Individual Differences in Working Memory Capacity Predict Retrieval-Induced Forgetting

Alp Aslan and Karl-Heinz T. Bäuml
Regensburg University

Selectively retrieving a subset of previously studied information enhances memory for the retrieved information but causes forgetting of related, nonretrieved information. Corresponding evidence comes from numerous studies using the so-called retrieval-practice task (Anderson, Bjork, & Bjork, 1994). In the standard variant of this task, participants study items from different semantic categories (e.g., Fruit–Orange, Fruit–Banana, Animal–Tiger, Animal–Lion) and subsequently perform selective retrieval practice on half of the items from half of the categories (e.g., Fruit–Orange). After a retention interval, a category-cued recall test for all previously studied items is administered. The typical finding in this test is that recall of the practiced items (Orange) is improved and recall of the unpracticed items (Banana) is impaired, relative to control items from unpracticed categories (Tiger, Lion). The latter finding is called retrieval-induced forgetting (RIF) and has proven to be a very robust and general effect. Indeed, RIF is not restricted to simple word lists but has also been found using more complex stimuli and applied settings, including propositional (Anderson & Bell, 2001) and visuospatial materials (Ciranni & Shimamura, 1999), eyewitness (Shaw, Bjork, & Handal, 1995) and autobiographical memory (Barnier, Hung, & Conway, 2004), and even foreign-language-acquisition scenarios (Levy, McVeigh, Marful, & Anderson, 2007).

A prominent account of RIF assumes that RIF arises from inhibitory executive-control processes during retrieval practice. The proposal is that during retrieval attempts, the not-to-be-retrieved (unpracticed) items interfere and, to overcome the interference, are inhibited (for reviews, see Anderson, 2003; Bäuml, 2008). Several lines of evidence support the inhibitory executive-control account of RIF. One key feature of the account is that it attributes RIF to active suppression of an item’s representation itself rather than to changes in the item’s associative structure. Access to an inhibited item, therefore, should be impaired regardless of which memory test or retrieval cue is used to probe the item. Consistently, RIF has been found in a wide range of memory tests, including word-stem completion (e.g., Anderson et al., 1994; Bäuml & Aslan, 2004), item recognition (e.g., Hicks & Starns, 2004; Spitzer & Bäuml, 2007), and tests using so-called independent probes, that is, novel retrieval cues not used until the test phase of the experiment (e.g., Anderson & Spellman, 1995; Saunders & MacLeod, 2006). Moreover, a recent fMRI study reported increased brain activity during recall of the unpracticed items in the test phase that was predictive of RIF (Wimber et al., 2008). Because the activity was found in brain areas typically associated with the controlled retrieval of particularly weak memories, the finding agrees with the inhibitory view that retrieval practice reduces unpracticed items’ memory representation.

Several other findings also support the view that RIF is the result of (inhibitory) executive-control processes. In particular, inhibitory executive control is generally considered a frontally mediated, resource-demanding process (A. R. A. Conway & Engle, 1994; Kane & Engle, 2002). In line with this view, recent neurocognitive work reported prefrontal activations during the retrieval-practice phase that predicted later forgetting of the unpracticed information (Johansson, Aslan, Bäuml, Gabel, & Mecklinger, 2007; Kuhl, Dudukovic, Kahn, & Wagner, 2007; Wimber, Rutschmann, Greenlee, & Bäuml, 2009). Other behavioral work showed that RIF is diminished when a secondary task is performed during retrieval practice (Román, Soriano, Gómez-Ariza, & Bajo, 2009).
Although the inhibitory executive-control view of RIF is consistent with many findings in the literature, the view has not been without criticism. In fact, some studies failed to find RIF with independent probes, leading some researchers to either doubt the finding that RIF is cue independent (Camp, Pecher, & Schmidt, 2007; Perfect et al., 2004; Williams & Zacks, 2001) or to express more basic objections against the independent-probe procedure itself (Camp, Pecher, Schmidt, & Zeelenberg, 2009). Jakab and Raaijmakers (2009) failed to find a role of item strength in RIF, which is in conflict with the inhibitory view that more RIF should arise for strong than for weak unpracticed items (but see Anderson et al., 1994; Bäuml, 1998). Finally, several studies reported intact RIF in (groups of) individuals who were supposed to have deficits in inhibitory executive control, including various clinical populations (M. A. Conway & Fthenaki, 2003; Moulin et al., 2002; Racsmány et al., 2008) and (healthy) young children (FORD, Keating, & Patel, 2004; Zellner & Bäuml, 2005); if RIF reflected the operation of controlled inhibitory processes, RIF should be reduced or even be absent in these individuals.

In light of such findings, some authors questioned the inhibitory executive-control view of RIF and instead proposed an interference-based explanation of the effect (Camp et al., 2007; Jakab & Raaijmakers, 2009). According to this proposal, retrieval practice strengthens the association between an item and its retrieval cue, so that at test, the (stronger) practiced items come to mind persistently and block access to the (relatively weaker) unpracticed items, thus causing RIF without inhibition. In contrast to the inhibitory executive-control account, the interference account thus assumes that RIF reflects a more passive and automatic side effect of increased interference at test owing to the prior strengthening of the practiced items. However, the interference account cannot easily explain the finding of RIF in interference-free memory tests, like recognition or independent-probe tests, and it has difficulties accommodating the above-mentioned findings from the neurocognitive work.

Our aim in the present study was to further evaluate the inhibitory executive-control account of RIF by approaching RIF from an individual-differences perspective (see also Levy & Anderson, 2008). Previous work indicates that individuals differ largely in their capability for inhibitory executive control. Specifically, it has been argued that the efficiency of inhibitory control processes is related to individuals’ working memory capacity (WMC), so that individuals with higher WMC are better able than individuals with lower WMC to deal with interference and inhibit task-irrelevant information (Redick, Heitz, & Engle, 2007). Consistently, measures of WMC have been found to predict performance in a number of cognitive tasks that are supposed to require controlled inhibition, including the Stroop task (Kane & Engle, 2003), the antisaccade task (Kane, Bleckley, Conway, & Engle, 2001), and the directed-forgetting task (Aslan, Zellner, & Bäuml, 2010). Following the view that WMC is related to the efficiency of inhibitory control processes, the inhibitory executive-control account of RIF predicts that individuals with higher WMC should show more RIF than individuals with lower WMC.

In the present study, we tested this prediction by examining young adults’ performance in the retrieval-practice task and relating it to their performance in the operation span (OSPAN) task (Turner & Engle, 1989). The OSPAN task is a widely used tool in individual-differences research; it requires participants to simultaneously store and process information and provides reliable and valid measures of individuals’ WMC (see A. R. A. Conway et al., 2005). If RIF reflected a controlled and resource-demanding inhibitory process, a positive relationship between RIF and WMC should arise, and individuals with higher OSPAN scores should show more RIF than individuals with lower OSPAN scores. It is important to note that because the supposed relationship between RIF and WMC depends on the purity of the inhibition measure, we followed previous work (e.g., Spitzer & Bäuml, 2007) and assessed RIF using an interference-free recognition test.

Method

Participants

One hundred sixty-eight adults ($M = 24.1$ years, $SD = 4.4$ years) took part in the study. They were tested individually.

Retrieval-Practice Task

Materials. Twelve exemplars from each of 12 semantic categories were drawn from published word norms (Mannhaupt, 1983; Scheithe & Bäuml, 1995). Within a category, each item had a unique word stem. The three exemplars with the highest and the three exemplars with the lowest word frequency within each selected category were never studied but were used as lures in the recognition test. From the remaining six items of each category, the three exemplars with the higher word frequency were used as target items, and the three exemplars with the lower word frequency were used as nontarget items. The separation of materials into target and non-target items followed previous work (Aslan & Bäuml, 2010; Spitzer & Bäuml, 2007) and was intended to boost the RIF effect for target items; in fact, previous work provided evidence that, compared with high-frequency exemplars, low-frequency exemplars are less susceptible to RIF or not susceptible at all (Anderson et al., 1994; Bäuml, 1998).

Design. The experiment consisted of three main phases: a study phase, an intermediate retrieval-practice phase in which the non-targets from half of the categories were practiced, and a final test phase in which a recognition test was administered. Retrieval practice created four types of studied items: Practiced categories contained practiced non-target items ($RP^+$) and unpracticed target items ($RP^-$); unpracticed categories contained (unpracticed) non-target items ($NP^+$) serving as controls for $RP^+$ items, and (unpracticed) target items ($NP^-$) serving as controls for $RP^-$ items. Retrieval practice further created two types of (nonstudied) lure items: lures from practiced categories ($RP$ lures) and lures from unpracticed categories ($NP$ lures). Across participants, all non-targets served equally often as $RP^+$ and $NP^+$ items, and all targets served equally often as $RP^-$ and $NP^-$ items. Likewise, all lure items served equally often as $RP$ and $NP$ lures. Thus, the relevant comparisons regarding retrieval-induced enhancement ($RP^+ \text{ vs. } NP^+$) and RIF ($RP^- \text{ vs. } NP^-$) were not affected by item characteristics.

Procedure

Study phase. The 72 (12 × 6) items of the study list were presented successively at a 3-s rate on a computer screen, each
item together with its category name (e.g., Fruit–Orange). Items were presented in a fixed random order to one half of the participants and in the reverse order to the other half. Following the last item of the study list, participants engaged in a 30-s backward-counting distractor task.

Retrieval-practice phase. Following the distractor task, a written category-cued word-stem completion test was administered. Participants were given a sheet of paper containing the word stems of 18 (6 × 3) nontarget items, each stem together with its category name (e.g., Fruit–Orange), and were asked to complete the stems with the corresponding items from the study list. Immediately after this first retrieval-practice cycle, a second, identical practice cycle was conducted on a second sheet of paper. The order of the word stems on both sheets was random with the restriction that no two items from the same category were presented successively.

Test phase. Following another 1-min backward-counting distractor task, participants engaged in a written old–new recognition test. The test sheet contained the 72 previously studied items and 72 nonstudied lure items, intermixed in random order. Participants were asked to go through the items at their own pace and to judge whether each item was old (i.e., had been previously studied) or new (i.e., had not been studied). The order of the items was counterbalanced such that the mean position of each item type was equal across participants.

Working Memory Task

The working memory task was administered after the RIF experiment. Participants’ WMC was assessed with a German version of the OSPAN task (Turner & Engle, 1989; see also Aslan et al., 2010). The OSPAN task required participants to solve arithmetic equations while trying to remember unrelated words. Each trial consisted of a certain number (varying between 2 and 6) of successively presented equation–word pairs: for example, \(8 \div 4 + 3 = 5\) moon. Participants had to read each equation aloud, verify whether it was correct by saying “yes” or “no,” and read the to-be-remembered word (moon) aloud. Participants were urged to respond quickly and, immediately after the response, the next equation–word pair was presented. Following the last equation–word pair, participants were asked to recall the to-be-remembered words in correct order. There were three repetitions of each set size (2–6), leading to a maximum OSPAN score of 60. The span score was defined as the number of recalled words on correct sets. A set was counted as correct if all of the presented words from that set were recalled in correct order (Turner & Engle, 1989; see A. R. A. Conway et al., 2005, for a review of scoring methods).

Results

Retrieval-Practice Effects

In the retrieval-practice task, participants, on average, successfully completed 91.6% of the nontarget items’ word stems. The results of the final recognition test are shown in Table 1. Preliminary analyses revealed that false-alarm rates were comparable for lures from practiced categories (Rp lures) and lures from unpracticed categories (Nrp lures: 13.6% vs. 14.0%), \(p > .50\). To examine the memorial effects of retrieval practice, we assessed individuals’ recognition accuracy for the various item types using \(d' = \frac{z_{\text{hit rate}} - z_{\text{false-alarm rate}}}{\sqrt{2}}\) as the dependent measure. Overall, retrieval practice had the expected effects and improved memory for the practiced (nontarget) items, \(d'(Rp+) = 2.73\) versus \(d'(Nrp+) = 1.79, \tau(167) = 16.1, SE = 0.06, p < .001,\) and caused RIF for the unpracticed (target) items, \(d'(Rp-) = 1.73\) versus \(d'(Nrp-) = 1.92, \tau(167) = 3.8, SE = 0.05, p < .001.\) As expected, retrieval-induced enhancement, defined as \(d'(Rp+) - d'(Nrp+)\), was correlated to the success rate during retrieval practice, \(r = .21, p < .1;\) however, it was not correlated to RIF, defined as \(d'(Nrp-) - d'(Rp-)\), \(r = -.08, p > .20,\) which is consistent with prior work (e.g., Hanslmayr, Staudigl, Aslan, & Bäuml, 2010; Staudigl, Hanslmayr, & Bäuml, 2010).

Table 1

<table>
<thead>
<tr>
<th>Item type</th>
<th>Target</th>
<th>Nontarget</th>
<th>Lure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieval practice status</td>
<td>%</td>
<td>SE</td>
<td>%</td>
</tr>
<tr>
<td>Practiced</td>
<td>64.3</td>
<td>1.6</td>
<td>89.7</td>
</tr>
<tr>
<td>Unpracticed</td>
<td>71.5</td>
<td>1.3</td>
<td>66.9</td>
</tr>
</tbody>
</table>

Note. “Old” responses to target and nontarget items reflect hits; “old” responses to lure items reflect false alarms. Retrieval practice created four types of studied items: Practiced categories contained practiced nontarget items (Rp+) and unpracticed target items (Rp−); unpracticed categories contained (unpracticed) nontarget items (Nrp+) serving as controls for Rp+ items and (unpracticed) target items (Nrp−) serving as controls for Rp− items. Retrieval practice further created two types of (nonstudied) lure items: lures from practiced categories (Rp lures) and lures from unpracticed categories (Nrp lures).

Relationship Between WMC and the Retrieval-Practise Effects

Regarding WMC, participants’ mean OSPAN score was 27.2 (SD = 12.5, range 6–60), a value well comparable to prior work (e.g., Aslan et al., 2010). Preliminary analyses revealed that WMC was related to neither individuals’ success rates (\(p > .50\)) nor individuals’ false-alarm rates for Rp lures and Nrp lures (both ps > .15).

To examine the relationship between WMC and the two retrieval-practice effects, we regressed individual retrieval-induced enhancement scores and individual RIF scores separately on the individual WMC score. The resulting scatterplots together with the best-fitting linear regression lines are shown in Figure 1. Regarding retrieval-induced enhancement (see Figure 1A), the regression line was essentially horizontal. In fact, the slope of the regression line did not differ from zero, \(b = -.002, SE = .005, p > .70,\) indicating that the amount of retrieval-induced enhancement did not vary with individuals’ WMC. In contrast, regarding RIF (see Figure 1B), the slope of the regression line was positive and

1 This procedure created primacy and recency items on the one hand and middle items on the other. Fortunately, retrieval practice effects do not vary with items’ serial list position (Jakab & Raatjmakers, 2009).
differed significantly from zero, $b = .018$, $SE = .018$, $p < .001$, reflecting the fact that the amount of RIF increased reliably with increasing WMC. The correlation between RIF and WMC ($r = .35$) was significantly larger than the correlation between retrieval-induced enhancement and WMC ($r = -.03$), $p < .001$.2

Discussion

Using a relatively large sample of participants ($N = 168$), in this study, we examined the relationship between WMC and the two memorial effects of selective retrieval practice. We found no relationship between WMC and the beneficial effect of retrieval practice, with high-WMC and low-WMC individuals showing equivalent retrieval-induced enhancement of practiced items. In contrast, we found a positive relationship between WMC and the detrimental effect of retrieval practice, with high-WMC individuals, as compared with low-WMC individuals, showing more RIF of unpracticed items.

The results of the present study are consistent with previous individual-differences research and with the inhibitory account of RIF. The inhibitory account of RIF assumes that RIF arises as an aftereffect of executive-control processes during retrieval practice, suppressing the not-to-be-retrieved (unpracticed) items’ memory representation (Anderson, 2003; Bäuml, 2008). Previous individual-differences research revealed that individuals with higher WMC are generally better able than individuals with lower WMC to exert inhibitory control on interfering material (for a review, see Redick et al., 2007). The inhibitory account of RIF, therefore, predicts that individuals with higher WMC should show more RIF than individuals with lower WMC, who may show no RIF at all. Indeed, this is exactly what we found in the present study.

The finding that RIF is related to WMC also suggests that RIF should be reduced if individuals’ attention was divided during retrieval practice and, on a neurocognitive level, that RIF should be accompanied by prefrontal brain activations during retrieval practice (Curtis & D’Esposito, 2003; Kane & Engle, 2002). Corresponding results have been reported in recent work. Román et al. (2009) found RIF to be diminished when individuals engaged in a secondary task during retrieval practice. Using both fMRI and EEG measures of brain activity, several groups of researchers observed prefrontal brain activations during retrieval practice that were predictive of individual RIF performance (Johansson et al., 2007; Kuhl et al., 2007; Staudigl et al., 2010; Wimber et al., 2009). Together, these findings are in line with the present indication that WMC is related to RIF.

Although the present results agree with the inhibitory account of RIF, they are hard to reconcile with the (noninhibitory) interference account (e.g., Camp et al., 2007; Jakab & Raaijmakers, 2009). The interference account assumes that RIF arises at test as a result of increased interference from the (stronger) practiced items, blocking access to the (relatively weaker) unpracticed items. Following this account, RIF should be restricted to interference-rich memory tests, like free or category-cued recall, and should be absent in interference-free memory tests, like item recognition or independent-probe tests. The present finding of reliable overall RIF in item recognition, therefore, is inconsistent with the interference account (for previous demonstrations of RIF in item recognition, see Hicks & Starns, 2004; Spitzer & Bäuml, 2007).

Moreover, because the interference account assumes that RIF is a direct side effect of the strengthening of the practiced items, the account predicts that the two effects of retrieval practice should go hand in hand and behave concordantly. The finding that RIF and retrieval-induced enhancement were not correlated and WMC pre-

---

2 Because hit rates for practiced ($Rp^+$) items were relatively high in the present study, we checked whether ceiling effects may have masked possible correlations between retrieval-induced enhancement and RIF or between retrieval-induced enhancement and WMC. Doing so, we repeated the above-reported analyses, excluding all participants who were at ceiling with their $Rp^+$ hit rate. Examination of this restricted sample revealed a correlation between RIF and WMC, $r = .27$, $p < .01$; no correlation between retrieval-induced enhancement and WMC, $r = -.01$, $p > .90$; and no correlation between retrieval-induced enhancement and RIF, $r = -.00$, $p > .90$. In particular, the correlation between RIF and WMC was larger than the correlation between retrieval-induced enhancement and WMC, $p < .05$. These findings mimic those of the total sample, suggesting that the null correlations between retrieval-induced enhancement and the two other variables (i.e., RIF and WMC) were real and not the result of ceiling effects for the practiced ($Rp^+$) items.
dicted RIF but not retrieval-induced enhancement thus provides another challenge for the interference account. The finding, however, fits well with previous behavioral studies, showing that RIF and retrieval-induced enhancement are uncorrelated (Hanslmayr et al., 2010; Staudigl et al., 2010), RIF can occur without retrieval-induced enhancement (Gómez-Ariza, Lechuga, Pelegrina, & Bajo, 2005; Storm, Bjork, Bjork, & Nestojko, 2006; Veling & van Knippenberg, 2004), and retrieval-induced enhancement can occur without RIF (Anderson et al., 1994; Bäuml & Kuhbandner, 2007; Kössler, Engler, Riether, & Kissler, 2009). The finding is also in line with recent neurocognitive work reporting dissociable neural correlates for RIF and retrieval-induced enhancement (Kuhl, Kahn, Dudukovic, & Wagner, 2008; Spitzer, Hanslmayr, Opitz, Mecklinger, & Bäuml, 2009; Wimber et al., 2008).

The present finding that RIF is reduced in individuals with lower WMC contrasts with the results of previous developmental and clinical studies reporting efficient RIF in young children (Ford et al., 2004; Zellner & Bäuml, 2005) and certain clinical populations, like patients with frontal lesions (M. A. Conway & Fihenaki, 2003), schizophrenia (Racsmany et al., 2008), and Alzheimer’s disease (Moulin et al., 2002). Indeed, because WMC is generally reduced in these subject groups, as compared with healthy young adults (Baddeley, Bressi, Della Sala, Logie, & Spinell, 1991; Curtis & D’Esposito, 2003; Lee & Park, 2005; Siegel, 1994), at first glance, the finding of efficient RIF in these (low-WMC) individuals disagrees with the present results.

However, in contrast to the present study, in which we used recognition testing, most of the previous clinical and developmental studies assessed individuals’ RIF using interference-contaminated recall tests, in which the practiced items may easily interfere with the recall of the unpracticed material. Given that individuals with low WMC are particularly vulnerable to interference effects (e.g., Kane & Engle, 2000), the observed forgetting in these studies may not have been caused by efficient inhibition during retrieval practice but rather may have been due to (exaggerated) interference effects at test. Consistently, when using item-specific memory tests that arguably circumvent interference effects, like word-list completion and recognition, several more recent studies reported inefficient RIF in individuals who were supposed to have reduced WMC, including attention-deficit/hyperactivity disorder patients (Storm & White, 2010), schizophrenic patients (Soriano, Jiménez, Román, & Bajo, 2009), and young children (Aslan & Bäuml, 2010).

In sum, this study is the first one that examined the relationship between WMC and the two memorial effects of selective retrieval practice. Although we found no relationship between WMC and the beneficial effect of retrieval practice, we found a positive relationship between WMC and the detrimental effect of retrieval practice. The present results are in line with previous individual-differences work that suggests a close link between WMC and inhibitory capability. In particular, our findings support the inhibitory executive-control view of RIF.

References


Received June 11, 2010
Revision received August 6, 2010
Accepted August 6, 2010