were elaborated in the current series of experiments, and most strikingly in Experiment 3. In this experiment, we observed interference due to items that were never presented before but that were semantically related to previously relevant items. Finally, we also showed that *n*-back lure interference is

controlled by prioritizing recollection over familiarity in a way that is reminiscent of the conflict adaptation mechanism described by the CMH (Botvinick et al., 2001). Combined, our findings fit well with the view that protecting the contents of WM against interference is a fundamental responsibility of WM updating (Bunting et al., 2006; Kessler & Meiran, 2008; Szmalec et al., 2009) and even of WM capacity more generally (Engle, 2005; Jonides & Nee, 2006; Oberauer, 2005).

Control of Interference During Updating

The present study describes how competition between familiarity and recollection causes interference in the *n*-back updating task and how this interference is dealt with by the cognitive control system. Although previous studies have argued that WM updating should involve a kind of interference control mechanism (e.g., Szmalec et al., 2009) or a global process responsible for protecting the materials held in WM against interference (Kessler & Meiran, 2008), the nature and the functional characteristics of this interference control mechanism were not further specified. The present study shows that interference during WM updating in fact seems to be counteracted by a functionally similar mechanism as the one responsible for the control of cognitive conflict in a variety of response-congruency tasks (cf. Botvinick et al., 2001). Note, however, that this functional similarity does not necessarily imply that the same neural mechanism is responsible for control of interference in different domains (e.g., Nelson, Reuter-Lorenz, Sylvester, Jonides, & Smith, 2003).

A schematic summary of the interference control mechanism involved in the *n*-back task is represented in Figure 7. In essence, this schema is the *n*-back version of a Stroop-like conflict monitoring model according to the CMH. The left half of the figure illustrates a 2-back mismatch. In this trial, both the familiarity and recollection process activate the mismatch response unit, so no interference occurs. The right half of the figure shows a lure trial. On lure trials, the familiarity signal activates the match response
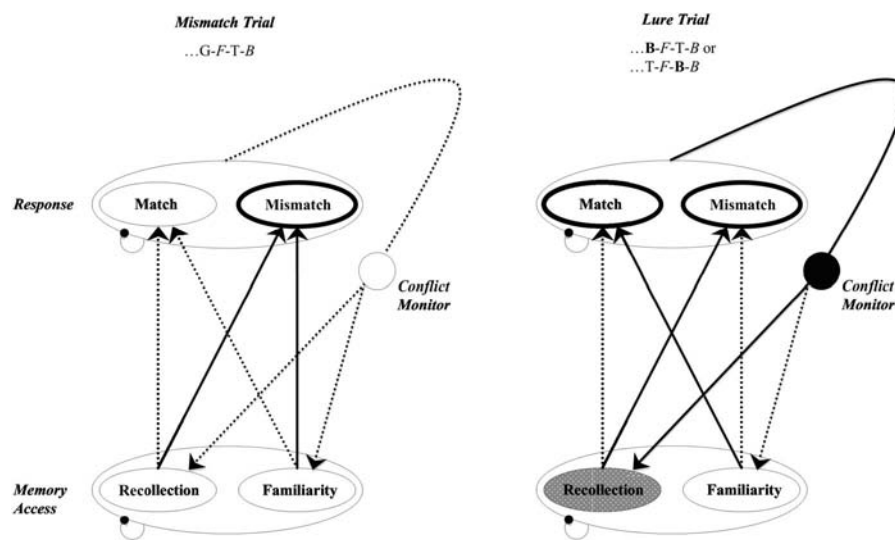

Figure 7. Schematic representation of the cognitive control mechanism responsible for top-down control of lure interference in the *n*-back updating task (example: 2-back variant).

unit, whereas the recollection process activates the mismatch response unit, causing a conflict in the response layer. This conflict is detected by a conflict monitoring unit (Botvinick et al., 2001), which subsequently biases the recollection process in a top-down fashion so that the appropriate mismatch response can be emitted. Empirical evidence for this top-down biasing of recollection was provided in Experiment 4 by means of a Gratton-like behavioral adaptation effect.

But how is this biasing of recollection achieved at the level of WM? One possible solution would be to try to establish stronger bindings and to adjust the threshold for retrieval of LTM representations in the DA region of WM (cf. Oberauer, 2009). With a low threshold, there is a high infiltration of LTM information in the DA region, and therefore, a strong influence of familiarity signals. If this information turns out to be irrelevant or misleading, proactive interference will arise. By raising the threshold, the LTM representations must attain a much higher level of activation in order to be retrieved in the DA region. This way, WM can shield the contents of the DA region against intrusions from LTM.

To the best of our knowledge, this is the first study to find evidence for Gratton-like conflict or interference adaptation effects in the context of memory recognition processes. This implies that the demonstration of the Gratton effect is not restricted to congruency tasks, where controlled processing is required to make sure that subjects can respond on the basis of relevant stimulus features while ignoring irrelevant stimulus features. The control mechanism behind the Gratton effect thus seems to be functionally analogous to the ones responsible for biasing different ways to access the contents of WM (i.e., recollection and familiarity). This opens new theoretical and empirical perspectives on the study of the relation between two major domains of cognitive science, namely cognitive control and memory.

The *n*-Back Task

The current results also have implications for the use of the *n*-back task as a measure of WM performance. The *n*-back task is currently one of the most popular measures of WM in cognitive neuroscience, particularly in studies investigating the neural substrates of WM (Owen, McMillan, Laird, & Bullmore, 2005; Smith & Jonides, 1997). This popularity mainly originates from the *n*-back task's high face validity as a WM task, although the studies that have directly investigated the relation between the *n*-back task and other WM measures always revealed weak correlations (Kane et al., 2007; Oberauer, 2005). The detailed *n*-back task analyses conducted in the current study show that the kind of processes that are involved in the *n*-back task, or in other words what the *n*-back task measures, depends heavily on whether or not the procedure contains lure trials. If lure trials are not involved, the task can be performed mainly by relying on familiarity matching without measurable requirements for explicit recall or cognitive control (see also Bunting et al., 2006). When the proportion of lure trials is higher, consciously controlled recollection of the items is privileged and cognitive control resources are heavily taxed to protect the contents of memory against interference. It is thus important to realize that the *n*-back task may be a measure of cognitive control in addition to WM performance, as also supported by its modest correlation with pure WM measures (see Kane et al., 2007, for a more detailed discussion) and rather strong correlation with Stroop

interference (Kwong See & Ryan, 1995). Future studies should be aware that the presence or absence of lure trials in the *n*-back procedure has important implications for the cognitive processes of interest: An *n*-back procedure with and one without lure trials are almost two different tasks in terms of what they measure.

Working Memory Updating: Executive Function or Task?

Finally, what do the present findings regarding interference control during updating tell us about the concept of executive control in WM? Executive control was initially conceptualized as a single control entity, such as the central executive (Baddeley, 1986), responsible for producing goal-directed behavior. This concept, however, was so poorly specified that it seemed as if a hidden "homunculus" performed the critical interventions. An important theoretical step forward was made by Miyake et al. (2000), who questioned the unity/diversity within executive control and proposed to fractionate executive control in a taxonomy of separable and relatively autonomous executive control functions, such as inhibition, task switching, or updating. Although the notion of autonomous executive control functions has gained much popularity over the years, the empirical evidence for this theoretical position is not that strong in our view. On the one hand, correlations between executive function tasks are generally moderate (e.g., Miyake et al., 2000), which can actually be interpreted as evidence for both sides of the unity/diversity debate. On the other hand, the fractioning program of a unitary control system does not seem to end with functions like updating or inhibition of shifting because it seems that even more lower level control processes, such as response selection, might be the primitives of executive control (e.g., Szmalec, Vandierendonck, & Kemps, 2005; Szmalec et al., 2008; Vandierendonck, Szmalec, Deschuyteneer, & Depoorter, 2007).

The present data clearly suggest that the executive demands of the *n*-back updating task are strongly related to the requirement of protecting WM against interference during updating. This interference seems to be dealt with by a conflict monitoring loop, which is functionally similar to the one put forward by the CMH (Botvinick et al., 2001). During updating, the control mechanism is responsible for biasing recollection to protect WM against proactive interference. As we argued before, this biasing can be achieved by adjusting retrieval thresholds in LTM. Therefore, we believe that the empirical evidence so far supports the view that WM updating is in the first place a task demand (e.g., the demand to update a representation in WM, as in the *n*-back task) that involves several cognitive processes (e.g., binding in WM, threshold adaptation in LTM) and that is controlled by a conflict monitoring mechanism, operationalized by the Gratton effect, which monitors for suboptimal performance (or conflict, interference) in the cognitive system. Our view thus promotes the redefinition of a central executive control entity into a clearly described conflict monitoring and control mechanism that may be operating in a variety of task contexts, rather than by a taxonomy of potentially separable and autonomous executive functions (see also Braver, Gray, & Burgess, 2007; Szmalec et al., 2009).

References

Baddeley, A. (1986). *Working memory.* Oxford, England: Oxford University Press, Clarendon Press.

Bartha, M. C., Martin, R. C., & Jensen, C. R. (1998). Multiple interference effects in short-term recognition memory. *American Journal of Psychology, 111,* 89–118.

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 317–333.

Botvinick, M., Braver, T., Barch, D., Carter, C., & Cohen, J. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108,* 625–652.

Braver, T., Gray, J., & Burgess, G. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 76–106). Oxford, England: Oxford University Press.

Bunting, M., Cowan, N., & Saults, J. S. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology, 59,* 1691–1700.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Collette, F., & Van Der Linden, M. (2002). Brain imaging of the central executive component of working memory. *Neuroscience & Biobehavioral Reviews, 26,* 105–125.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin, 104,* 163–191.

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

Cowan, N. (2005). *Working memory capacity.* New York: Psychology Press.

De Deyne, S., & Storms, G. (2008). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods, 40,* 198–205.

Engle, R. (2005, March). *Working memory capacity and inhibition.* Paper presented at the 2005 meeting for the Place of Inhibitory Process in Cognition, Arlington, TX.

Gardiner, J., Gawlik, B., & Richardson-Klavehn, A. (1994). Maintenance rehearsal affects knowing, not remembering; elaborative rehearsal affects remembering, not knowing. *Psychonomic Bulletin & Review, 1,* 107–110.

Glanzer, M., & Cunitz, A. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 351–360.

Goethe, K., & Oberauer, K. (2008). The integration of familiarity and recollection information in short-term recognition: Modeling speed–accuracy trade-off functions. *Psychological Research, 72,* 289–303.

Gratton, G., Coles, M., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General, 121,* 480–506.

Gray, J., Chabris, C., & Braver, T. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6,* 316–322.

Hamilton, A., & Martin, R. C. (2007). Proactive interference in a semantic short-term memory deficit: Role of semantic and phonological relatedness. *Cortex, 43,* 112–123.

Henson, R. N. A. (1999). Coding position in short-term memory. *International Journal of Psychology, 34,* 403–409.

Hockey, R. (1973). Rate of presentation in running memory and direct manipulation of input-processing strategies. *Quarterly Journal of Experimental Psychology, 25,* 104–111.

Jonides, J., & Nee, E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience, 139,* 181–193.

Jonides, J., Schumacher, E., Smith, E., Lauber, E., Awh, E., Minoshima, S., & Koeppe, R. (1997). The task-load of verbal working memory affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience, 9,* 462–475.

Kane, M., Conway, A., Miura, T., & Colflesh, G. (2007). Working memory, attention control and the *n*-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 615–622.

Kerns, J., Cohen, J., MacDonald, A., Cho, R., Stenger, V., & Carter, C. (2004, February 13). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303,* 1023–1026.

Kessler, Y., & Meiran, N. (2006). All updateable objects in working memory are updated whenever any of them is modified: Evidence from the memory updating paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 570–585.

Kessler, Y., & Meiran, N. (2008). Two dissociable updating processes in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1339–1348.

Kwong See, S., & Ryan, E. (1995). Cognitive mediation of adult age differences in language performance. *Psychology and Aging, 10,* 458–468.

Lewandowsky, S., Brown, G. D. A., Wright, T., & Nimmo, L. M. (2006). Timeless memory: Evidence against temporal distinctiveness models of short-term memory for serial order. *Journal of Memory and Language, 54,* 20–38.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 817–835.

Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Nelson, J. K., Reuter-Lorenz, P. A., Sylvester, C. Y. C., Jonides, J., & Smith, A. D. (2003). Dissociable neural mechanisms underlying response-based and familiarity-based conflict in working memory. *Proceedings of the National Academy of Science in the USA, 100,* 11171–11175.

Ng, H. L. H., & Maybery, M. T. (2002). Grouping in short-term verbal memory: Is position coded temporally? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55*(A), 391–424.

Notebaert, W., Gevers, W., Verbruggen, F., & Liefooghe, B. (2006). Top-down and bottom-up sequential modulations of congruency effects. *Psychonomic Bulletin & Review, 13,* 112–117.

Notebaert, W., & Verguts, T. (2008). Cognitive control acts locally. *Cognition, 106,* 1071–1080.

Oberauer, K. (2005). Binding and inhibition in working memory—Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General, 134,* 368–387.

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation: Advances in Research and Theory, 51,* 45–100.

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review, 115,* 544–576.

Owen, A., McMillan, K., Laird, A., & Bullmore, E. (2005). The *n*-back working memory paradigm: A meta-analysis of normative studies. *Human Brain Mapping, 25,* 46–59.

Oztekin, I., & McElree, B. (2007). Retrieval dynamics of proactive interference: PI slows retrieval by eliminating fast assessment of familiarity. *Journal of Memory and Language, 57,* 126–149.

Page, M., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761–781.

Smith, E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33,* 5–42.

Stürmer, B., Leuthold, H., Soetens, E., Schröter, H., & Sommer, W. (2002).

Control over location-based response activation in the Simon task: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1345–1363.

Szmalec, A., Demanet, J., Vandierendonck, A., & Verbruggen, F. (2009). Investigating the role of conflict resolution in memory updating by means of the one-back choice RT task. *Psychological Research, 73,* 390–406.

Szmalec, A., & Vandierendonck, A. (2007). Estimating the executive demands of a one-back choice RT task by means of the selective interference paradigm. *Quarterly Journal of Experimental Psychology, 60,* 1116–1139.

Szmalec, A., Vandierendonck, A., & Kemps, E. (2005). Response selection involves executive control: Evidence from the selective interference paradigm. *Memory & Cognition, 33,* 531–541.

Szmalec, A., Verbruggen, F., Vandierendonck, A., De Baene, W., Verguts, T., & Notebaert, W. (2008). Stimulus ambiguity elicits response conflict. *Neuroscience Letters, 435,* 158–162.

Vandierendonck, A., Szmalec, A., Deschuyteneer, M., & Depoorter, A.

(2007). Towards a multicomponent view of executive control. The case of response selection. In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 247–259). Oxford, England: Oxford University Press.

Verbruggen, F., & Logan, G. (2009). Automaticity of cognitive control: Goal priming in response-inhibition paradigms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1381–1388.

Verbruggen, F., Notebaert, W., Liefooghe, B., & Vandierendonck, A. (2006). Stimulus and response conflict-induced cognitive control in the flanker task. *Psychonomic Bulletin & Review, 13,* 328–333.

Verguts, T., & Notebaert, W. (2008). Hebbian learning of cognitive control: Dealing with specific and nonspecific adaptation. *Psychological Review, 115,* 518–525.

Verguts, T., & Notebaert, W. (2009). Adaptation by binding: A learning account of cognitive control. *Trends in Cognitive Sciences, 13,* 252–257.

Yonelinas, A. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517.

Appendix A

Analyses for Experiment 1

For each value of set size *n*, the data were analyzed by means of a 2 (modality: verbal and visuospatial) × 4 (trial type) ANOVA with repeated measures on the last factor. Power calculations, based on Cohen's (1988) power tables, indicate that the tests of difference between both stimulus modalities were sufficiently powerful to detect a moderate sized effect (all values = .99)

Variable	Reaction time			Accuracy		
	df	F	η_p^2	df	F	η_p^2
n = 1 (1-back)						
Trial type	3, 57	18.72***	.50	3, 57	11.69***	.38
Modality	1, 19	0.00	.00	1, 19	2.54	.12
Trial Type × Modality	3, 57	0.42	.02	3, 57	1.30	.06
n = 2 (2-back)						
Trial type	3, 57	46.88***	.71	3, 57	35.57***	.65
Modality	1, 19	1.66	.08	1, 19	1.25	.06
Trial Type × Modality	3, 57	0.71	.04	3, 57	2.58	.12
n = 3 (3-back)						
Trial type	3, 57	15.63***	.45	3, 57	16.17***	.46
Modality	1, 19	1.66	.08	1, 19	9.96**	.34
Trial Type × Modality	3, 57	0.02	.00	3, 57	3.84*	.17

Note. Trial type conditions differed as a function of *n*: match, mismatch, *n* + 1 lure, and *n* + 2 lure for 1-back; match, mismatch, *n* + 1 lure, and *n* − 1 lure for 2-back; match, mismatch, *n* − 1 lure, and *n* − 2 lure for 3-back.
* *p* < .05. ** *p* < .01. *** *p* < .001.

(*Appendices continue*)

Appendix B

Analyses for Experiment 2

The 2-back data were analyzed by means of a 2 (ISI: short and long) × 4 (trial type: match, mismatch, $n + 1$ lure, $n - 1$ lure) repeated measures ANOVA

Variable	Reaction time			Accuracy		
	df	F	η_p^2	df	F	η_p^2
Trial type	3, 75	56.70***	.69	3, 75	9.41***	.27
ISI	1, 25	4.09	.14	1, 25	0.04	.00
Trial Type × ISI	3, 75	14.38***	.36	3, 75	1.91	.07

Note. ISI = interstimulus interval.
* $p < .05$. ** $p < .01$. *** $p < .001$.