Running head: COMPONENTS OF UPDATING

The Components of Working Memory Updating: An Experimental Decomposition and

Individual Differences

Ullrich K. H. Ecker and Stephan Lewandowsky

University of Western Australia

Klaus Oberauer

University of Bristol

Abby E. H. Chee

University of Western Australia

Word count: 13909 excluding references (approximate count due to use of LaTeX)

Ullrich Ecker

School of Psychology

University of Western Australia

Crawley, W.A. 6009, AUSTRALIA

Telephone: +618 6488 3266

Fax: +618 6488 1006

E-mail: ullrich.ecker@uwa.edu.au

Abstract

Working memory updating (WMU) has been identified as a cognitive function of prime importance for everyday tasks, and has also been found to be a significant predictor of higher mental abilities. Yet, little is known about the constituent processes of WMU. We suggest that operations required in a typical WMU task can be decomposed into three major component processes, namely retrieval, transformation, and substitution. We report a large-scale experiment that instantiated all possible combinations of those three component processes. Results show that the three components make independent contributions to updating performance. We additionally present structural equation models that link WMU task performance and working memory capacity (WMC) measures. These feature the methodological advancement of estimating interindividual covariation and experimental effects on mean updating measures simultaneously. The modeling results imply that WMC is a strong predictor of WMU skills in general, although some component processes—in particular substitution skills—were independent of WMC. Hence, the reported predictive power of WMU measures may largely rely on common WM functions also measured in typical WMC tasks, although substitution skills may make an independent contribution to predicting higher mental abilities.

The Components of Working Memory Updating: An
Experimental Decomposition and Individual Differences

Working memory (WM) can be characterized as a mechanism, or collection of mechanisms, dedicated to holding selected representations available for further cognitive processing. For example, mental arithmetic would be impossible without storage of intermediate sums in working memory (Fürst & Hitch, 2000). One implication of the involvement of WM in many cognitive tasks is that the content of WM must be continuously updated. For instance, a simple task like counting sheep in a paddock consists of updating a number in working memory with each new object counted (Garavan, 1998). If we count ewes and rams separately, then two numbers must be continuously updated and retained, and that number rises to three if we also keep track of lambs. Likewise, when reading a story, we follow the unfolding course of events by building a situation model that is updated according to the events described (Morrow, Greenspan, & Bower, 1987). While driving, every glance into the rearview mirror updates our representation of the surrounding traffic, thus providing "situational awareness" (Gugerty, 1997). Language comprehension also depends on rapid updating of syntactic representations, for instance when encountering garden-path sentences, such as "while the scientists explored the cave remained undiscovered" (Christianson, Hollingworth, Halliwell, & Ferreira, 2001).

Updating of WM has long been recognized as an important topic of study in cognitive psychology and the neurosciences. One of the first experiments on WM updating was conducted by Yntema and Mueser (1962), who presented participants with a long list of words and asked people to remember the last exemplar in each of a varying number of categories. In the related running memory task, participants are asked to recall the last n items of a long list. This task has been used by Morris and Jones (1990) to study

updating within the "central executive" of Baddeley's working memory model (Baddeley, 1986). It should be noted, however, that the utility of the running memory task as a tool to investigate updating (e.g., Postle, 2003) has recently been questioned (e.g., Bunting, Cowan, & Saults, 2006; see also Geiger & Lewandowsky, 2008).

WM updating has also been identified as one of three correlated factors that capture individual differences in executive functions (Miyake et al., 2000). Indeed, the "updating" factor has been found to be the only executive function that predicts fluid intelligence (Chen & Li, 2007; Friedman et al., 2006). Accordingly, difficulties in WM updating—especially the inability to inhibit information that is no longer relevant—have been proffered as the cause of poor comprehension skills (Carretti, Cornoldi, De Beni, & Romano, 2005; Palladino, Cornoldi, De Beni, & Pazzaglia, 2001). In the neurosciences, Donchin and Coles (1988) proposed that one of the most prominent ERP components, the P300, reflects updating of WM.

The obvious and acknowledged importance of WM updating stands in contrast to our limited understanding of the underlying processes (for similar arguments, see Kessler & Meiran, 2008; Oberauer & Kliegl, 2001). Kessler and Meiran (2008) argued that updating must involve distinct sub-processes because working memory needs to be stable and flexible at the same time. Specifically, when updating information, WM must be protected against interference by outdated information, but it must also allow for the modification of memorial content when appropriate. Those two competing demands are unlikely to be met by a single unitary process, although at present nothing is known about what constituent processes might govern WM updating. Our study seeks to make the first step towards filling this void by (a) experimentally manipulating the components involved in a comprehensive updating task and (b) considering how those components relate to individual differences in WM capacity.

WM Updating and Individual Differences

The consideration of individual differences has a long history in WM research (Daneman & Carpenter, 1980; Fry & Hale, 1996; Just & Carpenter, 1992), and the pattern of correlations between tasks can address issues relating to the number and/or nature of processes or components involved in working memory. For instance, Oberauer and colleagues (Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000; Oberauer, Süß, Wilhelm, & Wittman, 2003) have used such an approach to decompose the working memory capacity construct into functional components (simultaneous storage and processing, relational integration/coordination, and supervision/speed) and content facets (verbal/numerical vs. spatial WM), which were then used to investigate the relations between functional WM components and higher cognitive functions such as reasoning (Oberauer, Süß, Wilhelm, & Wittmann, 2008) or attention (Buehner, Krumm, & Pick, 2005).

We already noted that WM updating (WMU from here on) has been found to predict fluid intelligence (Friedman et al., 2006); however, the issue is far from settled. For instance, in their studies of individual differences in executive functions, Miyake et al. (2000) and Friedman et al. (2006) measured two kinds of executive functions—task shifting and inhibition—by subtracting an experimental condition requiring those functions from a control condition not requiring them. WMU, in contrast, was measured simply by the accuracy on WM tasks that required updating, such as a running memory task (termed "letter task" in their paper). These tasks obviously do not supply a pure measure of WMU, but additionally involve other WM functions such as memory maintenance. Therefore, we do not know whether the correlation between the "updating" factor and fluid intelligence (Friedman et al., 2006) reflects a relationship between intelligence and people's ability to update WM, or simply a correlation between intelligence and WM capacity (WMC from here on). Recently, Schmiedek, Hildebrandt, Lövdén, Wilhelm, and

Lindenberger (in press) have argued that common WM abilities underlie performance in both WMC (e.g., complex-span tasks) and WMU tasks, and that hence both types of task are strongly related and predict higher cognitive abilities to a similar degree.

In contrast, some researchers have claimed that there is only a weak link between WMU and WMC; for instance, Radvansky and Copeland (2001) reported that the ability to update the situational representation of a verbally described, unfolding event was unrelated to WMC. Moreover, it has been suggested that WMU abilities (as opposed to WMC) are relatively preserved in old age (Radvansky & Dijkstra, 2007), but disproportionately affected in schizophrenia (Van Raalten, Ramsey, Jansma, Jager, & Kahn, 2008).

Hence, the relationship between WMU and WMC remains to be examined: On the one hand, prominent theories of WM that assume a tight link between WMC and executive functions (Baddeley, 1986; Engle, Kane, & Tuholski, 1999; Kane et al., 2004) should predict that updating processes are closely related to measures of WMC because updating is regarded as one executive function (Miyake et al., 2000). This is supported by recent evidence (Schmiedek et al., in press). On the other hand, WMU and WMC may be dissociable dimensions of mental ability, as suggested for instance by the results of Radvansky and colleagues (Radvansky & Copeland, 2001; Radvansky & Dijkstra, 2007).

To address this issue, our study included three different working memory tasks, thus permitting a reliable estimate of each participant's WMC. These WMC estimates could then be statistically related to the individual component processes of WMU[1] that were isolated by our experimental manipulations.

Components of WM Updating

In the absence of any relevant precedent or prior theorizing, we began our examination of the components of WMU by conducting a task analysis of previously-used

experimental updating paradigms. This task analysis identified three putative

sub-processes; namely, retrieval, transformation, and substitution. Our experiment

orthogonally manipulated these three components, thus permitting their empirical

identification and assessment of their interrelationship.

First, updating may or may not involve the requirement to retrieve information. For

example, suppose a restaurant manager is told that "Twenty patrons are expected

tonight." This information would need to be updated when the new information "Actually,

it will be 5 patrons more than the initially expected 20" becomes available; however, in

this instance no retrieval of the initial information is required because it is present in the

updating prompt. By contrast, the prompt "Actually, it will be 5 patrons more than

initially expected" requires retrieval of the initial number to ensure accurate updating.[2]

Second, updating may or may not involve transformation. Both the examples given

above involved transformation of information—adding 5 to 20 in order to arrive at the

final result of 25. However, the need for transformation is not inevitable; for example, the

prompt "Actually, it will be 25 patrons" permits updating without the need to transform

earlier information.

Finally, updating may or may not involve a substitution of information. The three

preceding examples all replaced information (25 instead of 20 patrons), but again this is

not inevitable. For example, one may receive the information "Actually, we are expecting

5 groups of 4 patrons", or "Three patrons canceled their reservation, but we got 3 new

bookings". Although these prompts will likely elicit some transformative processing (e.g.,

multiplying 4 by 5), the result does not, however, require a substitution of the old

information.

The various WM updating tasks used in the literature employ these three processes

to varying degrees. Some of the most commonly used updating tasks are summarized in

Table 1. For each task, we have identified the extent to which it involves the three distinct

sub-processes just discussed. For example, the n-back task, in which people are presented with a series of items and have to judge for each one whether or not it is identical to the item presented n steps before, clearly involves a retrieval component—because one has to remember at least the last n items and retrieve the nth one back at every step—but it does not involve any transformation because the information must be retrieved in a form that is identical to the form in which it was presented. By contrast, the Memory Updating (MU) task involves both retrieval and transformation. In the MU task, people have to hold in memory a set of items (typically single digits), each presented in a separate frame, and are then presented with a prompt (e.g., "$? + 1$") directing them to retrieve the content of one particular frame and transform this information as per the prompt, and from then on retain the result of that transformation in memory.

The table clarifies that there is considerable cognitive heterogeneity among updating tasks. Some tasks involve all three processes whereas others require only a subset of two of our processes. Moreover, these three component processes are not necessarily unique to updating tasks; for instance, retrieval by definition obviously plays a role in virtually all tests of memory. It follows that the heterogeneity in the observed relationship between updating and WMC may partially stem from differences between the various WMU tasks. In particular, the different tasks may tap the proposed underlying sub-processes of updating in different ways or to different extents. In order to shed more light on these component processes and their relations to higher cognitive functioning, we aimed to decompose the processes involved in a single updating task and to investigate their contribution to updating performance, and also their covariation with general WM capacity.

The Present Study

The present study investigated the three component processes potentially required for performance in WMU tasks—retrieval, transformation, and substitution. Our goal was to isolate these components from other processes and variables and to observe their interactions. To this end, we used a variant of the MU task previously used by Salthouse, Babcock, and Shaw (1991) and Oberauer et al. (2000). Participants memorized three letters and then performed a sequence of operations that selectively required updating of one of the memoranda, and that variously did or did not require retrieval, transformation, and substitution. Hence, this task—described in detail below—involved multiple processes. In order to assess the interplay among the processes, we used a multi-level regression approach.

In addition, we combined those experimental manipulations with psychometric analysis methods to isolate the variance of each of these components of WMU and relate them to WMC. To do so, participants also completed a battery of working memory measures in a separate session. The covariation of WMU components and WMC was investigated with structural equation models.

Method

Participants

One hundred and four third-year psychology students from the University of Western Australia participated in the two experimental sessions for partial course credit. Data from 7 subjects were discarded either due to technical failure (2 subjects) or because their performance in the WMU experiment was classified as an outlier (5 subjects; see below for details). The analyses are therefore based on the data from the remaining 97 participants (13 males, age range 19-41, mean age 21.2 years).

Apparatus

Both experimental sessions were controlled by a Matlab program designed using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Subjects were tested individually in sound-attenuated booths. Subjects sat about 70 cm from a 17 inch TFT monitor.

Working Memory Capacity Measures

The first experimental session lasted 30 minutes and involved measurement of each participant's WMC using a battery of three tasks.

Stimuli, Design, and Procedure

The three WMC tasks were taken with slight modifications from Oberauer (2005). The tasks were: Operation Span (OS), Sentence Span (SS), and Spatial Short-Term Memory (SSTM), and they were always administered in this order.[3]

Operation span (OS). This task was originally designed by Turner and Engle (1989). On each trial, participants saw an alternating sequence of arithmetic equations (e.g., $3 + 2 = 5$) and the to-be-remembered consonants (all consonants were used except Q and Y, but subjects were not made aware of this additional constraint). Participants had to judge the correctness of each equation, and encode the following consonant for later serial recall.

Commencement of a trial was indicated by a fixation cross presented for 1.5 seconds. Then, the first equation appeared centrally on the screen. It disappeared when subjects made a response or after the maximum response time of 3 s had elapsed. Subjects used the "?/" and "Z" keys to make "Yes, this is correct" and "No, this is not correct" responses, respectively. Keys were labeled with "Y" and "N" accordingly.

After the equation disappeared, a consonant was presented centrally for 1 second. After a 100 ms blank interval, the next equation appeared. This sequence repeated 4 to 8

times, depending on list length. Following list presentation, recall of the letters was prompted with a question mark and a blinking underscore. Subjects then typed the remembered series of letters in their order of presentation; all letter keys were accepted but non-letter keys were blocked. Every typed letter appeared next to the question mark for 200 ms. Subjects had to type as many letters as were actually presented in the trial. They were informed that the order of letters mattered, and were hence instructed to guess if necessary rather than skip letters they could not remember. There was no timing constraint for recall. The inter-trial interval was 500 ms; there was a self-paced break after every 3 trials.

List length (i.e., the number of equations and letters) ranged from 4 to 8. There were 15 trials altogether, 3 trials per list length. The first equation operand was randomly drawn from the 1 to 10 range, and the second operand was drawn from the -9 to $+10$ range (excluding 0), and results ranged from $+1$ to $+20$. Half of the equations were correct. Three practice trials (with list lengths 3, 4, and 5) preceded the experimental trials.

Letter sequences, equations, and trial order were placed in one random order that remained the same for all subjects. The consonant lists for each trial contained no repetitions, and the random generation of lists was repeated until no common acronyms (such as "JFK") occurred.

Sentence span (SS). This task was a variant of the task originally designed by Daneman and Carpenter (1980) and was very similar to the *OS* task just described, except that the processing task was to judge the meaningfulness of sentences. There was an equal number of meaningful and meaningless sentences, and regardless of meaningfulness, there were 3 types of sentence structures. About a third of the meaningful sentences were so-called garden-path sentences that required updating of an initial parsing solution within the sentence (e.g., "As Toby sang a song played on the radio"). They were paired with

structurally analogous sentences that were meaningless regardless of how they were parsed (e.g., "As Dan drank the milk rolled over the hill"). Another third of the sentences had the same structure but were disambiguated early by an additional pronoun with reference to the object, thus avoiding the garden path (e.g., "While Susan wrote the letter it fell off the table" vs. "As the chef stirred the soup it veered into the ditch"). In the remaining third of the sentences, the pronoun referred to the subject ("While Lisa drank the water she drove down the street" vs. "As the man walked the poodle he slept calmly").

All sentences contained between 8 and 11 words, and sentence length, use of "while" vs. "as," and use of concrete first names were broadly counterbalanced across sentence categories. Because this processing task was more difficult than processing of equations, the maximum response time to sentences was set to 5 seconds, and list lengths only ranged from 3 to 7. The practice phase comprised 3 trials of set sizes 2, 3, and 4. In all other respects, the SS and OS tasks were identical.

Spatial short-term memory (SSTM). This task followed closely the original version by Oberauer (1993). Participants had to remember the location of a number of dots in a 10×10 grid. Following central presentation of an "Alert!" message for 1 second, the grid was shown, and a variable number of solid dots appeared, one by one, in individual cells of the grid for 900 ms each (inter-stimulus interval 100 ms).

Subjects were instructed to remember the spatial relations between the dots. That is, absolute dot positions were irrelevant, only the overall pattern of dots was to-be-remembered. After all dots were presented, subjects were cued to reproduce the pattern of dots. The cue was presented using a pattern mask that was the same size as the grid.

Subjects reproduced the remembered pattern of dots by clicking the cells using a standard computer mouse. Clicking on a dot again made it disappear, and subjects were allowed to correct the generated dots until they were satisfied with their response. They

proceeded to the next trial by clicking a "Next" button at the bottom of the screen. No feedback was given.

The number of to-be-remembered dots varied from 2 to 6, and there were two spatial distance conditions. In the "near" conditions, all dots of a trial occurred within a 5×5 area (i.e., a quarter of the grid), in the "far" condition, dots could be placed anywhere in the grid. There were 3 trials for each combination of set size and spatial distance condition, and hence 30 trials altogether. Dot positions were generated at random, with the constraint that no dots appeared in corner positions to avoid verbal coding. Again, the order of trials and dot sequences were fixed for all subjects.

<div align="center">Working Memory Updating</div>

The aim of the second experimental session was to disentangle the three presumed updating processes retrieval (R), transformation (T), and substitution (S).

Design and Stimuli

We modified the Memory Updating (MU) task originally designed by Salthouse et al. (1991) and adapted for psychometric purposes by Oberauer et al. (2000). In this task, subjects encode an initial set of items and subsequently update them repeatedly. Hence each MU trial consists of an encoding phase, an updating phase, and typically a final recall phase. We created 8 experimental conditions by fully crossing the 3 factors R, T, and S while keeping other variables constant.

On each trial, subjects were presented with an initial set of 3 letters (ranging from A-Z), each presented in a separate rectangular frame on the screen (in a single row). To avoid re-use of the same letter during updating, the minimal alphabetic distance between the three input letters was 5 (i.e., with A in frame 1, the closest frame 2 and 3 letters were V and F, F and K, or V and Q; A and Z were considered neighbors because the alphabet "wrapped around", see below).

Thus, participants always had to remember (and then update) three items. Previous research (cf. Oberauer et al., 2000) has shown this set size to provide an intermediate difficulty level ideal for the study of individual differences, as it avoids floor or ceiling effects. Oberauer (2002) additionally showed that the latency of individual updating steps is affected by whether or not they involve switching from one element in WM to another. Because we were not interested in those effects here, we held switching constant by moving to a new (randomly chosen) item on each step.

All updating operations were cued by displaying the appropriate prompt in the frame that was to be updated. Each trial involved 6 updating steps. Participants typed the result of the update (i.e., by pressing the corresponding letter on the keyboard) at every updating step. The 6 updating steps were followed by a recall test of the final contents of all frames.[4]

There were 8 different updating conditions. From these conditions, all possible subsets of 6 were implemented across trials, resulting in 28 $[\binom{8}{6}]$ trials. This implies that no updating condition was repeated within a trial, and that each condition occurred 21 $(28*6/8)$ times in total. The conditions were created by fully crossing the three factors R, T, and S in a within-subjects design. We identify conditions by referring to these 3 letters; for example, the experimental condition involving all three processes is labeled **R-T-S**; the condition featuring only a transformation is labeled r_{no}-**T**-s_{no}, etc. The experimental conditions are summarized in Table 2, which presents the stimuli shown on a given updating step assuming that the letter "C" is the currently remembered content of the targeted frame.

Conditions involving *retrieval* required subjects to retrieve the most recent letter of the cued frame from memory to perform the current operation. In contrast, this letter was provided with the cue in the no-retrieval conditions, meaning that the operation could be executed without retrieving the previous content of the frame. *Transformation* conditions

involved a transformation of the selected letter by alphabet arithmetic. Only positive operations of +1 and +2 were used (with equal probability), and the alphabet "wrapped around", such that $Z + 1 = A$. *Substitution* conditions resulted in the replacement of memory content with new information, whereas the outcome of no-substitution steps was identical to the information already held in memory.[5]

Condition r_{no}-t_{no}-s_{no} involved neither of the three processes, and the currently remembered letter was simply presented again. This condition can be regarded as a baseline. Condition r_{no}-t_{no}-**S** was identical to baseline except that a different letter was presented, which replaced the one in memory (pure S without R or T). In condition **R**-t_{no}-s_{no}, a "?" prompted subjects to retrieve the currently held letter and report it (pure R without T or S). Conditions **R**-**T**-**S**, r_{no}-**T**-**S**, and r_{no}-**T**-s_{no} involved standard alpha-arithmetic operations. For instance, in **R**-**T**-**S**, subjects had to add a number to whatever they currently remembered for that particular frame. In r_{no}-**T**-s_{no}, there was a transformation, but because its result was identical to the remembered letter, there was no substitution. Condition **R**-**T**-s_{no} was designed as a transformation that does not involve substitution and hence does not change memory content ("? + 0"). In condition **R**-t_{no}-**S**, subjects were presented with an arrow that linked one frame with another, indicating that they should retrieve and then copy the letter from one frame to the other as indicated by the arrow, thus requiring a retrieval and a substitution but no transformation (other than a spatial "relocation"). Note that conditions **R**-t_{no}-**S** and **R**-**T**-s_{no} were required to permit the orthogonal combination of all three experimental variables, although this mandated the use of operations that differed from those in the other conditions (e.g., "+0" to avoid a substitution). We deal with the implications of these design decisions during data analysis.

Procedure

The trial sequence (of two sample trials) is illustrated in Figure 1: Each trial was initiated by a key press. The starting letters were presented in their frames, all at once, for 2 seconds. Then the frames were cleared and the first updating instruction (as summarized in Table 2) was displayed in a randomly selected frame. Participants had to carry out the operation required by the instruction, type the result within a specified time limit (explained below), and remember the result as the new content of that frame from there on. If no response was detected within the time limit, an error was recorded and the next step commenced with display of a new instruction in a new frame. No feedback was given after a response, and responses were not echoed on the screen. The instruction for each step was presented immediately after the response of the previous step. After 6 such steps, final recall was signaled by pink question marks presented one by one in each of the frames in random order, and participants were required to type the remembered content for that frame.

The timing constraint for recall was 5 seconds per frame. Feedback was provided after the final memory test at the end of each trial, showing participants the number of correct responses they obtained out of a total of three. The inter-trial interval was 2.5 seconds.

Before the experimental trials began, participants were given a number of practice trials (three with only one frame, three with two, and four with three frames). All eight conditions were covered in the practice trials. There was a break after the practice phase, and also halfway through the session. The session took approximately 25 minutes.

To account for individual differences in typing speed, time limits for typing of the results of individual processing steps were calculated on an individual basis. The time limit was three times the median of RTs in the three-frame practice trials, calculated separately for conditions involving and not involving transformation (pilot testing

suggested that transformation would have a significant impact on RT). Mean deadlines were 7.94 s and 4.91 s for transformation and no-transformation steps, respectively. This manipulation ensured that all participants were given sufficient time to complete the operations during the experiment, without however providing them with excess slack time during which to rehearse.

Results

Working Memory Capacity Measures

Dependent measures in all WMC tasks were final correct-in-position recall scores (with the exception of the span task processing components, where the dependent measure was simply the rate of correct responses). All WM tasks were scored using partial credit scoring (cf. Conway et al., 2005). For instance, a subject correctly remembering 5 out of 6 letters in a span-task trial would score 5/6 on that trial, and the total score would be calculated as the mean of these partial scores across trials. Chance performance level in all tasks was close to zero (e.g., $1/21 = .05$ in the span tasks, as no vowels were used). Descriptive results of the WMC battery are summarized in Table 3. The task inter-correlations are given in Table 4.

Working Memory Updating

As mentioned earlier, data from 5 participants were excluded from all analyses because they failed to comply with updating instructions in Condition **R-T-S** of the WMU session, as revealed by an updating performance of below .2 in that condition (no such problems occurred in the other conditions). The mean performance of the 5 eliminated subjects was .191, which was 2.7 standard deviations below the condition mean.

Descriptive Data

Mean updating performance and RTs are shown in Table 5. RT data are based on correct updating steps only. Responses with RTs below 300 ms were excluded from consideration.

Regression Analyses

We first analyzed the data via multi-level regression. Multi-level regression (also known as "hierarchical" regression or "mixed-effects" modeling; see, e.g., Pinheiro & Bates, 2000) permits an aggregate analysis of data from all participants without confounding within- and between-subject variability, and has been used previously to analyze data in short-term memory (for details, see e.g., Lewandowsky, Brown, Wright, & Nimmo, 2006). We used the lmer function of the R programming language (R Development Core Team, 2008; Bates, 2007) for the multi-level regressions; *p*-values were obtained via Markov Chain Monte Carlo simulations (5000 samples) using the mcmcpvalue function (Plummer, Best, Cowles, & Vines, 2008).

In both accuracy and RT analyses, the independent variables (R, T, and S) were dummy coded (e.g., 0 and 1 for absence and presence, respectively, of a substitution). The variable T, however, was further decomposed. First, we used dummy variables T1 and T2 to accommodate the fact that the difficulty of a transformation varies with the size of the operand (Zbrodoff, 1999). We used T1 as the baseline, so all transformation conditions were dummy-coded with T1 = 1, and all remaining (non-transformation) conditions were set to T1 = 0. Transformations involving a +2 operation were additionally coded as T2 = 1. That is, T1 and T2 were coded incrementally, so a +2 operation would be coded as 1 on both T1 and T2 factors whereas a +1 would be coded as 1 and 0, respectively. For the zero-transformation (? + 0) condition **R**-**T**-s$_{no}$, we further introduced a dummy-coded variable T0. Although technically involving a transformation, this condition required little

more than retrieval of the current frame content, and therefore is likely to take less time, and incur fewer errors, than adding 1. T0 was set to 1 for condition **R**-**T**-s$_{no}$ and was 0 for all other conditions (so a +0 operation was coded 1, 1, 0 on T0, T1, and T2).

Finally, condition **R**-t$_{no}$-**S** involved not only retrieval and substitution, but also—unlike the other conditions—an attentional shift from one frame to another. Therefore, a switch parameter SW was incorporated to account for these extra frame switching costs. Similar to T0, this parameter was dummy coded 1 for condition **R**-t$_{no}$-**S** but was 0 for all others.

¿From our task analyses, the most straightforward and parsimonious assumption regarding the 3 processes retrieval, transformation, and substitution would be to assume "chained independence": a serial chain of principally independent processes. That is, one needs to retrieve information before one can transform it, and one needs to await the result of the transformation before one can use this result to substitute outdated information. Despite this "chaining" due to obvious sequential constraints, we assume the processes per se to be independent. In terms of RT, such independence would be reflected in an additive regression model with no interactions. If one process takes x ms and another process takes y ms, then the two together should take $x + y$ ms. In contrast, in terms of updating accuracy, independence would best be captured in a multiplicative model. If one process has a success probability of x and another process has a success probability of y, the probability that both will succeed is $x * y$. The probability of completing the whole step accurately is the product of the success probabilities of all component processes involved.

In the RT model, the beta weights of R, S, and T1 represent the estimated durations of retrieval, substitution, and the +1 transformation, respectively. In the accuracy model, these beta weights represent the estimated success probabilities of the same processes. The T2 and T0 codes require a somewhat different interpretation because they do not represent processing steps but rather serve to modify T1. In the RT model,

T2 represents the additional time it takes to add 2 over and above adding 1, and T0 represents how much less time +0 takes compared to +1. Therefore, we expect that the beta weight of T0 will be negative, whereas all other beta weights should be positive.

Likewise, in the accuracy model, T2 represents the proportional decrease of success probability for adding 2 relative to the success probability of adding 1, whereas T0 represents the proportional increase of success probability for the +0 zero-transformation relative to the success probability of +1. Hence, T0 is expected to have a beta weight > 1, whereas all other beta weights are expected to be < 1.[6]

No interaction terms were included into the regression models to begin with. In all models, the independent variables were coded as both fixed and random effects, where the former represent the overall effects of those manipulations in the experiment (hence fixed effects estimate the group mean of each experimental effect and the intercept) and the latter represent individual variation in responsiveness to the experimental manipulations (hence random effects estimate individual differences in the size of experimental effects, and of the intercept, expressed as standard deviations). We only report the fixed effects here because individual variation is examined later, in the section on structural equation modeling.

The updating accuracy (UpdAcc) model was:

$$UpdAcc = .992 * .886^{R} * .914^{T1} * 1.065^{T0} * .934^{T2} * .972^{S} * .965^{SW}.$$

Note that the dummy variables are exponents in this case; thus a process not involved, coded as 0, figures as 1 in the chain of products and thereby has no effect on the predicted accuracy; an involved process is coded as 1 and hence contributes to the extent specified by the beta weight. This is particularly important to understand the effects of +0, +1, and +2 transformations: A "standard" +1 transformation will reduce baseline accuracy by the factor .91 according to the model (T0 and T2 drop out of the equation); a +2

transformation would reduce accuracy by the factor .85 ($.91 * .93$; T0 drops out), and a $+0$ "transformation" would reduce accuracy only by .97 ($.91 * 1.07$; T2 drops out).[7] This model yielded $AIC = -638.9$, $BIC = -464.9$, $logLik = 354.5$, coefficient of determination (squared correlation of fitted and observed values across all data points) $CoD = .76$. The analysis revealed a significant contribution of R ($t = -8.74$, $p < .001$), T1 ($t = -6.12$, $p < .001$), T0 ($t = 2.39$, $p < .01$), T2 ($t = -4.28$, $p < .01$), S ($t = -3.10$, $p < .01$), but not SW ($t = -1.36$, $p = .08$).

For the RT model, instead of estimating a switch parameter (SW) for condition **R**-t$_{no}$-**S**, we maximized parsimony by using an existing empirical estimate of the time taken to switch attention from one location to another, provided by Garavan (1998). This estimate of 483 ms was subtracted from the condition **R**-t$_{no}$-**S** mean of each subject before fitting the RT model, thus saving a free parameter that otherwise would have been required for SW.

The resulting updating RT (UpdRT) model was:

$$UpdRT = 1.300 + .039\, R + 1.201\, T1 - .941\, T0 + .579\, T2 + .296\, S.$$

Here, the dummy variables can be interpreted as factors; again, they are 1 if the process is involved in a certain condition, and 0 otherwise. The multilevel analysis with $AIC = 1357$, $BIC = 1491$, $logLik = -651.4$, $CoD = .91$, revealed a significant contribution of T1 ($t = 29.02$, $p < .001$), T0 ($t = -13.32$, $p < .001$), T2 ($t = 14.77$, $p < .001$), and S ($t = 12.11$, $p < .001$), but no effect of R ($t < 1$).

We next examined whether the fit of these models could be improved by adding pairwise interactions between the three experimental variables R, S, and T1.[8] Using likelihood ratio tests, we determined that for both the accuracy and the RT model, there was no case in which the addition of interactions (we added between one and three interaction terms) improved the fit of the model, all $\chi^2 < 3.4$, $p > .05$, suggesting that our

three experimental variables were independent.

Summary of regression models. We fit regression models to both accuracy and response times. The analyses revealed that the effects of transformation, retrieval, and substitution were independent and showed no sign of interaction for either accuracy or RT. In terms of updating accuracy, the largest effect was associated with retrieval, which introduced errors, leading to a drop in performance by over 10 % whenever information had to be retrieved. Perhaps surprisingly, this performance drop was not mirrored in response latency, as we found that the requirement to retrieve information from memory did not slow people at all.

By contrast, the requirement to substitute information in memory was a time-consuming process, taking about 300 ms irrespective of whether the new letter was generated by a transformation or simply presented. Unlike retrieval, substitution of information had only a small effect on the accuracy of updating. The fact that substitution did not affect accuracy much is perhaps not surprising when we consider that errors in substitution primarily affect the accuracy on future steps, not necessarily on the present step. People can enter the correct letter and then fail to establish that letter as the new content of the current frame in their memory, which would lead to an error on the next time they retrieve the content of the same frame.

Finally, as expected, transformations had a large impact on RT—the more demanding the transformation of memory content, the more time it required. Transformations also contributed to errors, especially when the transformation involved moving 2 steps up in the alphabet. We defer further discussion of these results until all analyses have been presented.

Linking WMC and WM Updating

To investigate the relation between working memory capacity and updating, we used structural equation modeling (SEM). SEM is typically used to capture individual differences and correlational dependencies between latent variables, without regard to experimental manipulations or differences between means. In the present case, we extended this standard approach by also modeling *mean* RT and accuracy for each experimental condition. We thus constructed an SEM model that concurrently captured both inter-individual variation *and* experimental effects. The general approach has been outlined in Oberauer, Wilhelm, and Schmiedek (2005).

On the basis of the preceding regression analyses, we initially focused on models that preserved the independent (additive) structure among experimental variables as far as possible (note that again, a technically additive model was applied to log-transformed accuracies to capture the multiplicative relationship).

The SEM models for accuracy and RT are depicted in Figures 2 and 3, respectively. To maximize readability, the figures show all significant **standardized** weights between latent variables (in bold, drawn next to the appropriate links) and all estimated *unstandardized* means of the latent variables (in italics, drawn inside the variables themselves; note that latent means not shown in the figure were set to zero). To prevent clutter, no other estimates are shown in the figures. Those remaining estimates are instead presented in Tables 6, 7, and 8.

The models consisted of two measurement models—one for the WMC part and one for the updating part. The WMC part had 3 manifest (i.e., observed) variables—one for each task of the WMC test battery—which were linked to a single *WMC* latent factor. The updating part had 8 observed variables—one corresponding to each of the 8 experimental conditions—that were connected to latent factors corresponding to the experimental variables R, T, and S, and a general factor accounting for the general level of

performance (this corresponds to the intercept in regression).

Measurement Models

   Measurement model for working memory capacity. In the WMC measurement model, the mean of *WMC* was set to 0 and its variance to 1, to provide a scale for the factor and allow for identification. This also made the unstandardized solution more interpretable, which we were particularly interested in as it provides intercept and slope estimates on the original scale of measurement (see below for further details).

   Measurement model for working memory updating. Models with the same basic structure were applied to the mean RTs and log-transformed accuracies, as well as their variances and covariances, in the eight design cells. In Figures 2 and 3 these 8 manifest variables are labeled by the experimental condition code they refer to (e.g., **R**-**T**-s$_{no}$).

   To confirm the reliability of measures obtained from the experimental WM updating task, we split the 28 trials into 4 sets of 7 and calculated Cronbach's α on the means of each subset of trials. This revealed an α of 0.96 for the RT data and an α of 0.89 for the accuracy data, demonstrating high internal reliability.

   There were four latent variables. One of them reflected baseline performance common to all design cells, and therefore all manifest variables had loadings on it. The remaining three factors reflected the experimental manipulations (R, T, and S). The factor R received loadings from the four manifest variables representing performance in conditions that required retrieval (i.e., conditions **R**-**T**-**S**, **R**-t$_{no}$-**S**, **R**-**T**-s$_{no}$, and **R**-t$_{no}$-s$_{no}$), but not from the remaining four manifest variables representing conditions without retrieval (implying that their loadings were fixed to 0). In the same way, variables reflecting conditions with transformations received loadings on factor T, and variables with substitution received loadings on factor S.

Constraints on the Models

We imposed a number of strong constraints on our models that are typically absent in SEM applications. First, all loadings between manifest and latent variables in the WMU measurement models were fixed to 1. In addition, the intercepts of the manifest variables were fixed to zero, as were the means of the error terms associated with them.

The rationales for these constraints were as follows. First, the additivity suggested by the regression is captured by the fixed loadings because they imply that each factor has the same effect on all conditions that load on it—for instance, substitution increases RT by the same amount in all four design cells involving substitution. Second, fixing the intercepts of manifest variables and their error means to zero implies that no individual condition had a mean higher or lower than that predicted from the additive model.

Thus, the estimated mean of the General factor represents the baseline level of performance (i.e., the intercept of the additive model), and the estimated means of the three factors that represent components of updating reflect the mean effect of each experimental manipulation. Correspondingly, the estimated variances of the factors reflect the individual differences in baseline performance and in the magnitude of the experimental effects, respectively. For both RTs and accuracies we had to relax these strong constraints of the purely additive model at a few points, either based on theoretical considerations or, on a few occasions, based on deviations of the data from the additive model, as detailed below.

Finally, in the overall structural models presented below, the co-variances between the latent updating variables and the latent *WMC* variable reveal the extent to which individual variation in working memory capacity relates to variation among individuals' responses to our experimental updating manipulations.

The Accuracy Model

For the accuracy model, the link between T and the manifest **R**-**T**-s$_{no}$ variable was allowed to vary freely, reflecting the assumption that the "zero-transform" condition **R**-**T**-s$_{no}$ entailed a smaller effect of T than the other three conditions involving a transformation. This represented a slight departure from a strict additive structure. Moreover, to achieve an acceptable model fit, correlations were introduced between error terms that arose from highly similar experimental conditions *e2* and *e6*, and also *e6* and *e7*. Finally, we set the variance of S to zero because when it was freely estimated, the variance was slightly (and non-significantly) negative. The implication of a zero variance estimate is that S contributed a constant effect to performance without accounting for any inter-individual variability.

We first fit the updating measurement model to the experimental data, to ensure model identification (identification of the WMC measurement model was trivial because that model had 0 *df*s). The fit of the updating measurement model was acceptable, $\chi^2(24) = 43.59$; comparative fit index $(CFI) = .96$; root-mean-square error of approximation $(RMSEA) = .09$, standardized root-mean-square residual $(SRMR) = .052$. After this was established, the measurement models were combined to yield the structural model.

Initially, *WMC* was linked to all updating-component factors, and links were then removed if they were insignificant and did not contribute to the model fit. Likewise, new links that were theoretically and empirically justifiable were added to improve the overall fit. The fit of the final accuracy model—depicted in Figure 2—was acceptable, $\chi^2(45) = 65.27$; $CFI = .97$; $RMSEA = .068$; $SRMR = .051$. The unstandardized coefficients of the accuracy model are summarized in the top panel of Table 8. To evaluate model fit correcting for non-normality of the data, we performed a Bollen-Stine bootstrap (with $N = 1000$ samples), which indicated a good model fit.

We first consider the pattern of correlations among latent variables (shown in Figure 2). *WMC* was found to correlate positively with overall updating accuracy (*GenAcc*), implying that people with high WMC have higher baseline performance on the WMU task. *WMC* also correlated positively with R and T. This implies that the accuracy cost incurred by the need to retrieve or transform information declined with increasing WMC. Furthermore, R was positively correlated with T and *GenAcc*. Hence, higher costs of retrieval were typically accompanied by higher transformation costs, and lower levels of baseline performance. Although S contributed significantly to mean WMU accuracy, it had no variance and therefore could not correlate with *WMC*.[9]

Now consider the pattern of estimated means for the latent variables (shown inside the variables in Figure 2) which permit reconstruction (i.e., prediction from the model) of all observed performance variables. For example, if an operation involves T and S but not R, the estimated updating accuracy can be calculated by adding the (negative) contributions of T and S to the baseline performance variable *GenAcc*—any additional need for processing is assumed to reduce both accuracy (towards 0) and log-accuracy (towards $-\infty$). Hence, we can calculate the log-accuracy in this case as $-.06 + (-.11) + (-.01) = -.18$. We can then retransform the log-value into an accuracy estimate of $exp(-.18) = .84$ (alternatively, one could first retransform each log-operand and then multiply the resulting variables as in the corresponding regression analysis)—which corresponds closely to the observed level of performance (which was .87 in this case; see Table 5). The mean absolute deviation of the SEM estimates from the observed values across all conditions was $< .05$ accuracy units, suggesting that the SEM model captured mean performance in all 8 cells with good precision.

The RT Model

The principal architecture of this model was identical to that of the accuracy model. Again, the link between T and the manifest $\mathbf{R}$-$\mathbf{T}$-s$_{no}$ RT variable (i.e., the T0 transformation) was freely estimated.

Error terms associated with the manifest variables were fixed to have a mean of zero, with the exception of *e3*, which was set to .483—the estimate of attentional-shift cost provided by Garavan (1998) that was used earlier in the regression analysis.[10] Again, in order to achieve an acceptable model fit, correlations were introduced between error terms that arose from highly similar experimental conditions, namely *e4* and *e8*, and also *e2* and *e6*.

Again, the memory-updating measurement model was fit to the data first, and the fit was found to be acceptable, $\chi^2(24) = 63.28$; $CFI = .94$; $RMSEA = .131$, $SRMR = .075$. After this was established, the measurement models were combined to yield the structural model. The fit of the final RT model—depicted in Figure 3—was acceptable, $\chi^2(47) = 87.23$; $CFI = .94$; $RMSEA = .094$; $SRMR = .084$. The unstandardized coefficients of the RT model are summarized in the lower panel of Table 8. A Bollen-Stine bootstrap ($N = 1000$) again indicated good model fit.

Latent correlations are shown in Figure 3. A positive correlation was found between *GenRT* and *T*. Hence, longer transformation times were associated with longer baseline performance latencies. *WMC* correlated negatively with overall latency (*GenRT*; this covariance was marginally significant after removing the nonsignificant link between *WMC* and *T* but significant before), implying that people with higher WMC had lower baseline RTs. Although *S* contributed significantly to WMU mean latency and had a significant variance term, this experimental effect did not covary with *WMC*.

The estimated means shown in the figure again permit accurate reconstruction of all observed data. For example, if an operation involves T and S but not R, the RT predicted

by the model is calculated as the $GenRT$ mean plus T and S means, or

$1.275 + 1.470 + .335 = 3.08$—the observed level of performance was 3.01 seconds (see

Table 5). The mean absolute deviation of the SEM estimate from the observed values

across all conditions was 232 ms. The only conditions whose estimates deviated

considerably from the data were conditions **R**-t$_{no}$-**S** and **R**-**T**-s$_{no}$. The mean deviation

across the remaining 6 conditions was only 30 ms.

Summary of SEM Modeling

Our SEM models are readily summarized: First, they confirmed the independence of

experimental variables that was suggested by the regression analyses. At the level of

means, they provided an excellent fit of the data from the updating session. With the

exception of RT estimates for conditions **R**-t$_{no}$-**S** and **R**-**T**-s$_{no}$, SEM mean predictions

were close to the observed condition means, with average deviations of 30 ms and below

.05 accuracy units, respectively. Mirroring the regression analyses, we again found that

transformations had a strong impact on both accuracy and especially RT, whereas

substitutions had small but reliable effects, and retrieval had a large effect on accuracy

but no effect on RT.

Second, the models captured the variance among individuals' WMC and related it

to the variation in the magnitude of the experimental effects. That is, in the accuracy

model, WMC correlated positively with overall accuracy on the updating task ($GenAcc$)

as well as retrieval success (R) (and also transformation success reflected in T, although

this relation was apparently less reliable). In the RT model, a relation between $GenRT$

and T could be observed, such that the higher the baseline RT reflected in $GenRT$, the

larger the impact of T. The RT model additionally showed that WMC correlated

negatively with overall updating latency $GenRT$. In both models, S contributed

significantly to WMU performance, but did not covary with WMC.

A possible criticism of our SEM models might point to the fit statistics reported above, which did not always satisfy the criteria for a "good" fit (e.g., the requirement that $RMSEA < .06$, Hu & Bentler, 1999). In response, we emphasize that in order to capture both the inter-individual (co-)variation in WMC and WMU *and* the effects of our experimental manipulations, we made use of novel and highly constrained models (e.g., loadings for the S, T, and R latent variables were fixed at 1). This reduces the number of degrees of freedom, and hence heuristics for "good fits", which are obtained from Monte Carlo simulations involving models of variance-covariance matrices with unconstrained loadings, cannot be directly transferred to our study.

Nonetheless, with the aid of modification indices, the fit of both models could be improved considerably, to the point where they met fairly stringent goodness-of-fit criteria, by introducing additional covariance links between error terms in the WMU measurement model.[11] None of those modifications altered the basic psychological structure of the models; in all cases, the relevant weights between latent variables retained their significance. These changes did, however, come at the cost of losing parsimony and interpretability, and therefore we do not focus on those best-fitting models. We conclude that our models capture the relationships between experimental variables and individual differences in a highly parsimonious way while achieving acceptable fit.

General Discussion

The present study makes three principal contributions.

At the level of methodology, we have utilized structural equation modeling (SEM) in a novel way, by modeling individual differences as well as experimental manipulations simultaneously. Both inter-individual variability and experimental effects on performance measures (RT and accuracy) were described by parsimonious models.

At an empirical level, we presented the first examination of the basic component

processes that govern performance in common working memory updating tasks. Our putative decomposition of WMU task processes into three distinct components—retrieval (R), transformation (T), and substitution (S)—was tested by examining every possible combination of these three components. The data indicate that these component processes make distinct and independent contributions to WMU performance.

Finally, we used the decomposition of WMU to analyze the relationship between these processing components and working memory capacity. As noted at the outset, both the view that WMC and WMU are strongly related and the view that they form dissociable dimensions of mental ability have been represented in the literature (e.g., Friedman et al., 2006; Radvansky & Copeland, 2001). Our study may help resolve this inconsistency because while we found that WMC did predict performance in the experimental WMU task to some degree, we also found that some processing components involved in that task were independent of WMC.

We now take up those principal issues in turn.

Modeling Experimental Outcomes with SEM

The vast majority of SEM applications focus on capturing individual variation alone. That is, conventional SEM models seek to describe the variance-covariance structure among a number of measured (manifest) variables by postulating one or more unobserved (latent) variables, each of which captures variance that is shared by several manifest variables. One often-cited benefit of SEM is that latent variables are free of measurement error that besets interpretation of pairwise correlations (cf. Blunch, 2008).

Although data collection for those studies often involves computer-controlled laboratory tasks, these studies typically are not experiments in the strict sense because no independent variables are manipulated. Recently, Oberauer et al. (2005) suggested that SEM constitutes a powerful tool also in an experimental context, because it permits

modeling of the experimental effects not only at the level of means (as in the conventional ANOVA) but also at the level of variances and covariances.

The present study takes up this suggestion, and we therefore highlight one more conclusion from our analysis that became possible only because of our use of SEM. At the outset, we conducted a task analysis that putatively identified three *potential* component processes of WMU. Our experiment then sought to examine those processes by manipulating independent variables that selectively tapped each putative component. The unique contribution of the SEM analysis is that it identified those putative components as valid latent constructs—that is, notwithstanding the strong constraint that all loadings (barring the necessary exceptions noted earlier) were set to unity, our SEM models fit the data well. Because the models, unlike conventional ANOVA, also captured the individual variation and covariation of the effects of our experimental variables, the models support the psychological reality of our putative constructs with greater confidence than would be possible based on a conventional means-only analysis. In a nutshell, the factors S, T, and R in our SEM models identify psychological processes rather than just independent variables in an experiment. We can therefore now turn to a characterization of those component processes.

Component Processes

The analyses revealed specific costs associated with each component of WMU. Retrieval and transformation, and to a lesser degree substitution, affected the accuracy of processing steps in our WMU task. In contrast, transformation and—again to a lesser degree—substitution, but not retrieval, determined WMU response times. Hence, transformation of memory contents seems to be the main determinant of processing in a generic WMU paradigm, retrieval seems to be error-prone but not time-consuming, and substitution also has a significant, albeit smaller, impact on WMU task performance.

As the fit of both the linear regression models and the structural equation models was good without including any interactions among components, we suggest that the contributions of retrieval, transformation, and substitution to WMU are orthogonal and independent. This observation has considerable theoretical implications.

A priori, there are three possible scenarios describing the relationship between component processes. First, the processes can run completely in parallel, starting at the same time. For our updating paradigm, this scenario can be ruled out on logical grounds: information must first be retrieved before it can be transformed or used for substitution, and it must be transformed before the result of the transformation can be used for substitution. The second scenario is that processing stages are strictly serial—meaning that the next process does not start before the preceding one is completed. This scenario implies that the time used by one component does not influence the time required for subsequent processing components, so that their contributions to RTs are additive (and their accuracies multiplicative). The third scenario is a cascade of processing such that later stages can be initiated before the completion of earlier stages (cf. McClelland, 1979; Meyer, Irwin, Osman, & Kounois, 1988; Townsend & Fific, 2004; Ratcliff, 1978). Many cognitive operations likely run in cascades (for instance, naming and speech production, Navarrete & Costa, 2005; Roelofs, 2008).

The additivity of the three experimental variables on RTs favors a strict serial model over a cascade model. If processes are cascaded, interactions between corresponding parameters are more likely to occur (although this need not *necessarily* be the case, cf. McClelland, 1979). For example, if both transformation and substitution are required for a particular WMU operation, and substitution could start significantly before transformation is completed (i.e., partial information delivered by transformation is already fed into the substitution process), then T and S should interact under-additively: both together should take less time than the sum of the time taken to transform and to

substitute separately, because part of the substitution would already be completed during transformation. No such interactions were observed.

Our conclusions must be tempered by two potential problems: First, the additivity among updating components was based on a limited set of experimental conditions (8 in total), including two that required special consideration during analysis owing to their unique properties (conditions $\mathbf{R}$-t$_{no}$-$\mathbf{S}$ and $\mathbf{R}$-$\mathbf{T}$-s$_{no}$). It is possible that other sets of operations will yield different outcomes; nonetheless, until demonstrated otherwise, we propose that the components of working memory updating operate in a strictly sequential, stage-like manner.

Second, one could argue that our decomposition of WMU processes is too coarse. Indeed, it is plausible to further deconstruct some of the components identified here. For example, the substitution process may in turn involve the *removal* of old information followed by the *addition* of new information. Precedents for this possibility exist; for example Postle (2003) described "discarding" as a potential memory-updating process. Instantiations of such a targeted removal process can be found in models that embody the notion of response suppression (e.g., Lewandowsky & Farrell, 2008). Similarly, Oberauer and Vockenberg (2009) suggested that binding processes—i.e., binding a piece of information to a certain position in a spatial or temporal mental "map"—lie at the heart of memory updating. If a chunk of information is bound to a particular position in mental space (e.g., this could be an actual spatial location, a temporal position, or some idiosyncratic marker), a substitution would thus require "unbinding" of old information followed by renewed binding of new information.

Likewise, our transformation process can be further analyzed into a retrieval component (albeit from long-term memory), which provides the next letter(s) of the alphabet, plus a counting component that signals when to stop retrieving more letters. We therefore expect that the component processes identified here will be subject to

further examination and refinement; nonetheless, the fact that our data conformed to a purely additive model suggests that any further decomposition of processes would only take place *within* the proposed components (e.g., substitution being partitioned into removal and addition).

<div align="center">*Instantaneous Retrieval?*</div>

One perhaps surprising aspect of the data is that the retrieval component did not affect processing times. This finding seems to imply that retrieval from memory is as fast as the visual encoding of new information. It is worth noting that this finding replicates previous research that found no effect of retrieval on latency of WMU (Bialkova & Oberauer, in press).

What does this result tell us about the structure and representations of working memory? Clearly the **R** conditions of our task requires selective access to the letter in a particular frame, and given that each updating step used a different frame from the one preceding it, access to a new letter is required on every step. Of course, no process can be instantaneous. Hence, the fact that our retrieval manipulation did not incur a time cost implies that access to a new letter occurs not only when the new letter is needed as input for processing, as in the **R** conditions, but not in the r_{no} conditions when the letter is not required as input. This conclusion suggests that objects in WM are tightly bound composites of content elements (in our case, letters) and the context by which they can be retrieved or accessed (in our case, the spatial location of the frames), such that retrieving or accessing one piece of information involves obligatory retrieval of, or access to, the whole object. As a consequence, all conditions in our experiment induced access to a new object merely by bringing a new context into the focus of attention. We now discuss two models that can accommodate this finding.

Oberauer (2002, 2003, 2006) proposed that there are three functionally distinct

layers of representation in working memory: (1) the activated part of long-term memory, which holds information that is likely to be relevant for the current task, (2) the region of direct access, which holds a limited set of representations for ongoing processing, and (3) the focus of attention, which holds the one piece of information already selected for the next cognitive operation. In this model, the focus of attention can only select an item from the region of direct access. Making a representation available for processing therefore involves two steps. The first is to retrieve a representation from (activated) long-term memory, thereby bringing it into the region of direct access; the second step is accessing one representation among several candidates in the direct-access region. Within this model, we can explain our results as follows: All three letters are continuously held in the direct-access region. Every updating step involves presentation of a stimulus in a frame different from the preceding frame, and the new frame serves as an automatic trigger for the focus of attention to focus on that frame. Shifting the focus incurs a time cost, as reported by Garavan (1998), but this time is a constant in our experiment because all conditions involved a shift to a new frame. The critical assumption is that moving the focus of attention to a new frame means not merely to move visual attention to a new location in space (a process that would require an order of magnitude less time than the 500 ms for a focus-switch in working memory), but rather involves focusing on a new object in WM. That object is a composite consisting of the spatial location of the frame and the letter currently bound to it; hence it is immaterial whether or not that letter has to be "retrieved" for the next processing step as it is already in the focus of attention.

An alternative model was proposed by McElree based mainly on examination of the time-course of retrieval (McElree & Dosher, 1989; McElree, 2001, 2006). This model postulates that there are only two representational states: (1) passive representations and (2) a single item in the focus of attention. The latter item can be processed very quickly, whereas items outside of attention are retrieved in a time-consuming fashion with constant

retrieval speed. Similar to Oberauer's model, this view can also handle the present data by assuming that whenever people shift their focus of attention to a new frame (as they had to at all steps in our experiment), they automatically retrieve the corresponding letter (whether needed or not).

The difference between the models is that whereas Oberauer assumes that all 3 items are in the direct-access region (hence requiring "access" rather than "retrieval"), in McElree's model each updating step involves retrieval from long-term memory rather than access to an object in working memory. Because retrieval in McElree's model is from long-term memory, it follows that in order to accommodate our results, retrieval of object-location bindings from long-term memory must be automatic and obligatory. Although this stance has been taken—for instance, Logan (1988, 2002) has argued that attending to a specific object in space and retrieving information about it rely on the same basic process—we argue that there is ample evidence that *associative* retrieval from long-term memory is not automatic.

For example, Nobel and Shiffrin (2001) made a strong point that both associative recognition and cued recall differ from single-item recognition in that they involve a slow, non-automatic search process. Using a response deadline procedure, Gronlund and Ratcliff (1989) demonstrated that item information is usually available before associative information can be accessed, arguing for a different time course of retrieval for items and "bound" information, such as pairs of items. The neuropsychology literature has corroborated these assumptions, and suggests that automatic retrieval from long-term memory is limited to the recognition of single items (or highly unitized chunks). By contrast, associative retrieval of arbitrary bindings, such as object–location, word–temporal position, or face–name pairings, rest on slower retrieval processes that have been shown to be strategically controlled (Ecker, Zimmer, & Groh-Bordin, 2007; Herron & Rugg, 2003; Mecklinger, 2000; Meiser, Sattler, & Weißer, 2008). For instance, Ecker et al.

(2007) have shown that only intrinsic item features (such as an object's color or font) but not context features (such as background or location information) automatically affect object recognition, whereas contextual information is selectively retrieved via controlled recollection if relevant for the task. These controlled binding processes draw on distinct brain activity in the hippocampus and prefrontal cortex (Aggleton & Brown, 1999; Cansino, Maquet, Dolan, & Rugg, 2002; Rugg, Fletcher, Chua, & Dolan, 1999), and in line with this, Van Petten, Senkfor, and Newberg (2000) demonstrated that recognition of objects-in-locations—as compared to mere object recognition—elicits additional late prefrontal brain activity, interpreted as controlled retrieval and evaluation of source information. Moreover, damage to hippocampal or prefrontal brain areas leads to impairments in memory that requires controlled across-domain integration—including recall, source memory, and associative recognition—but typically spares single item recognition (Janowsky, Shimamura, & Squire, 1989; Kessels, Hendriks, Schouten, Asselen, & Postma, 2004; Mayes et al., 2004; Vargha-Khadem et al., 1997).

To conclude, we suggest that in order to explain our finding of obligatory retrieval of location-bound information at each updating step, the McElree model needs to make assumptions about associative long-term memory retrieval that are difficult to reconcile with research in that area. By contrast, the Oberauer model avoids these issues by means of the intermediate stage of direct access. We therefore consider the latter model to be more in line with our results than McElree's model: People can apparently hold and directly access 3 pieces of information in working memory. One prediction following from the Oberauer model but not the McElree model is that a substantial time-cost of **R** should emerge once retrieval from long-term memory is required. This could be achieved by working with memory sets far exceeding the capacity of the direct-access region, or by having people retrieve information that has already been removed from the direct-access region (Oberauer et al., 2005).

If three items are available in the direct-access region, why did our R conditions incur a cost in accuracy? If a focus shift entails obligatory retrieval of an item, whether needed or not, why is responding so much more accurate when the item is (re-)presented? The reason is simple: Memorized representations of the three letters are not always accurate—either because a previous updating step went wrong, because the incorrect letter is accessed (e.g., the frame-1 letter instead of the frame-2 letter), or because the object, which was previously outside the focus of attention, has been degraded by the time it is accessed. Recent research suggests that any degrading is most likely caused by interference rather than temporal decay (cf. Lewandowsky, Oberauer, & Brown, 2009). As a consequence, when an item is placed in the focus of attention by the obligatory process just discussed, it may simply be the wrong letter, or people might be unable to recall a letter at all. In the **R** conditions, this would lead to an error, whereas in the r_{no} conditions all required information is presented on the screen, so that failure to access the correct object in WM has no consequences for the accuracy of the response entered.

The Role of WMC for WMU

One main aim of this study was to examine the relationship between WMC and WMU component processes, using structural equation models. Not surprisingly, *WMC* was related to the latent factors that represented baseline performance in the WMU task (i.e., *GenRT* and *GenAcc*); these factors simply represent performance on a working memory task. Hence, people with a high WMC generally perform better and faster on a WMU task because performance on any WM task reflects variance due to sources that are common to WMC tasks and all conditions of our WMU task, such as memory maintenance and protection from interference, or generic processing speed. Of greater theoretical interest is the relation between WMC and the specific factors representing experimental effects; we discuss these in turn.

Transformation

Transformation had the strongest impact on WMU performance overall; conditions featuring a transformation were significantly slower and less accurate than conditions without transformation requirements. The size of this impact varied consistently with the size of the transformative operation, replicating the known effects of "problem size" with alpha-arithmetic operations in untrained subjects (e.g., Logan, 1988).

WMC and T covaried significantly in the accuracy model (but not the RT model). It could thus be concluded that WMC predicts the accuracy (but not the speed) of manipulating representations in working memory, and that it is this variance that primarily underpins the relation between WMC and any updating task that involves transformation.

In our view, transformation is a process that is not specifically related to WMU or WMC. Rather, transformation might primarily reflect general processing speed, a generic and well established individual-differences factor (cf. Chen & Li, 2007). This conclusion is supported by the significant correlation between T and $GenRT$ in the RT model. It is therefore likely that the $WMC\text{--}T$ relation in the accuracy model in part reflected individual differences in general processing efficiency. [12]

Retrieval

WMC strongly predicted the accuracy cost of retrieval, such that people with lower WMC were more likely to fail in retrieving the correct item. This finding echoes recent reports that WMC predicted the accuracy of retrieval in an updating task (Unsworth & Engle, 2008). These authors argued that low-span subjects have difficulties to correctly select items from outside the focus of attention during retrieval. An alternative interpretation is that people with low WMC suffer more interference between representations in WMC already during encoding, for instance through overlaying

distributed representations in a common weight matrix (Lewandowsky & Farrell, 2008) or feature overwriting (Oberauer & Kliegl, 2006).

Given that the retrieval component was associated with a zero mean time cost, the null covariation between R and WMC in the RT model is not unexpected. As argued above, retrieval of, or access to, a new memory item is likely to be involved in all experimental conditions, and therefore, individual differences in retrieval time or access time are captured in our generic $GenRT$ factor (among other processing aspects).

Substitution

Finally, substitution had a small but reliable impact on WMU performance. Moreover, individual differences in the S factor were not correlated with WMC. In the accuracy model the variance in S was too small for a covariance with WMC to become significant. For the RT model, however, the non-significant correlation is more informative because the variance of the S latent variable was significant (.03, $SE = .01$, $p < .001$). This could imply that—along the lines of a possible further decomposition of the substitution process as sketched earlier—both the encoding of a new letter into working memory and the implied removal of a previous letter from working memory may be unrelated to WMC.

This conclusion would be consistent with previous research. Using a visual short-term recognition task, Vogel, McCollough, and Machizawa (2005) found that individuals with high and low capacity on that task did not differ in their efficiency of encoding items but rather differed in the ability to prevent irrelevant input from being encoded. Likewise, Oberauer (2005) found that WMC was uncorrelated with the speed of removing a partial list from working memory in a verbal recognition task. We therefore conclude that unlike the other two components transformation and especially retrieval, substitution (as measured in our study) uniquely contributes variance to WMU

independent of WMC.

However, we must note that our measure of S may not reflect all sub-processes involved in substitution. S reflects the difference between an updating step that changes the memory content and one that does not. It is conceivable that in both conditions people actively replace their old representations by a new representation, except that in the no-substitution condition the new representation happens to have the same content as the old one. In that case, the active replacement process would not be reflected in S. Rather, S would reflect the cost of having to establish a new representation and overcoming the proactive interference from a different old representation, or in other words, the benefit from having an old representation that matches perfectly with the new one. We stress here that the mechanism generating proactive interference is not necessarily the same that disrupts retrieval. For instance, retrieval failures could be mostly due to gradual degradation of representations through feature overwriting. Proactive interference, in contrast, comes about because old features that should be removed from working memory are difficult to eliminate at will.

Despite this limitation, our finding that substitution is unique to WMU may shed some light on the ambivalent outcomes mentioned at the outset, which variously did or did not report an association between WMC and WMU. Specifically, if WMU tasks are utilized that draw heavily upon substitution (as opposed to retrieval and transformation) skills, then updating performance should be more likely to be independent of WMC than if the WMU tasks draw on retrieval and transformation.

For example, there is evidence that updating, compared to WMC, is disproportionately more impaired in schizophrenia (Galletly, MacFarlane, & Clark, 2007; Van Raalten et al., 2008). These results were based on tone-detection tasks that involved the counting of a specific tone or the detection of immediate tone repetition. Counting a single tone requires no retrieval because the current counter value is always in the focus of

attention, and transformation likewise contributes little variance because counting is highly overlearned. The main determinant of updating success therefore is likely to have been the substitution component. Thus, in light of our results, we suggest that substitution—as opposed to WMU more generally—is impaired in schizophrenia more than general WMC.

Relatedly, Oberauer (2005) found that removing items from WM was not reliably affected by age, whereas WMC in general decreases with age (e.g., Gilchrist, Cowan, & Naveh-Benjamin, 2008). Thus, one sub-process likely to be involved in substitution (viz., removal from memory) appears to be spared by healthy aging, whereas general WMC is not. Both results are consonant with the idea that substitution (which can be further differentiated into removal followed by encoding, as noted earlier) is independent of other measures of WMC.

Conversely, many previous studies that reported correlations between updating and other indicators of cognitive performance (Chen & Li, 2007; Friedman et al., 2006; Carretti et al., 2005; Palladino et al., 2001) utilized updating measures that drew heavily upon retrieval. In fact, in those four studies, WMU was measured almost exclusively with running-memory and keep-track tasks. The former requires retrieval of the last x items from an ongoing list with unknown list length and the latter requires the retrieval of the most recently presented exemplars of a finite number of categories (see also Table 1). Although these tasks involve substitutions, they crucially depend on retrieval. It follows that their correlation with WMC is not unexpected and likely due to the retrieval demands entailed by those tasks. It also follows that the power of WMU tasks to predict higher cognitive function (e.g., fluid intelligence) may rest partially or completely on the involvement of regular WM processes such as retrieval in the updating task.

Concluding Remarks on the WMC–WMU Relation

To conclude, after we decomposed performance in a working memory updating task, we identified substitution as the only process that unequivocally contributes variance that is unique to WMU. Retrieval is part and parcel of every WMC measure, and consequently our retrieval factor correlated with the WMC factor. The correlation of WMC and transformation skills was more ambiguous and may reflect individual differences in generic processing efficiency, although transformation may also play a more specific role in measures of WMC that require simultaneous maintenance and processing (Oberauer, 2002). Thus, with the exception of substitution, performance on updating tasks can largely be explained on the basis of component processes—transformation and especially retrieval—that are known attributes of WMC and/or processing speed (cf. Schmiedek et al., in press).

It follows that if researchers want to establish updating of working memory as a construct that reflects moderately general and stable individual differences in an executive process that is distinguishable from other WM and speed-related processes (i.e., retrieval, transformation, and perhaps some processes involved in WM maintenance such as refreshing, distractor inhibition, etc.), they will have to focus on the component of variance reflected in our substitution factor, and control for all other components. Future research could look at more heterogeneous samples—where there could be more interindividual variance in substitution performance—and should also determine the potential power of substitution performance to predict higher cognitive functions (e.g., fluid intelligence) in itself.

Conclusions

In summary, we reported the first study that systematically decomposed processes in a typical working memory updating task into constituent processes. Our results suggest

that the processes of retrieval, transformation, and substitution make distinct and independent contributions to performance on updating tasks.

We further found a selective relationship between working memory capacity and processes involved in working memory updating tasks, whereby only the process of substitution seems to be unique to paradigms for investigating updating. This potentially reconciles inconsistencies in the literature, as some studies may have conflated updating measures with more capacity-related factors.

References

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the
hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences*, *22*, 425-444.

Baddeley, A. (1986). *Working memory.* New York: Oxford University Press.

Bates, D. B. (2007). lme4: Linear mixed-effects models using s4 classes [Computer
software manual]. (R package version 0.99875-9)

Bialkova, S., & Oberauer, K. (in press). Direct access to working memory contents.
Experimental Psychology.

Blunch, N. J. (2008). *Introduction to structural equation modelling using SPSS and
AMOS.* London: Sage.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Buehner, M., Krumm, S., & Pick, M. (2005). Reasoning=working memory≠attention.
Intelligence, *33*, 251-272.

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

Cansino, S., Maquet, P., Dolan, R. J., & Rugg, M. D. (2002). Brain activity underlying
encoding and retrieval of source memory. *Cerebral Cortex*, *12*, 1048-1056.

Carretti, B., Cornoldi, C., De Beni, R., & Romano, M. (2005). Updating in work memory:
A comparison of good and poor comprehenders. *Journal of Experimental Child
Psychology*, *91*, 45-66.

Chen, T., & Li, D. (2007). The roles of working memory updating and processing speed in
mediating age-related differences in fluid intelligence. *Aging, Neuropsychology, and
Cognition*, *14*, 631-646.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles
assigned along the garden path linger. *Cognitive Psychology*, *42*, 368-407.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle,

R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769-786.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, *19*, 450-466.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 357-374.

Ecker, U. K. H., Zimmer, H. D., & Groh-Bordin, C. (2007). Colour and context: An ERP study on intrinsic and extrinsic feature binding in episodic memory. *Memory & Cognition*, *35*, 1483-1501.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (p. 102-134). London: Cambridge University Press.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*, 172-179.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*, 237-241.

Fürst, A. J., & Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory & Cognition*, *28*, 774-782.

Galletly, C. A., MacFarlane, A. S., & Clark, C. R. (2007). Impaired updating of working memory in schizophrenia. *International Journal of Psychophysiology*, *63*, 265-274.

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition*, *26*, 263-276.

Geiger, S. M., & Lewandowsky, S. (2008). Temporal isolation does not facilitate forward serial recall—or does it? *Memory & Cognition*, *36*, 957–967.

Gilchrist, A. L., Cowan, N., & Naveh-Benjamin, M. (2008). Working memory capacity for spoken sentences decreases with adult ageing: Recall of fewer but not smaller chunks in older adults. *Memory*, *16*, 773-787.

Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 846-858.

Gugerty, L. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied*, *3*, 42-66.

Herron, J. E., & Rugg, M. D. (2003). Strategic influences on recollection in the exclusion task: Electrophysiological evidence. *Psychonomic Bulletin & Review*, *10*, 703-710.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia*, *27*, 1043-1056.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122-149.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189-217.

Kessels, R. P. C., Hendriks, M. P. H., Schouten, J., Asselen, M. van, & Postma, A. (2004). Spatial memory deficits in patients after unilateral selective amygdalohippo-campectomy. *Journal of the International Neuropsychological*

Society, *10*, 907-912.

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

Kirchner, W. K. (1958). Age differences in short term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*, 352-358.

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

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. *The Psychology of Learning and Motivation*, *49*, 1-48.

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences*, *13*, 120-126.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492-527.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*, 376-400.

Mayes, A. R., Holdstock, J. S., Isaac, C. L., Montaldi, D., Grigor, J., Gummer, A., et al. (2004). Associative recognition in a patient with selective hippocampal lesions and relatively normal item recognition. *Hippocampus*, *14*, 763-784.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287-330.

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

McElree, B. (2006). Accessing recent events. *The Psychology of Learning and Motivation*, *46*, 155-200.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory:

The time course of recognition. *Journal of Experimental Psychology: General*, *118*,

346-373.

Mecklinger, A. (2000). Interfacing mind and brain: A neurocognitive model of recognition

memory. *Psychophysiology*, *37*, 565-582.

Meiser, T., Sattler, C., & Weißer, K. (2008). Binding of multidimensional context

information as a distinctive characteristic of remember judgments. *Journal of

Experimental Psychology: Learning Memory and Cognition*, *34*, 32-49.

Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounois, J. (1988). The dynamics of

cognition and action: Mental processes inferred from speed-accuracy decomposition.

Psychological Review, *95*, 183-237.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager,

T. D. (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.

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the

central executive. *British Journal of Psychology*, *81*, 111-121.

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessability and situation

models in narrative comprehension. *Journal of Memory and Language*, *26*, 165-187.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further

evidence for a cascade model of lexical access. *Journal of Memory and Language*,

53, 359-377.

Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall.

Journal of Experimental Psychology: Learning, Memory, and Cognition, *27*,

384-413.

Oberauer, K. (1993). Die Koordination kognitiver Operationen - eine Studie über die

Beziehung zwischen Intelligenz und "working memory" (The coordination of cognitive operations - a study on the relation of intelligence and "working memory"). *Zeitschrift für Psychologie, 201*, 57-84.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 411-421.

Oberauer, K. (2003). Selective attention to elements in working memory. *Experimental Psychology, 50*, 257-269.

Oberauer, K. (2005). Control of the contents of working memory—a comparison of two paradigms and two age groups. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 714-728.

Oberauer, K. (2006). Is the focus of attention in working memory expanded through practice? *Journal of Experimental Psychology: Learning Memory and Cognition, 32*, 197-214.

Oberauer, K., & Kliegl, R. (2001). Beyond resources: Formal models of complexity effects and age differences in working memory. *European Journal of Cognitive Psychology, 13*, 187-215.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55*, 601–626.

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and Individual Differences, 29*, 1017-1045.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31*, 167-193.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2008). Which working

memory functions predict intelligence? *Intelligence*, *36*, 641-652.

Oberauer, K., & Vockenberg, K. (2009). Updating of working memory: Lingering
bindings. *Quarterly Journal of Experimental Psychology*, *62*, 967-987.

Oberauer, K., Wilhelm, O., & Schmiedek, F. (2005). Experimental strategies in
multivariate research. In A. Beauducel, B. Biehl, M. Bosniak, W. Conrad,
G. Schönberger, & D. Wagener (Eds.), *Multivariate research strategies. Festschrift in
honor of WernerW. Wittmann* (p. 119-149). Aachen, Germany: Shaker.

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and
updating processes in reading comprehension. *Memory & Cognition*, *29*, 344-354.

Pelli, D. G. (1997). The video toolbox software for visual psychophysics: Transforming
numbers into movies. *Spatial Vision*, *10*, 437-442.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus*. New York,
NY: Springer.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2008). coda: Output analysis and
diagnostics for mcmc [Computer software manual]. (R package version 0.13-2)

Postle, B. R. (2003). Context in verbal short-term memory. *Memory & Cognition*, *31*,
1198-1207.

R Development Core Team. (2008). R: A language and environment for statistical
computing [Computer software manual]. Vienna, Austria. Available from
`http://www.R-project.org` (ISBN 3-900051-07-0)

Radvansky, G. A., & Copeland, D. E. (2001). Working memory and situation model
updating. *Memory & Cognition*, *29*, 1073-1080.

Radvansky, G. A., & Dijkstra, K. (2007). Aging and situation model processing.
Psychonomic Bulletin & Review, *14*, 1027-1042.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59-108.

Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning

using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 353-368.

Rugg, M. D., Fletcher, P. C., Chua, P. M. L., & Dolan, R. J. (1999). The role of the prefrontal cortex in recognition memory and memory for source: An fMRI study. *NeuroImage*, *10*, 520-529.

Salthouse, T. A., Babcock, R. L., & Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working memory. *Psychology and Aging*, *6*, 118-127.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (in press). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Townsend, J. T., & Fific, M. (2004). Parallel versus serial processing and individual differences in high-speed search in human memory. *Perception and Psychophysics*, *66*, 953-962.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *49*, 446-468.

Unsworth, N., & Engle, R. W. (2008). Speed and accuracy of accessing information in working memory: An individual differences investigation of focus switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 616-630.

Van Petten, C., Senkfor, A. J., & Newberg, W. M. (2000). Memory for drawings in locations: Spatial source memory and event-related potentials. *Psychophysiology*, *37*, 551-564.

Van Raalten, T. R., Ramsey, N. F., Jansma, J. M., Jager, G., & Kahn, R. S. (2008). Automatization and working memory capacity in schizophrenia. *Schizophrenia Research*, *100*, 161-171.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic

and semantic memory. *Science*, *277*, 376-380.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500-503.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors*, *5*, 7-17.

Yntema, D. B., & Mueser, G. E. (1962). Keeping track of variables that have few or many states. *Journal of Experimental Psychology*, *63*, 391-395.

Zbrodoff, N. J. (1999). Effects of counting in alphabet arithmetic: Opportunistic stopping and priming of intermediate steps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 299-317.

Author Note

Notes

[1] In the following, whenever we refer to the "component processes of WMU", it would be formally more correct to use the term "component processes potentially required for performance in WMU tasks", as not all updating necessarily involves all three sub-processes described in the following paragraph.

[2] We use the term "retrieval" in a theoretically neutral manner, to refer to whatever cognitive processes are required to use information that is not presented. In some models, the actual process may be more accurately described as "access" to already-retrieved information; we will take up those theoretical nuances after all the data have been presented.

[3] The WMC task battery in fact also included a fourth task—the original version of the Memory Updating task as used by Oberauer (e.g., Oberauer et al., 2000; Oberauer, 2002). However, as this task served as a blueprint for our experimental task (explained below), we do not consider it further; including it into the estimation of WMC does not substantially alter our conclusions.

[4] Because our focus was exclusively on updating, we do not report the final recall data. The recall test was included only to ensure that people remembered all three items until after the last updating step.

[5] We stress that we define substitution as the replacement of memorized information (which is bound to a frame location) by the result of an updating step in that frame. One could argue that our r_{no}-**T** conditions (additionally) involve a second type of substitution, namely replacing the memory content with the operand of the transformation before actually beginning the transformation. We do not agree with this for two reasons. First, in that case our r_{no}-**T**-**S** condition would involve another substitution process and should hence be slower than our **R**-**T**-**S** condition—which it was not (see below). Second, we conducted an additional experiment with 25 subjects that compared two conditions, both

of which required transformation and substitution but no retrieval. One condition exactly replicated the r_{no}-**T-S** condition of the present experiment (i.e., the operand of the transformation differed from the current memory content, so when remembering "C", the transformation was either $A + 1$ or $B + 2$), the other used the current memory content as the transformation operand (i.e., when remembering "C", the transformation was either $C + 1$ or $C + 2$). These two conditions differed neither in RT (difference 13 ms) nor accuracy (difference .01 accuracy units). Hence, if there was no retrieval involved, it did not matter whether subjects used the letter they kept in memory or a newly presented letter when performing the transformation.

[6] Technically, for the accuracy model, we used an additive regression model but first log-transformed the accuracy data so that multiplicative factors were transformed into additive factors. After the regression, the estimated weights were then re-transformed via an exponential transformation to facilitate interpretation. It is those re-transformed estimates that are presented in the regression model.

[7] Coding each type of transformation with such an incremental combination of variables has the advantage that the significance of each variable's effect over and above the others can be assessed — e.g., whether or not a +2 transformation significantly reduces accuracy over and above a +1 transformation — although a discrete estimation of each transformation type would also be possible and equally appropriate.

[8] Interactions involving T0 and T2 could not be assessed because they were not fully crossed with S and R, respectively. We regard T1 as the baseline that represents the effect of a typical transformative operation (incrementing the letter by one), whereas T0 and T2 only reflect savings or costs, respectively, involved in specific deviations from that baseline; therefore we regard interactions involving T0 and T2 as less important. We also could not assess the 3-way interaction of R, S, and T1 because that interaction is redundant with T0 (i.e., the 3-way interaction term captures a deviation of a single cell

from the pattern expected by the full model without the 3-way, and T0 captures exactly the same), thereby making the model unidentifiable.

[9] We confirmed this pattern by a "reverse" path model that predicted WMC from the WMU components. We calculated this with paths (instead of correlations) going from R, T, S, and $GenAcc$ to WMC. We found that R accounted for most of the variance in WMC, T accounted only for an insignificant amount, and S accounted for none (the weight from $GenAcc$ to WMC was also nonsignificant; variance of the disturbance variable on WMC was significant, implying that there was unexplained variance left in WMC). This corroborated our finding that retrieval but not substitution is related to WMC, while the relationship between transformation and WMC seems to straddle the statistical detection boundary.

[10] If this parameter was freely estimated, the estimate was .514.

[11] Specifically, the accuracy model's fit could be further improved by adding a covariance link between error terms $e1$ and $e8$, $\chi^2(44) = 59.94$; $CFI = .97$; $RMSEA = .061$; $SRMR = .052$; the RT model's fit could be improved by linking $e2$–$e6$, $e3$–$e4$, $e3$–$e8$, $e4$–$e8$, and $e6$–$e8$, $\chi^2(44) = 64.04$; $RMSEA = .069$, $CFI = .97$, $SRMR = .082$.

[12] The assumption that the variance in T is general variance is consistent with the nonsignificant WMC–T relation in the supplementary directional path model, because in this model the different WMU components compete with each other as WMC predictors, and apparently T and $GenACcc$ share the same variance with WMC.

Table 1

Processes Involved in Common Working Memory Updating Tasks

Task	Source	Retrieval	Transformation	Substitution
n-back	Kirchner (1958) McElree (2001)	+	−	+
Keep-track	Yntema (1963) Miyake et al. (2000)	+	−	+
Running memory	Morris & Jones (1990) Friedman et al. (2006)	+	−	+
MU	Salthouse et al. (1991) Oberauer et al. (2000)	+	+	+
Tone-repetition detection	Galletly et al. (2007)	−	−	+
Deviant-tone counting	Van Raalten et al. (2008)	−	+	+
Matrix updating	Chen & Li (2007)	+	−	+

Legend. MU, Memory Updating task; −, no involvement; +, involvement.

Note. In the keep-track task, subjects are given a number of categories (e.g., tools, animals) and are then presented with a series of exemplar words (e.g., hammer, dog) and the task is to always remember the most recently presented exemplar for each category. In the running memory task, subjects are presented with an item list of unpredictable length and have to remember the last n items of the list. The tone-repetition detection is a simple task in which subjects monitor a series of tones and respond whenever a target tone is immediately repeated. The deviant-tone counting task involves keeping a running count of rarely occurring tones in a sequence. The matrix updating task requires subjects to mentally move memorized dots across the cells of a matrix. The *n*-back and MU tasks are explained in the text. Task labels are not necessarily those used in the cited studies.

Table 2

Conditions Used in the Working Memory Updating Session, with Examples of Updating Prompts

	T		t_{no}	
	R	r_{no}	**R**	r_{no}
S	?+1 or ?+2	A+1 or B+2	$\rightarrow$	X
s_{no}	?+0	A+2 or B+1	?	C

Legend. **R**, Retrieval involved; **T**, Transformation involved; **S**, Substitution involved; r_{no}, Retrieval not involved; t_{no}, Transformation not involved; s_{no}, Substitution not involved.

Note. Examples of prompts assume that the letter "C" is currently memorized in the frame being updated. See text for details.

Table 3

Descriptive Statistics of the Working Memory Tasks

Task	Mean	Range	Skewness	Kurtosis
OS	.69 (.013)	.23 − .94	-1.10	2.03
OS_{pt}	.91 (.006)	.71 − 1.00	-.87	.45
SS	.63 (.018)	.25 − .96	-.22	-.78
SS_{pt}	.79 (.009)	.47 − .92	-1.22	1.81
SSTM	.85 (.004)	.73 − .94	-.24	-.24

Legend. OS, Operation Span; SS, Sentence Span; $_{pt}$ denotes processing tasks; SSTM, Spatial Short-Term Memory.

Note. Standard errors in parentheses.

Table 4

Working Memory Tasks Intercorrelations

Task	OS	OS$_{pt}$	SS	SS$_{pt}$	SSTM
OS	1.00				
OS$_{pt}$	.41	1.00			
SS	.50	.31	1.00		
SS$_{pt}$	.12	.36	.21	1.00	
SSTM	.30	.37	.27	.09	1.00

Legend. OS, Operation Span; SS, Sentence Span; $_{pt}$ denotes processing tasks; SSTM, Spatial Short-Term Memory.

Table 5

Accuracy of Updating (in Upper Rows) and Reaction Times (RTs in ms, in Lower Rows) in the Working Memory Updating Task

	T		t_{no}	
	R	r_{no}	**R**	r_{no}
S	.76 (.016)	.87 (.011)	.82 (.014)	.98 (.005)
	3115 (63)	3009 (69)	2141 (50)	1642 (37)
s_{no}	.85 (.013)	.89 (.011)	.87 (.012)	.99 (.005)
	1599 (35)	2803 (69)	1297 (24)	1272 (29)

Legend. **R**, Retrieval involved; **T**, Transformation involved; **S**, Substitution involved; r_{no}, Retrieval not involved; t_{no}, Transformation not involved; s_{no}, Substitution not involved. See text for details.

Note. Standard errors in parentheses.

Table 6

Standardized Weights of WMU Measurement Models

	Accuracy							
Factor	**R-T-S**	r$_{no}$-**T-S**	**R**-t$_{no}$-**S**	r$_{no}$-t$_{no}$-**S**	**R-T**-s$_{no}$	r$_{no}$-**T**-s$_{no}$	**R**-t$_{no}$-s$_{no}$	r$_{no}$-t$_{no}$-s$_{no}$
GenAcc	.21	.36	.25	.82	.29	.35	.30	.88
R	.49		.58		.68		.70	
T	.30	.51			.04	.49		
S	0	0	0	0				
	RT							
Factor	**R-T-S**	r$_{no}$-**T-S**	**R**-t$_{no}$-**S**	r$_{no}$-t$_{no}$-**S**	**R-T**-s$_{no}$	r$_{no}$-**T**-s$_{no}$	**R**-t$_{no}$-s$_{no}$	r$_{no}$-t$_{no}$-s$_{no}$
GenRT	.36	.34	.49	.62	.65	.33	.90	.81
R	.13		.17		.23		.32	
T	.66	.61			.24	.60		
S	.28	.25	.37	.47				

Legend. GenAcc, General Accuracy (latent factor); R, Retrieval (latent factor); T, Transformation (latent factor); S, Substitution (latent factor); GenRT, General RT (latent factor); R-T-S codes, refer to log-accuracy measures (upper panel) or RT measures (lower panel) of respective experimental conditions (manifest variables).

Note. The only unstandardized weights that were not fixed were the accuracy and RT weights linking T and the **R-T**-s$_{no}$ manifest variable. These were estimated as .10 (accuracy) and .20 (RT), respectively. See text for further details.

Table 7

Standardized and Unstandardized Weights of WMC Measurement Models

Factor	Model	OS	SS	SSTM
Accuracy				
WMC	Standardized	.72	.72	.36
	Unstandardized	.09	.12	.02
RT				
WMC	Standardized	.73	.69	.40
	Unstandardized	.09	.12	.02

Legend. WMC, Working Memory Capacity (latent factor); OS, Operation Span (manifest variable); SS, Sentence Span (manifest variable); SSTM, Spatial Short-Term Memory (manifest variable). See text for further details.

Table 8

Unstandardized Covariances of Latent Factors in Accuracy and RT Structural Equation Models

Factors	Estimate	CI	p
Accuracy			
WMC-GenAcc	.012 (.006)	.-001 − .041	.051
WMC-R	.062 (.014)	.034 − .095	$< .001$
WMC-T	.033 (.012)	.010 − .060	.005
GenAcc-R	.003 (.001)	.000 − .009	$< .001$
R-T	.006 (.001)	.003 − .011	$< .001$
e2-e6	.006 (.002)	.001 − .016	$< .001$
e6-e7	.003 (.001)	.001 − .006	$< .001$
RT			
WMC-GenRT	-.05 (.03)	-.107 − .008	.077
GenRT-T	.04 (.01)	.017 − .069	$<.001$
e2-e6	.07 (.02)	.037 − .105	.002
e4-e8	.01 (.01)	.002 − .025	.031

Legend. WMC, Working memory capacity; GenAcc, General Accuracy; R, Retrieval; T, Transformation; GenRT, General RT; e2, e4, e6, e7, and e8, error terms of Conditions r_{no}-**T**-**S**, r_{no}-t_{no}-**S**, r_{no}-**T**-s_{no}, **R**-t_{no}-s_{no}, and r_{no}-t_{no}-s_{no} performance variables; CI, 95% bias-corrected confidence interval (derived from 1000 bootstrap samples). See text for further details.

Note. In the RT model, the marginally significant WMC–GenRT covariance was significant with $p < .05$ before removing the nonsignificant WMC–T link and was thus retained. Standard errors in parentheses.
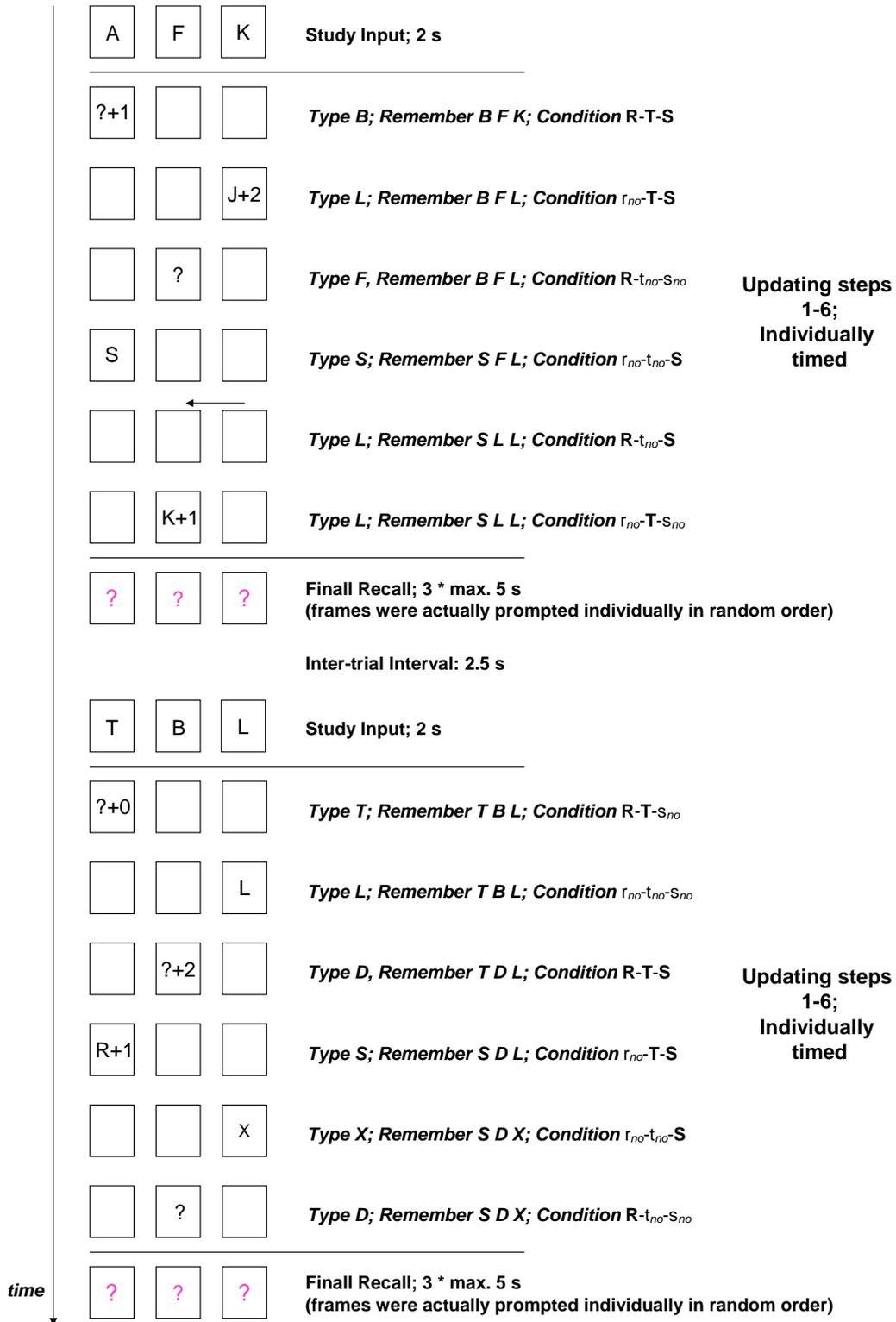
Figure Captions

Figure 1. Sample trial sequence for the WM updating task.

Figure 2. Graphical representation of the structural equation model for accuracy, showing the prediction of latent updating factors GenAcc (general accuracy), R (retrieval), T (transformation), and S (substitution) by a latent WMC factor. Manifest accuracy variables reflect log-transformed accuracy data referring to the eight respective experimental conditions. WMC-related manifest variables reflect mean performance in WMC tasks OS (operation span), SS (sentence span), and SSTM (spatial short-term memory). e1-11 = error variables. Estimated standardized weights (correlations, in boldface) are presented adjacent to latent connections. Estimated unstandardized means (in log-accuracy units, italicized) are shown inside the latent factors. Means of latent factors that are not given in the figure (error variables and WMC factor) were fixed at zero. Regression weights in the WMU measurement model were fixed at 1, with the exception of the link between T and the **R**-**T**-s$_{no}$ variable, which was freely estimated (dashed arrow with unstandardized estimate in italics). All estimated covariances provided in the figure are (marginally) significant, $p < .051$ (see upper panel of Table 8); all estimated means are significantly different from zero, $p < .001$. See text for further details.

Figure 3. Graphical representation of the structural equation model for RT, showing the prediction of latent updating factors GenRT (general RT), R (retrieval), T (transformation), and S (substitution) by a latent WMC factor. Manifest RT variables reflect RT data referring to the eight respective experimental conditions. WMC-related manifest variables reflect mean performance in WMC tasks OS (operation span), SS (sentence span), and SSTM (spatial short-term memory). e1-11 = error variables. Estimated standardized weights (correlations, in boldface) are presented adjacent to latent

connections. Estimated unstandardized means (in seconds, italicized) are shown inside the latent factors. Means of latent factors that are not given in the figure (error variables and WMC factor) were fixed at zero. Regression weights in the WMU measurement model were fixed at 1, with the exception of the link between T and the $\mathbf{R}$-$\mathbf{T}$-s$_{no}$ variable, which was freely estimated (dashed arrow with unstandardized estimate in italics). Error term e3 was fixed to an independently derived estimate. All estimated covariances provided in the figure are at least marginally significant, $p < .08$ (see lower panel of Table 8); all estimated means are significantly different from zero, $p < .001$, except the estimated mean of R, $p = .16$. See text for further details.
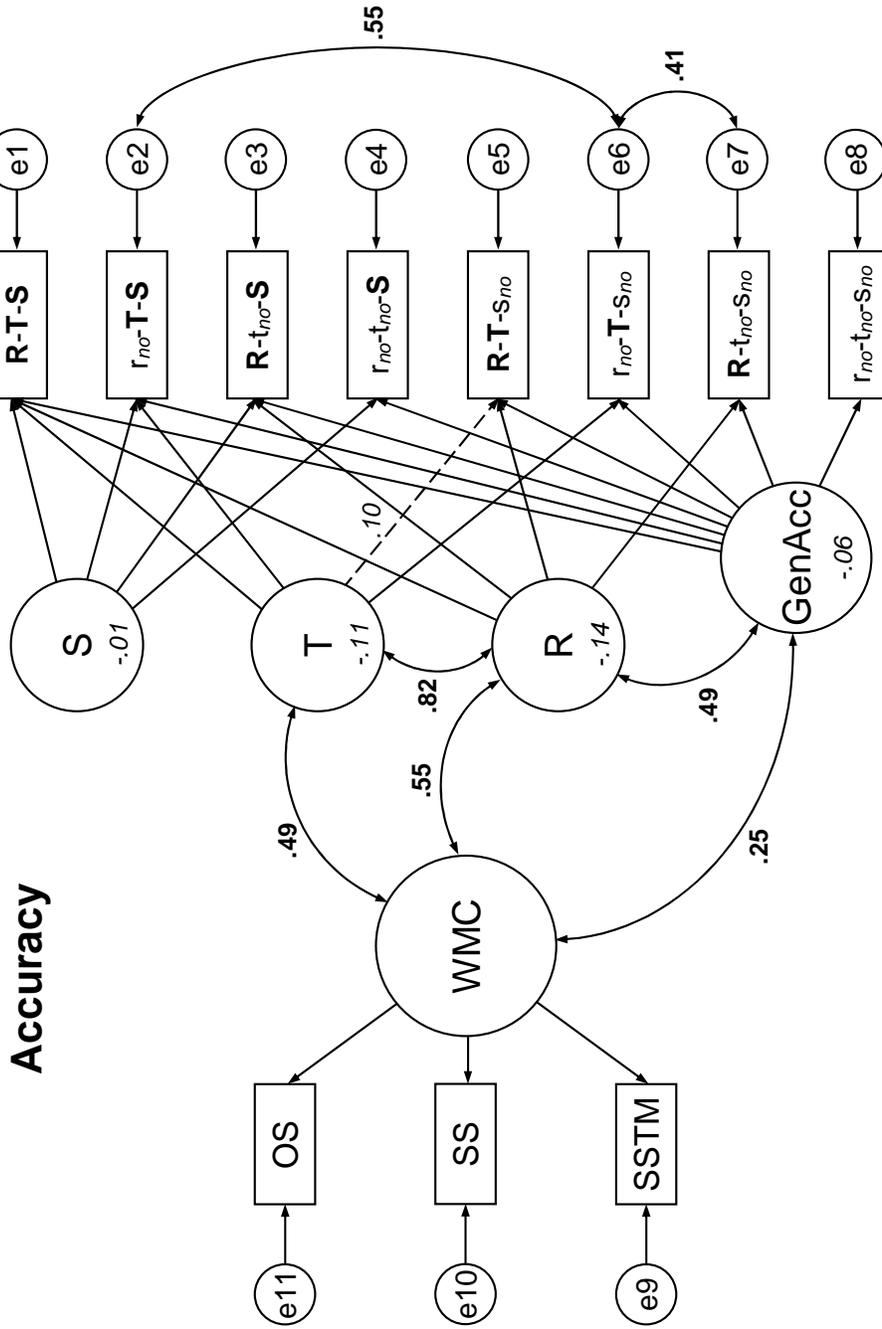
A	F	K	**Study Input; 2 s**

?+1			***Type B; Remember B F K; Condition*** R-T-S

		J+2	***Type L; Remember B F L; Condition*** r_{no}-T-S

	?		***Type F, Remember B F L; Condition*** R-t_{no}-s_{no}

**Updating steps
1-6;
Individually
timed**

S			***Type S; Remember S F L; Condition*** r_{no}-t_{no}-S

			Type L; Remember S L L; Condition R-t_{no}-S

	K+1		***Type L; Remember S L L; Condition*** r_{no}-T-s_{no}

?	?	?	**Finall Recall; 3 * max. 5 s** **(frames were actually prompted individually in random order)**

Inter-trial Interval: 2.5 s

T	B	L	**Study Input; 2 s**

?+0			***Type T; Remember T B L; Condition*** R-T-s_{no}

		L	***Type L; Remember T B L; Condition*** r_{no}-t_{no}-s_{no}

	?+2		***Type D, Remember T D L; Condition*** R-T-S

**Updating steps
1-6;
Individually
timed**

R+1			***Type S; Remember S D L; Condition*** r_{no}-T-S

		X	***Type X; Remember S D X; Condition*** r_{no}-t_{no}-S

	?		***Type D; Remember S D X; Condition*** R-t_{no}-s_{no}

time

?	?	?	**Finall Recall; 3 * max. 5 s** **(frames were actually prompted individually in random order)**

Accuracy

RT