Practice-related Changes in Perceptual Evidence Accumulation Correlate with Changes in Working Memory

Florian Schmiedek

Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

Martin Lövdén

Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

Department of Psychology, University of Gothenburg, Gothenburg, Sweden.

Roger Ratcliff

Psychology Department, The Ohio State University, Columbus, Ohio, USA.

Ulman Lindenberger

Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany, and

London, UK

Author Note

Correspondence concerning this article should be addressed to Florian Schmiedek, DIPF | Leibniz Institute for Research and Information in Education, Rostocker Str. 6, 60323 Frankfurt am Main, Germany.

E-mail: schmiedek@dipf.de

Repository link: https://osf.io/5qx7e/

Abstract

It has been proposed that evidence accumulation determines not only the speed and accuracy of simple perceptual decisions but also influences performance on tasks assessing higher-order cognitive abilities, such as working memory (WM). Accordingly, estimates of evidence accumulation based on diffusion decision modeling of perceptual decision-making tasks have been found to correlate with WM performance. Here we use diffusion decision modeling in combination with latent factor modeling to test the stronger prediction that practice-induced changes in evidence accumulation correlate with changes in WM performance. Analyses are based on data from the COGITO Study, in which 101 young adults practiced a battery of cognitive tasks, including three simple two-choice reaction-time tasks and three WM tasks, in 100 day-to-day training sessions distributed over six months. In initial analyses, drift rates were found to correlate across the three choice tasks, such that latent factors of evidence accumulation could be established. These latent factors of evidence accumulation were positively correlated with latent factors of practiced and unpracticed WM tasks, both before and after practice. As predicted, individual differences in changes of evidence accumulation correlated positively with changes in WM performance. Our findings support the proposition that decision making and WM both rely on the active maintenance of task-relevant internal representations.

Keywords: working memory, decision making, evidence accumulation, diffusion decision modeling, structural equation modeling, cognitive training.

Practice-related Changes in Perceptual Evidence Accumulation Correlate with Changes in Working

Memory

In an influential paper, Schmiedek, Oberauer, Wilhelm, Süß, and Wittmann (2007) demonstrated strong latent-factor correlations between working memory (WM) and simple two-choice decision tasks. Based on a diffusion model account (Ratcliff, 1978), they demonstrated that individual differences in the parameter capturing the efficiency of evidence accumulation in the decision tasks (i.e., the drift rate) could largely account for the specific relation of WM performance and individual differences in the skewness (i.e., the tau parameter of an ex-Gaussian analysis) of the decision tasks' response time (RT) distributions. As a possible explanation of the WM-drift rate relation, Schmiedek and colleagues (2007) argued that performance on the decision tasks may depend on the *active maintenance of procedural information*, that is, on establishing, maintaining, and updating stimulus-response bindings that are arbitrary in nature and vary from task to task – a function common to, and potentially supported by, the WM system. Here, we make use of an intensive training study, which allowed us to not only try to conceptually replicate the earlier findings of strong cross-sectional WM-drift rate correlations, but also to investigate how practice-related changes in WM performance and drift rates of decision tasks correlate with each other and how the cross-sectional WM-drift rate correlation changes due to the training intervention. In combination with the choice of a masking paradigm used for the perceptual decision tasks, this allowed us to investigate an alternative potential account of the involvement of WM in simple perceptual decision tasks. In this alternative account, the basic function of WM – to provide for flexible mental representations based on arbitrary bindings of pieces of information – is involved in the *active maintenance of declarative information* necessary for the decision process in cases where this information is not reliably and continuously provided by the environment.

In the following, we review the literature on relations of simple decision speed and higher-order cognitive abilities in general, on relations of WM and drift rates in specific, and on practice-

related changes in WM and drift rates, which we use as our main methodological approach to further explain the commonalities of WM and simple decision making. Results of our intensive training study indicate that drift rates correlate highly with WM performance at, both, pre- and posttest, with correlations tending to increase, and changes in WM being positively correlated with changes in drift rates. Together, these results provide support for the proposition that WM is involved in the active maintenance of declarative information when decision tasks are practiced intensively and stimulus information is masked.

The Worst Performance Rule: Attentional Control versus Evidence Accumulation Accounts

Investigations of the relation of simple decision speed (i.e., response times on simple perceptual two-choice decision tasks) and higher-order cognitive abilities (e.g., fluid intelligence) have a long tradition in cognitive and differential psychology (Deary, 2000; Jensen, 2006). In this area of research, a robust empirical phenomenon, coined the *worst performance rule* (Baumeister & Kellas, 1968; Larson & Alderton, 1990), has emerged. This rule describes the finding that correlations of individual differences in intelligence are stronger for the slower quantiles of RT distributions of simple decision tasks (see Coyle, 2003, for a review, and Schubert, 2019, for a meta-analysis). Two major theoretical accounts of the worst performance rule have been proposed.

The *attentional control account* (Larson & Alderton, 1990; McVay & Kane, 2012) suggests that people with lower levels of intelligence have greater difficulties in maintaining attentional control over task performance and therefore a higher likelihood of suffering lapses of attention. As such lapses of attention produce particularly long RTs, the slower quantiles of the RT distribution more strongly reflect such limitations in attentional control, and thereby correlate more highly with psychometric measures of intelligence than the faster quantiles.

Alternatively, the *evidence accumulation account* (Ratcliff, Schmiedek, & McKoon, 2008; Schmiedek et al., 2007), which we focus on in the present work, is based on the diffusion decision model (Ratcliff, 1978; Ratcliff & McKoon, 2008), a theoretical process model for two-choice

decision tasks. According to the diffusion model, variability in decision making reflects an inherently noisy process of evidence accumulation. The efficiency of evidence accumulation in the decision process is captured by the drift rate. If individual differences in drift rate contribute to individual differences in intelligence, then slower quantiles of decision making are expected to correlate more highly with intelligence (Ratcliff et al., 2008).

The evidence accumulation account of the link between perceptual decision making and intelligence is both more fundamental and more parsimonious than the attentional control account, as no qualitatively different states (i.e., controlled processing vs. attentional lapses) need to be assumed. Rather, occasional long RTs (which might subjectively very well be perceived as lapses) naturally result from a set of mechanisms that govern the full range of observed behavior and give rise to the right-skewed distribution that is typical for simple choice RT tasks.

The Role of Working Memory in the Evidence-Accumulation Account

Using a simplified version of the diffusion model (i.e., the EZ diffusion model, Ratcliff, 2008; Wagenmakers, van der Maas, & Grasman, 2007), Schmiedek and colleagues (2007) have shown empirically that a latent factor of drift rates, estimated for eight different two-choice decision tasks, correlated highly with a latent factor of fluid intelligence from a psychometric paper-and-pencil intelligence test. A similar pattern of (latent factor) correlations of drift rate with WAIS IQ measures has been reported by Ratcliff, Thapar, and McKoon (2010; 2011) and by Schmitz and Wilhelm (2016).

Importantly, Schmiedek et al. (2007) also found a high correlation ($r = .68$) of the drift rate factor with a latent factor of WM tasks. Given that associations between WM and fluid intelligence are known to be strong (Kane, Hambrick, & Conway, 2005; Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009; Wilhelm, Hildebrandt, & Oberauer, 2013; see also Schmiedek, Lövdén, & Lindenberger, 2014, who show a similar correlation using data from the same participants as those in the present investigation), these findings suggest that mechanisms

contributing to individual differences in simple decision making tasks might also contribute to individual differences in WM and fluid intelligence. Regarding the distinguishability of the attentional control and the evidence accumulation account, Schmiedek et al. (2007) also conducted a simulation study, which did show that the relation of individual differences in the skewness of the two-choice task RT distributions (represented by the ex-Gaussian parameter tau) and WM could almost fully be explained by individual differences in drift rate. Given this empirical finding and the theoretical parsimony of the evidence accumulation account, we do not attempt to further distinguish the two accounts in present work, but rather take the evidence accumulation account as our theoretical basis. This does not ultimately exclude the possibility that lapses of attention are an additional shared mechanism that contributes to the relation of perceptual decision tasks and WM (see, e.g., the recent work by Löffler, Frischkorn, Rummel, Hagemann, & Schubert, 2022, on the contribution of self-reported and EEG-based measures of lapses of attention to the explanation of the worst performance rule). Here, we focus on processes that might explain the high relation of drift rates and WM without resorting to attentional lapses.

As pointed out by Schmiedek et al. (2007), the binding hypothesis by Oberauer, Süß, Wilhelm, and Sander (2007) might provide such a common process. According to the binding hypothesis, the creation, maintenance, and updating of bindings between arbitrary pieces of information within temporary mental representations are fundamental functions of WM. Schmiedek et al. (2007) argued that the quality of bindings between stimuli and responses (e.g., the assignment of response categories to left and right keyboard buttons) may be one such representation contributing to performance on the two-choice decision tasks. Direct support for this hypothesis has been provided by Hülür and colleagues (Hülür, Keye-Ehing, Oberauer, & Wilhelm, 2019) who showed that a specific factor of tasks with arbitrary stimulus-response bindings could explain incremental variance in a WM factor above and beyond variance explained by a general factor of RTs on tasks with compatible or arbitrary bindings.

Note however, that Schmitz and Wilhelm (2016) reported comparatively lower correlations (r = .35 and .41) between latent factors of drift rates from simple decision tasks (search and comparison tasks) and WM than Schmiedek et al. (2007). Also, Dutilh and colleagues (2017) report rather moderate correlations between drift rates in two versions of a perceptual two-choice tasks and a composite score from a WM test battery (r = .24 and .28).

One factor contributing to these differences in correlations may be the complexity of the decision tasks. In Schmiedek et al. (2007), the eight tasks varied in complexity, from decisions as simple as identifying arrows as pointing upwards or downwards to decisions about words having one or two syllables and about filled patterns composing one or two coherent areas. In Schmitz and Wilhelm (2016), tasks also varied in complexity, and drift rates descriptively tended to be more highly related to WM for the more complex tasks. In contrast, Ester, Ho, Brown, and Serences (2014) investigated correlations of WM with drift rates (of a linear ballistic accumulator model; see Brown & Heathcote, 2008) and reported significant correlations for easier, but not for more difficult task conditions for a motion detection and a masked letter discrimination task. Finally, in Dutilh et al. (2017), only one task was used, which required participants to indicate whether a dot matrix contains more black or white dots. In a Bayesian analysis, the authors found inconclusive evidence for the hypothesis that drift rates for a more difficult version of this task (proportion of differently colored dots being 55:45) correlated more strongly with WM than an easier version (proportion of differently colored dots being 60:40).

While acknowledging that findings are mixed, we conclude that there is some indication that the more a task requires complex integration of information, the more WM may play a role in two-choice decision tasks beyond the maintenance of stimulus-response bindings, even when the information is available (on the screen) all of the time. This conclusion is corroborated by a recent study in which diffusion modeling was applied to more complex two-choice tasks, with much slower RTs (about 2-4 sec on average), than typical in extant applications of the diffusion model

(Lerche, von Krause, Voss, Frischkorn, Schubert, & Hagemann, 2020). At this extreme end of high complexity of two-choice tasks, a latent factor of drift rates for these "slow" tasks could explain an additional 10% of variance in fluid intelligence above a factor of drift rates of regular "fast" two-choice tasks.

In contrast to the above considerations, which either consider reasoning of some complexity being part of the decision process in the two-choice tasks, or postulate the involvement of WM in keeping procedural information (i.e., stimulus-response bindings) active, one can turn to perceptual decision-making tasks that require categorizations of minimal complexity, for example a letter discrimination task that only requires to decide whether a single letter is a consonant or a vowel. The diffusion process leading to such a categorization needs constant input of declarative stimulus information for the accumulation of evidence toward one of the response alternatives. This input is either constantly provided by the environment and can continuously be fed into the decision process through visual encoding – or it needs to be converted into a mental representation that maintains the declarative information required for the decision in an active state in case the stimulus is only present for a short period of time (Ratcliff & Rouder, 2000; Smith & Ratcliff, 2009). Active maintenance beyond sensation is a function provided by the WM system but is typically considered to be serving the solution of more complex tasks. In cases where the input from the environment is unstable, though, the visual WM system might serve to keep active the mental representation necessary for the information sampling that is fueling the diffusion process – and this might occur even for stimuli as simple as a single letter.

These considerations are well in accord with computational models of relevant neural processes. Critically, such models contain circuit architectures that serve both maintenance of contents in WM and representations of accumulated evidence in decision making tasks (Murray, Jaramillo, & Wang, 2017; Wang, 2008). An interesting question therefore is whether a relation of WM and drift rates of very simple perceptual decision tasks may be produced by the sheer

requirement of keeping the mental representation of the stimulus active. This requirement can be operationalized by masking stimuli very briefly after presentation (e.g., after < 100 ms; see Ratcliff & Rouder, 2000). Presentation times before mask onset need to be sufficiently long to give all necessary information the opportunity of being encoded but short enough to ensure that the bulk of the evidence accumulation process operates on a representation that is purely mental in the sense that it is no longer supported by the stimulus. In the present study, we aimed at achieving this balance by adjusting presentation times to individuals' performance levels. The three perceptual decision tasks required classifying letters (as consonants or vowels), digits (as odd or even), and figures (as symmetrical or asymmetrical), with all stimuli being based on the seven-segment display as used in "vintage" hand calculators (see Figure 1).

Before onset of the mask, this perceptual input needs to be transformed into a representation that binds the feature "shown" (versus "not shown") to the seven different line positions (e.g., the three horizontal lines and the two lines to the right). Based on these feature bindings, the declarative information of the stimulus identity has to be retrieved (e.g., the digit "3") and then to be categorized according to the task instruction (e.g., as "odd"). With each new trial, these bindings (i.e., the information which lines have been shown) have to be updated. In sum, successful performance on our masked perceptual decision tasks requires mental representations of procedural information (i.e., stimulus-response bindings) as well as of declarative information (i.e., perceptual feature bindings that code stimulus identity). While the procedural information is constant across trials, resulting in consistent stimulus-response mappings, the declarative information needs to be updated at each new trial – which potentially requires WM resources.

Practice-Related Changes and the Evidence-Accumulation Account

If stimulus-response bindings held in WM contribute to individual differences in drift rate, extended practice should reduce the link between drift rates and WM, given that that the associations of stimulus categories and response options become overlearned and automatized after

extended task exposure. In other words, the need to memorize stimulus-response bindings should diminish with practice, and so should the correlation of drift rate and WM.

However, if keeping active mental representations of declarative stimulus information is the common mechanism that produces the observed between-person correlations of WM performance and drift rates of simple decision tasks, then practice-related changes in this mechanism should be associated with both changes in WM performance and changes in drift rates. In this regard, training studies that include practice on WM tasks, simple decision tasks, or both, become instrumental in probing for the existence of a hypothesized shared mechanism. If practice in one domain leads to improvements of the shared mechanism, this should produce transfer to the other domain, and vice versa.

Investigated separately, practice-related improvements in WM performance are well documented (e.g., Adam & Vogel, 2018; Li et al., 2008; Redick et al., 2013) and there is also evidence for increases of drift rates with practice on simple decision tasks (Ratcliff, Thapar, & McKoon, 2006). Here, we make use of data from the COGITO Study (Schmiedek, Lövdén, & Lindenberger, 2010, 2020), a large training intervention in which 101 younger adults practiced a set of twelve cognitive tasks for 100 daily occasions. As these tasks included three WM tasks as well as three simple two-choice decision speed tasks, the study allowed investigating how individual differences in practice-related changes are related across the two task domains. Regarding changes in WM, results did show significant and strong improvements (Cohen's d for net effect sizes, comparing the training to a control group, ranging from 0.90 to 1.20; see Schmiedek et al., 2010) for the three practiced tasks of WM. For three transfer tasks of WM, which were based on the same paradigms as (but using different content material than) the practiced tasks, significant transfer could be demonstrated for one of the tasks (i.e., a spatial 3-back task). Importantly, combining the three WM transfer tasks into a latent factor and evaluating change from pretest to posttest using a latent change score model (McArdle, 2009; see section on Data Analysis), significant transfer could

be demonstrated at the ability construct level (Schmiedek et al., 2010). Furthermore, reliable individual differences were present in latent change score factors for, both, the practiced and the transfer tasks of WM and these latent change score factors were highly correlated (i.e., the factor correlation was not significantly different from 1.00). This pattern of results suggests that one or more common mechanisms underlie training-related changes in the WM practiced and the transfer tasks, with reliable individual differences in the size of these changes being present.

In the present work, we set out to relate these individual differences to individual differences in changes on the simple decision tasks (and associated drift rates), which were practiced in parallel to the WM tasks. These tasks required participants to make very simple decisions as to whether letters were consonants or vowels, digits were odd or even, and simple line figures were symmetric or asymmetric. Stimulus-response bindings were arbitrary, but should have been learned quickly and reliably during the long training phase of the study.

To sum up, the present research had the aim to extend our knowledge on the relation of processing efficiency on simple decision tasks in several important directions. We wanted to investigate: (a) whether individual differences in simple decision performance on tasks that require very basic classifications based on mental representations of masked stimuli are related across different tasks so that latent factors of diffusion model parameters (using the "full diffusion model" and not only the simplified EZ version) can be modelled; (b) whether latent factors of drift rates are related to latent factors of WM; (c) whether individual differences in practice-related changes in the drift rates of different tasks are correlated such that a common factor of drift rate changes can be extracted; and (d) whether such a latent change factor of drift rates is related to a latent change factor of WM.

Conceptually, the main goal of the present study was to discriminate among two competing and equally plausible explanations of the correlational link between individual differences WM and drift rates. According to the hypothesis that stimulus-response bindings tax WM resources,

correlations between drift rates and WM can be expected to decrease with extended practice, given that bindings are increasingly automatized. Also, individual differences in drift rate changes and WM changes are not necessarily expected to reflect the operation of a common mechanism, and hence might be low or absent. In contrast, according to the hypothesis that the active maintenance and updating of declarative stimulus information tax WM resources, correlations between drift rates and WM should continue to be high after extended practice, given that both sets of tasks continue to involve recurrent neural circuits that allow representations to be operated on (e.g., Murray, Jaramillo, & Wang, 2017; Wang, 2008). Also, to the extent that individuals differ in the strength of the effect that practice has on the functioning of this neural architecture, the hypothesis of an active maintenance of declarative stimulus information predicts that practice-induced changes in drift rate are correlated with concurrent changes in WM.

Method

Participants and Procedure

During the training phase, 101 younger adults (51.5% women, $M_{age} = 25.6$ years, $SD_{age} = 2.7$, range: 20-31 years) completed an average of 101 practice sessions ($SD = 2.6$, range: 87-109). Before and after the training, participants completed pre- and posttests during ten sessions that consisted of 2-2.5 hours of comprehensive cognitive test batteries and self-report questionnaires. On average, time elapsing between pre- and posttest was 197 days. Additional information on sample characteristics and study dropout can be found in Schmiedek et al. (2010) and Schmiedek, Bauer, Lövden, Brose, and Lindenberger (2010). The sample was quite representative regarding general cognitive functioning, as indicated by comparisons of Digit-Symbol performance with data from a meta-analysis (Schmiedek et al., 2010). Informed consent was provided by all participants and the study was approved by the local ethics committee.

Tasks

In each session during the training phase, participants practiced twelve different computerized tasks with two to eight blocks each. For perceptual speed, those were the three CRTs (digits: odd vs. even; letters: consonants vs. vowels; figures: symmetric vs. asymmetric) and three comparison tasks (two strings of digits/consonants, or two three-dimensional figures). The three CRTs used in the present work were based on the same stimulus layout – the seven lines of the digit "8" as displayed on "vintage" hand calculators. Stimuli were masked after brief onset intervals (12 ms, 24 ms, 47 ms, or 94 ms) with a stimulus that combined this "calculator 8" with lines extending in all 10 possible directions (see Figure 1).

For episodic memory, tasks required participants to memorize word lists, number-word pairs, or object positions in a grid. WM tasks were adapted versions of the alpha span, numerical memory updating, and spatial n-back tasks (for details of tasks not used in the present investigation, see Schmiedek et al., 2010). Difficulty levels for the CRTs, episodic memory, and WM tasks were individualized using different presentation/masking times based on pretest performance. For each task and each individual, mean accuracies for the different presentation conditions at pretest were fitted with exponential time-accuracy functions (including freely estimated parameters for onset, rate, and asymptote as well as a lower asymptote parameter fixed to different values for each task, which was .10 for memory updating, .50 for 3-back, and .00 for Alpha Span). The fitted values from these functions were used to choose presentation times that were clearly above random guessing but below some upper level. The upper level was defined by the midpoint between the lower asymptote level and perfect accuracy [e.g., (.10 + 1.0)/2 = .55 for Memory Updating], while the minimum level was defined by the midpoint between the lower asymptote level and the upper level [e.g., (.10 + .55)/2 = .325 for Memory Updating]. If performance was above the upper level for the second-fastest presentation time, the fastest presentation time was chosen even if predicted accuracy was below the minimum level for the fastest presentation time. For the Alpha Span task,

we deviated from the described procedure and chose .40 as the minimum level and .60 as the upper level on the basis of empirically observed time-accuracy functions.

For the CRTs, one "slow" and one "fast" condition was chosen for each participant, separately for each task. The algorithm used to choose these masking times attempted to pick conditions with an accuracy at pretest in the range of .625 to .75 for the "fast" and in the range of .875 and .95 for the "slow" condition (details on the algorithm are presented in Supplemental Figure S1). For the letters task, 47 participants got combinations of 12 ms/24 ms, 29 participants of 12 ms/47 ms, 8 participants 24 ms/94 ms, three participants 48 ms/94 ms as fast/slow masking times, while 14 participants (with very high accuracy levels across all masking time conditions at pretest) ended up having 12ms twice, both as slow and fast condition. For the digits tasks, the corresponding numbers of participants were 13 (12 ms/24 ms), 29 (12 ms/47 ms), 12 (12 ms/94 ms), 7 (24 ms/47 ms), 29 (24 ms/94 ms), 10 (47 ms/94 ms), and one participants with 12ms for both conditions. For the figures task, numbers of participants were 15 (12 ms/24 ms), 30 (12 ms/47 ms), 1 (12 ms/94 ms), 13 (24 ms/47 ms), 31 (24 ms/94 ms), 10 (47 ms/94 ms), and one participants with 12 ms for both conditions.

Transfer tasks included computerized tasks as well as 27 tasks from the paper-and-pencil *Berlin Intelligence Structure Test* (BIS test; Jäger, Süß, & Beauducel, 1997). The three near transfer WM tasks used in the present investigation were based on the same three paradigms as the practiced WM tasks, but used different content material.

Choice Reaction Tasks

The full set of stimuli used in each of the three CRTs as well as the masking stimulus used for all three tasks is shown in Figure 1. Below the stimuli, the German words for odd/even, consonant/vowel, and asymmetric/symmetric appeared in red/green color to the left/right. Responses were recorded with a customized button box containing red/green buttons to the left/right, so that the stimulus-response mapping was apparent and consistent across all testing

sessions. Masking times derived from the refreshment rate of the cathode ray tube monitors, with

the masking stimulus appearing one (12 ms), two (24 ms), four (47 ms), or eight (94 ms) screen

cycles after stimulus onset. If no response was made within 3000 ms, the next trial was presented.

At pretest, one of the ten daily testing sessions was dedicated to the CRTs. Participants worked on

the digits, figures, and letters tasks with large numbers of trials each, and breaks in between. For

each task, there were first two practice blocks of 40 trials without masking to familiarize

participants with the stimulus material. Then there were three practice blocks (not used in the

analyses) of 40 trials with masking, followed by 25 blocks of 40 trials with the four masking times

randomly distributed. For the total number of 1,000 trials per task, the numbers of trials for the

different masking time conditions ranged from 231 to 267. At posttest, the 3 x 25 blocks of test

trials were repeated using the same order and procedure. In the daily sessions during the training

phase, participants worked on two blocks of 40 trials for each of the three tasks (with individualized

masking times; see above), with these tasks being interspersed among the nine other tasks of the

daily testing sessions. Therefore the total number of trials during training for each task was about

8,000 (2 blocks x 40 trials x about 100 sessions).

Practiced WM Tasks

Alpha Span. Ten upper-case consonants were presented sequentially together with a

number located below the letter. For each letter, participants had to decide as quickly as possible

whether the number corresponded to the alphabetic position of the current letter within the set of

letters presented up to this step. Five of the ten items were targets. If position numbers were

incorrect (non-targets), they differed from the correct position by +/- one. At pretest, the Alpha

Span task was first practiced with six blocks of trials (not used for the analyses) for each of the four

presentation time conditions of 6000 ms, 3000 ms, 1500 ms, and 750 ms. ISI was 500 ms. Then,

after a break, 48 blocks with presentation times of 6000 ms (6 blocks), 3000 ms (6 blocks), 1500 ms

(6 blocks), 750 ms (6 blocks), 750 ms (6 blocks), 1500 ms (6 blocks), 3000 ms (6 blocks), and 6000

ms (6 blocks) were conducted, with a break after 24 blocks. At posttest, the exact same test blocks were repeated, without preceding practice blocks. The performance measure used in the analyses was based on the percentages of correct responses averaged across presentation time conditions. During the training phase, participants practiced eight blocks of trials at each daily occasion. Possible presentation times during training were 750, 1500, or 3000 ms.

Numerical Memory Updating. Participants had to memorize and update four one-digit numbers. In each of four horizontally placed cells, one of four single digits (from 0 to 9) was presented simultaneously for 4000 ms. After an ISI of 500 ms, a sequence of eight "updating" operations were presented in a second row of four cells below the first one. The updating operations were subtractions and additions from -8 to +8. The updating operations had to be applied to the digits memorized from the corresponding cells above and the new results then also had to be memorized. Each updating operation was applied to a cell different from the preceding one, so that no two updating operations had to be applied to one cell in sequence. The final result for each of the four cells had to be entered at the end of each trial. At pretest, the Memory Updating task was first practiced with memory load increasing from from two digits to four digits with two blocks of trials (not used for the analyses) for each of the four presentation time conditions of 5750 ms, 2750 ms, 1250 ms, and 500 ms. ISI was 250 ms. Then, after a break, 48 blocks with presentation times of 5750 ms (6 blocks), 2750 ms (6 blocks), 1250 ms (6 blocks), 500 ms (6 blocks), 500 ms (6 blocks), 1250 ms (6 blocks), 2750 ms (6 blocks), and 5750 ms (6 blocks) were conducted, with a break after 24 blocks. At posttest, the exact same test blocks were repeated, without preceding practice blocks. The performance measure used in the analyses was based on the percentages of correct responses averaged across presentation time conditions. During the training phase, participants practiced eight blocks of trials at each daily occasion. Possible presentation times during training were 750, 1500, or 3000 ms.

Spatial 3-Back. A sequence of 39 black dots appeared at varying locations in a four-by-four grid. For each dot, participants had to determine whether it was in the same position as the dot three steps earlier in the sequence or not. Dots appeared at random locations with the constraints that (a) 12 items were targets, (b) dots did not appear in the same location at consecutive steps, (c) exactly three items each were 2-, 4-, 5-, or 6-back lures, that is, items that appeared in the same position as they had 2, 4, 5, or 6 steps earlier. In the Spatial 3-Back task, presentation time was always fixed at 500 ms and the ISI was varied. At pretest, the task was first practiced with four blocks of trials (not used for the analyses) for each of the four ISI conditions of 3500 ms, 2500 ms, 1500 ms, and 500 ms. Then, after a break, 32 blocks with ISIs of 3500 ms (4 blocks), 2500 ms (4 blocks), 1500 ms (4 blocks), 500 ms (4 blocks), 500 ms (4 blocks), 1500 ms (4 blocks), 2500 ms (4 blocks), and 3500 ms (4 blocks) were conducted, with a break after 16 blocks. At posttest, the exact same test blocks were repeated, without preceding practice blocks. The performance measure used in the analyses was based on the percentages of correct responses on trials 4-39 averaged across presentation time conditions. During the training phase, participants practiced four blocks of trials at each daily occasion. Possible ISIs during training were 500, 1500, or 2500 ms.

WM Transfer Tasks

Animal Span. As in the alpha span task, a list of consecutively shown stimuli had to be ordered continuously. Instead of letters, six names of animals were shown one after the other, which had to be ordered by size and two-choice decisions on whether a given number corresponds to the current rank order of the present animal had to made. Presentation time was 3000 ms with an ISI of 1000 ms, resulting in a fixed presentation rate of 4000 ms. Eight blocks were conducted in total. Average accuracy was used as the performance score.

Spatial Memory Updating. In each block of this task, first a display of four 3x3 grids was shown for 4000 ms in each of which one black dot was present in one of the nine locations. Those four locations had to be memorized and updated according to shifting operations, which were

indicated by arrows appearing below the corresponding field. Presentation time of the arrows was 2750 ms with an ISI of 250 ms, resulting in a fixed presentation rate of 3000 ms. After six updating operations, the four grids reappeared and the resulting end positions had to be clicked on. After 12 practice blocks with memory load two, six test blocks with load two, six test blocks with load three, and twelve test blocks with load four were conducted and the average accuracy used for scoring.

Numerical 3-Back. As in the spatial version of the 3-back, two-choice decisions on whether the current stimulus matches the stimulus shown 3 steps earlier in the sequence had to be made. Instead of spatial positions, the 39 stimuli were one-digit numbers (1-9). Presentation time was 3000 ms with an ISI of 1000 ms, resulting in a fixed presentation rate of 4000 ms. Six blocks were conducted in total. Average accuracy was used as the performance score.

The Diffusion Model

The diffusion model is designed to explain decision-making in simple two-choice decisions. Decisions are made by a noisy process that accumulates information over time from a starting point z toward one of two decision boundaries, a and 0. When a boundary is reached, a response is initiated. Boundary separation represents how much evidence the subject requires in order to make a decision. The mean of the duration of the nondecision component is labeled T_{er}. Nondecision time represents the time taken for encoding the stimulus, extracting the decision relevant information to produce drift rate, and response output. Within trial variability (noise) in the accumulation of information results in processes with the same drift rate terminating at different times (producing RT distributions) and sometimes at the wrong boundary (producing errors; for detailed descriptions of the diffusion model, see Forstmann, Ratcliff, & Wagenmakers, 2016; Ratcliff & McKoon, 2008; Ratcliff, Smith, Brown, & McKoon, 2016).

The values of the components of processing are assumed to vary from trial to trial, under the assumption that subjects cannot accurately set the same parameter values from one trial to another (e.g., Laming, 1968; Ratcliff, 1978). Across-trial variability in drift rate is normally distributed with

SD η, across-trial variability in starting point is uniformly distributed with range s_z, and across-trial variability in the nondecision component is uniformly distributed with range s_t. Also, there are "contaminant" responses – slow outlier RTs as well as responses that are spurious in that they do not come from the decision process of interest (e.g., distraction, lack of attention). To accommodate these responses, we assume that, on some proportion of trials (p_o), a uniform distributed random delay between the minimum and maximum RT for the condition is added to the decision RT (see Ratcliff & Tuerlinckx, 2002). The assumption of a uniform distribution is not critical; recovery of diffusion model parameters is robust to the form of the distribution and to the form of the across-trial variability components (Ratcliff, 2008, 2013).

The values of all the parameters, including the variability parameters, are estimated simultaneously from data by fitting the model to all the data from all the conditions of each experiment. This was accomplished by minimizing a chi-square value with a general SIMPLEX minimization routine that adjusts the parameters of the model until it finds the parameter estimates that give the minimum chi-square value (see Ratcliff & Tuerlinckx, 2002). The data entered into the minimization routine for each experimental condition were the .1, .3, .5, .7, .9 quantile RTs for correct and error responses and the corresponding accuracy values. The quantile RTs and the diffusion model were used to generate the predicted cumulative probability of a response by that quantile RT. Subtracting the cumulative probabilities for each successive quantile from the next higher quantile gives the proportion of responses between adjacent quantiles. For the χ^2 computation, these are the expected values, to be compared to the observed proportions of responses between the quantiles (i.e., the proportions between 0, .1, .3, .5, .7, .9, and 1.0, which are .1, .2, .2, .2, .2, and .1) multiplied by the number of observations (and the probability for that condition). Summing over $(Observed-Expected)^2/Expected$ for all conditions gives a single χ^2 value to be minimized. Variability in the parameter estimates is much less than differences in the parameters across subjects so that individual differences are meaningful (Ratcliff & Childers, 2015).

Data Analysis

The relation of (changes in) WM and drift rates were analyzed with *latent change score models* (LCSM; McArdle, 2009; McArdle & Nesselroade, 1994), which allow to investigate practice-related changes free from measurement error and at the level of theoretical constructs (i.e., WM and efficiency of evidence accumulation in simple decision tasks) rather than the level of specific tasks (Noack, Lövdén, & Schmiedek, 2014). In these models, latent factors were defined by sets of three tasks (i.e., WM or CRT tasks). Average improvements at the latent factor level were captured by the means of latent change score factors. In order for these means to be interpretable, it is desirable that factor loadings and intercepts are constrained to be equal across occasions and experimental groups (i.e., strong measurement invariance), which was tested prior to further analyses. To estimate and test the correlations of changes in WM and in drift rates of CRTs, a series of bivariate LCSMs (see Kievit et al., 2018, for a tutorial on using bivariate LCSMs to estimate correlated change) were set up, each combining one LCSM for either practiced or transfer WM tasks and one LCSM for the drift rates of the three CRTs (see Figure 2) at one of the four masking time conditions, or using a composite score of the different conditions, resulting in ten separate models. The alpha level for significance tests was set to .05. Analyses were conducted in R (Version 4.1.0; R Core Team, 2022) using the lavaan package (Version 0.6-9; Rosseel, 2012) for structural equation modeling. The study and the analyses presented here were not preregistered. Study materials and analysis code are available on OSF (https://osf.io/5qx7e/). Data are available via a data sharing procedure set up at the Max Planck Institute for Human Development (see https://www.mpib-berlin.mpg.de/1291424/cogito).

Results

Descriptive Statistics

Working Memory

Average performance on the practiced WM tasks improved significantly from pretest to posttest for Alpha Span (M_{Pre} = .634; SD_{Pre} = .076; M_{Post} = .729; SD_{Post} = .108; $t[100]$ = 11.639; p < .05), Numerical Memory Updating (M_{Pre} = .639; SD_{Pre} = .116; M_{Post} = .802; SD_{Post} = .118; $t[100]$ = 16.907; p < .05), and Spatial 3-Back (M_{Pre} = .810; SD_{Pre} = .096; M_{Post} = .911; SD_{Post} = .104; $t[100]$ = 10.734; p < .05). Performance on the WM transfer tasks also improved significantly for Spatial Memory Updating (M_{Pre} = .693; SD_{Pre} = .116; M_{Post} = .802; SD_{Post} = .118; $t[100]$ = 16.907; p < .05) and Numerical 3-Back (M_{Pre} = .887; SD_{Pre} = .092; M_{Post} = .922; SD_{Post} = .086; $t[100]$ = 4.780; p < .05), but barely not for Animal Span (M_{Pre} = .845; SD_{Pre} = .116; M_{Post} = .867; SD_{Post} = .115; $t[100]$ = 1.97 n.s.). As reported in Schmiedek et al. (2010), these changes were significantly larger than those for a control group (not used in the present analyses) for the three practiced WM tasks and for Numerical 3-back.

Choice Reaction Tasks

As shown in Table 1, means and intraindividual SDs of the RT distributions significantly decreased from pretest to posttest for all tasks and all masking time conditions. Regarding accuracies, there was a conspicuous pattern of accuracy increasing significantly from pretest to posttest for the faster masking time conditions (the 12 ms condition for the letters and the digits task and the 12-47 ms conditions for the figures task) and decreasing significantly for the slower masking time conditions (the 47-94 ms conditions for the letters and the digits task and the 94ms conditions for the figures task). In combination, these patterns indicate that combinations of increases in drift rates (which lead to reductions in RT and increases in accuracy) and decreases in boundary separation (which lead to reductions in both, RT and accuracy) may have occurred from pretest to posttest.

Estimation of Diffusion Model Parameters

The fit of the diffusion model to data was good, as indicated by non-significant χ^2 values (see Table 1) and the agreement of model-predicted and observed quantile RTs for correct and error

responses (see Supplemental Figure S2). Table 1 contains descriptive statistics of the diffusion model parameter estimates. As expected, drift rates systematically varied with masking times at pretest ($F(3, 300] = 496.58, 453.34, 629.98$, for the letters, digits, and figures task, respectively) and at posttest ($F(3, 300] = 324.43, 256.15, 237.24$). Drift rates significantly increased from pretest to posttest for all tasks and masking time conditions (see Table 1), indicating that the information processing necessary for decision-making became more efficient with practice. Similarly, nondecision times significantly decreased for all tasks and masking time conditions. Boundary separation was significantly smaller at posttest for all three tasks, indicating that participants lowered their response criteria with practice. Estimates of the trial-to-trial variability in nondecision times and starting point were mostly smaller at posttest and exhibited a mixed pattern of changes for drift rates, with significant increases in variability for the digits tasks, significant reductions in variability for the figures tasks, and no significant changes for the letters task. Estimates for the proportion of contaminant responses were at 1% or lower and did not differ significantly between pretest and posttest. Overall, results indicate that the intensive practice produced substantial changes in all central parameters of the diffusion model, with improvements in drift rates and nondecision times and reductions in boundary separation being key indicators that information processing improved and that participants adapted their response criteria.

Correlations of Diffusion Model Parameters and WM Tasks

In our correlational analyses, we focused on the relations of drift rates (of different tasks, masking times, and measurement occasions) with each other and with performance on the WM tasks. At both, pretest and posttest, drift rates of the same task, but different masking time conditions, did show strong and significant correlations (ranging from .442 to .940) and (descriptively) simplex-like patterns of correlations being stronger for smaller differences in masking times (Table 2). At pretest, correlations of drift rates from different tasks and/or different masking time conditions were mostly significant (46 out of 48) and varied in size from .155 to .673.

At posttest, all these correlations were significant and ranged from .345 to .666. Drift rates from

corresponding task and masking time conditions all correlated significantly between pretest and

posttest, with these stability coefficients ranging in size from .259 to .523. Overall, these results

indicate (a) the presence of common variance across masking time conditions and tasks and (b)

combinations of stability and change that, together, render suitable the use of LCSMs, which allow

to capture the common variance of drift rates from different tasks and to estimate how much of this

common variance is stable and how much is changing from pretest to posttest.

Regarding correlations of drift rates and WM performance, at pretest, 48 of the 72

correlations between the six WM tasks and the twelve drift rate estimates (three tasks with four

masking time conditions each) were significant (Table 2) and the average of these correlations was

.214. At posttest, all but one of these correlations were significant and the average of the

correlations was .355. Comparing the full correlation matrix to one in which the 72 corresponding

correlations of drift rates and WM tasks at pre- vs. posttest (of which 70 were larger at posttest)

were constrained to be equal (using a structural equation model) resulted in a highly significant

difference in model fit (χ^2 [df = 72] = 107.77), indicating a reliable overall pattern of correlations

being higher at posttest. This pattern of increasing correlations between WM tasks and drift rates

from pre- to posttest indicates that individual differences in practice-related changes for both task

domains are positively correlated.

Regarding the other two central parameters of the diffusion model, boundary separation

and nondecision time, their relations to the WM tasks were considerably weaker (reported in

Supplemental Tables S1 and S2). Only 17 of the 72 WM tasks' correlations with nondecision time

parameters were significant at pretest (average r = -.090) and only eight at posttest (average r =

.018). Of the 18 correlations of WM tasks with the boundary separation estimates for the three

CRTs, only four were significant at pretest (average r = .086), and only one at posttest (average r =

-.004). These results confirm the expectation that measures of higher-order cognition, such as WM,

should be related to drift rates much more strongly than to the basic sensory and motor aspects

captured by nondecision time or to the strategic aspect of boundary setting (cf. Ratcliff et al., 2010,

2011).

Measurement Models for Diffusion Model Parameters

Before estimating correlations of latent (change score) factors of drift rates and WM tasks,

we compared series of models with successively imposed constraints on factor loadings and

intercepts to check whether the desired level of strong measurement invariance could be

implemented. To capture the common variance of drift rates from different tasks, but keep the

overall model from becoming too large and complex, we ran separate models for each masking time

condition. Based on modification indices, we decided to not only allow correlations of residuals of

the same tasks at pretest and posttest (to account for the temporal stability of task-specific aspects),

but also of the residuals of the letters and digits tasks (at pretest as well as at posttest). Given that

these two tasks both require the processing of symbolic information (rather than just figural

information as in the symmetry task) and that they have been shown to be associated with brain

activations in overlapping regions (i.e., the left inferior parietal lobule) in analyses with fMRI

versions of the letters and digits tasks done with a subsample of the present study sample (Kühn et

al., 2011), these residual correlations are also well justified on theoretical grounds.

As shown in Table 3, model fit for these models as indicated by CFI and SRMR was

acceptable for models with strong measurement invariance (i.e., factor loadings and intercepts

constrained to be equal at pretest and posttest) at all masking times. The RMSEA was larger than

conventional cutoff criteria for masking times of 24ms and 47ms, however. Directly comparing

models with configural and strong invariance resulted in nonsignificant tests for all but the masking

time condition of 47ms and for the averaged drift rates. Given that the RMSEA is known to be

problematic in models with small degrees of freedom (which was the case with df = 7 in our models

with strong measurement invariance), particularly when sample sizes are not very large (Kenny,

Kaniskan, & McCoach, 2015) and that models with strong measurement invariance are advantageous regarding the interpretability of estimates of latent change, we decided to consistently impose strong measurement invariance on LCMs for drift rates in all subsequent models. As shown in Table 4, standardized factor loadings were significant for all and medium to large in size for most of the task and masking time conditions.

Latent Factor Correlations at Pretest and Posttest

Before investigating (correlated) change with LCSMs, we did look at the correlations of latent drift rate factors and latent WM factors separately at pretest and posttest. As shown in Tables 5 and 7, correlations of latent factors for drift rates at different masking time conditions and the latent factors of practiced and transfer WM tasks were all significant, moderate to strong in size, and numerically larger at posttest. A direct comparison of the correlations at pretest and posttest, however, did not indicate that there was a significant increase in correlations after practice. To further solidify the descriptive observation of larger correlations at posttest, we calculated Bayes factors for competing hypotheses of correlations at posttest being larger, smaller, or the same as at pretest using composite scores of WM tasks and drift rates. Overall, those resulted in strong evidence for increasing (rather than decreasing or stable) correlations (see Supplemental Table S3).

Also descriptively, there was an apparent pattern of correlations increasing from the fastest (12 ms) to the second-but-slowest (47 ms) masking time and then decreasing again for the slowest (94 ms) masking time. In sum, there was consistent evidence of moderate to large correlations of latent drift rate factors with latent factors of practiced and transfer WM tasks. This was also apparent when drift rate estimates were averaged across masking time conditions (Tables 5 and 7). These findings agree with and add to the extant evidence that (latent factors of) drift rates of perceptual decision tasks are positively correlated with WM performance. They render the assumption that this correlation is produced by the requirement to keep stimulus-response bindings

active in WM unlikely, however, as this should lead to a decrease of correlations when these

bindings become automatized during the extensive training period between pretest and posttest.

Latent Change Score Models

To evaluate whether changes in WM performance and changes in drift rates were

associated, bivariate LCSMs were used. In a first step, we tested whether the means of the latent

change score factors of drift rates were significant, indicating reliable mean change common to the

drift rates of the three CRTs. The presence of highly reliable common mean change could be

confirmed for all masking time conditions (χ^2 values [df = 1] of 125.75, 134.01, 58.09, and 14.52,

for masking times of 12, 24, 47, and 94 ms, respectively) as well as for drift rate estimates averaged

across masking time conditions (χ^2 [df = 1] = 90.35). In a second step, we tested whether the

variances of the latent change score factors for drift rates were significant, that is, whether there was

evidence of reliable individual differences in the amount of change in drift rates. Comparing

LCSMs with or without the variance of the change factor (and the associated covariance of the

change with the pretest factor) fixed to zero, the resulting likelihood ratio test indicated significant

amounts of change variance for all masking time conditions (χ^2 values [df = 2] of 33.74, 24.70,

13.71, and 29.23, for masking times of 12, 24, 47, and 94 ms, respectively) as well as for drift rate

estimates averaged across masking time conditions (χ^2 [df = 2] = 31.60).

Results of bivariate LCSMs for practiced and transfer WM tasks are presented in Tables 6

and 8. Model fit was acceptable for the different models and estimates of the latent change-change

correlations of drift rates and WM were all positive, yet reached significance only for the

correlation of practiced WM tasks at the masking time condition of 24 ms ($r = .495$) and for drift

rate estimates averaged across masking time conditions ($r = .371$; see Figure 2). To illustrate this

relation using manifest variables, we created a vector plot (Supplemental Figure S3) and a scatter

plot (Supplemental Figure S4) of the bivariate changes using composite scores of WM tasks and

drift rates. In sum, the positive change-change correlations suggest that the mechanisms driving the improvements in WM performance and the increases in drift rates are partly shared.

Discussion

Overview

The results of the present study showed that (a) observed performance on the three CRTs could be fit by a full diffusion model and that drift rates of different tasks were correlated in a way that common drift rate factors could be modeled; (b) these drift rate factors did correlate significantly and substantially with latent factors of WM performance (on, both, practiced and transfer tasks) at pretest and at posttest; (c) there were significant average changes in drift rates as well as reliable individual differences in the common change captured by latent change factors, and (d) there was indication that these changes were systematically related to changes in WM performance. While positive answers to questions (a) to (c) were provided for all masking time conditions and practiced WM as well as transfer WM tasks, the positive answer to question (d) was restricted, as only the latent correlations for one of the drift rate factors based on different masking times and for the factor of drift rates averaged across masking time conditions with the factor of practiced WM tasks was significant.

Results at Pretest and Posttest

The chosen masking times had the desired effect of creating a range of accuracy from somewhat above chance level for the fastest to very high values for the slowest masking times. This provided the necessary number of error trials for diffusion model estimation and led to drift rates varying systematically across masking time conditions. Across masking times and tasks, drift rates were systematically correlated, so that common factors could be modeled. To reduce the overall complexity of models, we ran separate analyses for the different masking time conditions (or for drift rates averaged across masking time conditions). Within masking time conditions, we modeled common factors for drift rates of the different tasks to capture aspects of the decision process

efficiency that generalize across different stimulus domains (i.e., letters, digits, figures). The resulting drift rate factors did strongly correlate with WM factors of practiced and transfer tasks at pretest ($r = .522$ and $r = .592$, respectively, for averaged drift rates) and at posttest ($r = .566$ and $r = .714$, respectively, for averaged drift rates). While these latent factor correlations based on multiple tasks should not be compared to the correlations for single tasks in Dutilh et al. (2017) and Ester et al. (2014), they are higher than those reported for latent factors of drift rates in Schmitz and Wilhelm (2016) and comparable in size to the correlation reported in Schmiedek et al. (2007). How can these strong correlations be explained?

In comparison to some of the tasks used in the Schmiedek et al. (2007) study, we kept the complexity of the required decision in the three CRTs used here low. Also, the need to keep stimulus-response mappings in WM should play a minimal role, particularly after 100 days of training with these mappings being kept constant throughout. As the correlations of drift rates with WM tended to increase, rather than decrease, from pretest to posttest, the role of WM in CRT performance based on keeping active the procedural information of stimulus-response mappings can be considered negligible.

What is unique about our tasks and in our view plays a critical role in producing correlations with WM, however, is the fact that stimuli were masked after very brief time intervals. Less than 100ms after stimulus onset (the maximum masking time was 94ms), even very simple decision processes can only have started, but not yet have finished. After the mask has appeared, the sampling of information required for continuous evidence accumulation in the diffusion process therefore requires a mental representation of the stimulus from which declarative information can be sampled. As visual attention and visual WM are closely related in that "visual WM allows visual information to be actively extended in the absence of sensory input" (Chun, 2011, p. 1407; see also Chun, Golomb, & Turk-Browne, 2011), WM may serve this important role of establishing and keeping active the representation of the stimulus from which information can be sampled. Such

shared requirements of WM and decision-making tasks are also proposed in models that contain

overlapping frontoparietal circuits for WM maintenance and information accumulation in decision

making (Murray et al., 2017; Wang, 2008). Regarding the binding hypothesis of WM by Oberauer

et al. (2007), the critical set of arbitrary bindings to be held active in WM, and to be updated at each

new trial, is the binding of the seven line positions of the general ("hand calculator 8") stimulus

layout and the information of whether each of these lines has been part (i.e., shown in black) of the

stimulus just presented, or not. Interindividual differences in how efficiently the WM system can

create these bindings, and how reliably the mental representation build up on them can be kept

active during decision making, may then directly influence the size of the resulting drift rate –

producing a positive correlation of drift rate and WM.

Practice-Related Changes

After about 100 practice sessions, drift rates did improve on average for all tasks and

masking time conditions. Concurrently, nondecision times also improved and boundary separation

decreased. Together, this indicates that with extended practice, information processing did become

more efficient, both, at the decision and nondecision (sensory encoding and motor preparation and

execution) levels. Concurrently, participants tended to adjust their response criteria so that RTs

reduced considerably, while accuracy levels remained high.

Given that (a) both, WM performance and drift rates of the CRTs, substantially improved

with practice, (b) baseline levels of WM and drift rates were highly correlated, and (c) reliable

individual differences in change were present for the WM as well as for the drift rate factors, it was

of central interest whether these individual differences in practice-related changes in WM and drift

rates were also coupled with each other. If this is the case, it suggests that observable practice-

related improvements in both domains are driven by improvements of the same underlying shared

mechanisms that already produced the baseline correlations. These underlying improvements, in

turn, may be driven by practice on the WM tasks, by practice on the CRTs, or both acting in

concert. Alternatively, it may also be the case that improvements in each task domain were

separately produced by cognitive mechanisms, strategies, and motivational influences that were

specific to each task domain. The finding that correlations of drift rate and WM factors tended to be

higher (though not significantly so) at posttest than at pretest suggests that shared mechanisms

rather dominate domain-specific mechanisms.

As a direct test of the coupling of practice-related changes in WM and drift rates, the

estimate of the latent change-change correlation was significant only for practiced WM tasks and

for the masking time condition of 24 ms when estimated separately for different masking times.

When averaging across masking time conditions, the change-change correlation was significant for

the practiced tasks of WM, but not for the transfer tasks. Descriptively, and similar to the

correlations at pretest and at posttest, there was a clear pattern of the change-change correlations of

drift rates and WM increasing from very fast to intermediate masking times and then decreasing

again towards the slower masking times – for practiced WM tasks as well as for transfer WM tasks.

A straightforward explanation for this observed pattern of (non-)significant change-change

correlations is that the true correlations are strongest for intermediate masking times and statistical

power was only sufficient to detect the strongest correlation (of about .50). This is likely to be the

case, as the statistical power to detect correlated change is known to be lower than the power to

detect correlated levels (Hertzog, Lindenberger, Ghisletta, & von Oertzen, 2006). The finding that

the estimated change-change correlation of .495 for practiced WM tasks and a masking time of 24

ms was significant, but the estimated correlation of .499 for transfer WM tasks and a masking time

of 47 ms was not, can be explained by the fact that the reliability of the factors for the practiced

WM tasks was higher than for the practiced than for the transfer tasks (McDonald's ω = .69 and .83

at pretest and posttest for the factor of practiced WM tasks, and ω = .56 and .62 at pretest and

posttest for the factor of transfer WM tasks, respectively). This is likely due to the measures of

practiced tasks being based on composites of larger numbers of trials (aggregated across different presentation time conditions) than the transfer tasks.

If we assume that the inverse U-shaped relation of correlations being strongest for intermediate presentation times was indeed reliable, it may indicate that there is a "sweet spot" when presentation times are in the range of about 20ms-50ms. In this range, the presentation time is long enough to allow for encoding of at least some stimulus features (e.g., a subset of the up to seven segments of our stimuli), but still sufficiently short, such that the speed and reliability with which these features are transformed into a stable mental representation (i.e., the function we hypothesize to be supported by WM) is an important limiting factor in task performance. With longer presentation times, even less efficient WM systems are presumably able to create a sufficiently strong mental representation for the evidence accumulation process to work on, rendering individual differences in WM less relevant to performance on the perceptual decision tasks. Given that we lack the statistical power to confirm this non-linear pattern with the present data, confirmation of these considerations will require additional research.

The result of a reliable change-change correlation for the averaged drift rates and practiced WM tasks suggests that the same explanation that we offered for the correlations at pretest and posttest might also hold for the practice-related changes. The basic WM functions of establishing, keeping active, and updating mental representations composed of bindings between arbitrary elements of information may support the storage and processing requirements of the WM tasks, as well as the sampling of information from stimuli that have been masked after brief presentation. During the training phase of the study, both the practice on WM tasks and on the CRTs could have led to improvements of this basic WM function. Disentangling these different possibilities would require a training study design with different experimental groups, practicing either only on the WM tasks or on the CRTs. Not having implemented such separate training groups denotes a limitation of the present work and an interesting opportunity for further research.

Another limitation of our study is that we did not include task conditions of the CRTs without masking (apart from a few practice trials at the beginning of the study). This precludes the possibility to directly estimate the effect that masking has on the relation of WM and drift rates. As the CRTs without masking are very easy, however, huge numbers of trials would be necessary to allow for the estimation of diffusion model parameters. Therefore, a more focused correlational study on the relation of WM and drift rates estimated from similar tasks with and without masking would be desirable.

Conclusion and Outlook

The main goal of the present study was to discriminate among the accounts that (a) the maintenance of stimulus-response bindings versus (b) the fast creation, maintenance, and trial-to-trial updating of mental representations of stimuli are contributing to the correlational link between estimates of evidence accumulation based on simple decision-making tasks, on the one hand, and WM tasks, on the other. We found that correlations between individual differences in drift rates and individual differences in WM continue to be strong after extended practice, and that individual differences in drift rate changes are correlated with individual differences in WM changes. These results were obtained at the level of latent factors, which strengthens the reliability and generality of our findings (Little, Lindenberger, & Nesselroade, 1999). In line with neural inquiries into the computational basis of WM (e.g., Murray, Jaramillo, & Wang, 2017; Wang, 2008), our results support the account of active maintenance of declarative stimulus information, rather than an account attributing the main role of WM to the maintenance of procedural stimulus-response bindings. While we cannot exclude the possibility that additional mechanisms (like the propensity to have attentional lapses) also contribute to the correlations between WM and (particularly relatively slow) RTs, we hold that theory development should consider the involvement of WM mechanisms in even very simple perceptual decision tasks – particularly if the necessary information is not reliably and constantly available in the environment. Our results also suggest that

basic mechanisms of circuits of active maintenance might improve with extended practice. Future

work needs to test this hypothesis more directly at both behavioral and neural levels of analysis.

Context Paragraph

In 2007, one of us published an influential paper (Schmiedek et al., 2007) in this journal, in which a high relation of latent factors of intelligence and WM to a latent factor of drift rates of eight different two-choice tasks was demonstrated. Limitations of this work were that trial numbers allowed only the use of a simplified (i.e., the EZ) version of the diffusion decision model and that some of the two-choice tasks required decisions of non-trivial complexity. When planning the COGITO Study (Schmiedek et al., 2020), a study focusing on within-person variability and practice-related change in cognitive performance across 100 training sessions, we therefore included sufficient trial numbers for each of three very basic perceptual decision tasks at pre- and posttest to allow fitting the full diffusion model to individual participants' data, allowing us to again model relations at the latent factor level. Furthermore, the use of brief masks allowed us to evaluate the idea that active maintenance of mental representations is a mechanism shared by perceptual decision making and WM. Finally, the intensive training allowed us to address the critical question whether the association between WM and drift rates exists not only for individual differences at baseline, but also for individual differences in practice-related changes.

References

Adam, K. C. S., & Vogel, E. K. (2018). Improvements to visual working memory performance with practice and feedback. *PLoS One, 13,* e0203279. doi: 10.1371/journal.pone.0203279

Baumeister, A. A., & Kellas, G. (1968). Distribution of reaction times of retardates and normals. *American Journal of Mental Deficiency, 72*, 715–718.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive Psychology, 57*, 153–178. doi: 10.1016/j.cogpsych.2007.12.002

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia, 49*, 1407–1409. doi: 10.1016/j.neuropsychologia.2011.01.029

Chun, M. M., Golomb, J. A., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology, 62*, 73–101. doi: 10.1146/annurev.psych.093008.100427

Coyle, T. R. (2003). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence, 31*, 567–587. doi: 10.1016/S0160-2896(02)00175-7

Deary, I. J. (2000). *Oxford psychology series, No. 34. Looking down on human intelligence: From psychometrics to the brain.* Oxford University Press. doi: 10.1093/acprof:oso/9780198524175.001.0001

Dutilh, G., Vanderkerckhove, J., Ly, A., Matzke, D., Pedroni, A., Frey, R., Rieskamp, J., & Wagenmakers, E.-J. (2017). A test of the diffusion model explanation for the worst performance rule using preregistration and blinding. *Attention, Perception, and Psychophysics, 79*, 713–725. doi: 10.3758/s13414-017-1304-y

Ester, E. F., Ho, T. C., Brown, S. D., & Serences, T. C. (2014). Variability in visual working memory ability limits the efficiency of perceptual decision making. *Journal of Vision, 14*, 1–12. doi: 10.1167/14.4.2

Hertzog, C., Lindenberger, U., Ghisletta, P., & von Oertzen, T. (2006). On the power of

multivariate latent growth curve models to detect correlated change. *Psychological Methods, 11*, 244–252. doi: 10.1037/1082-989X.11.3.244

Hülür, G., Keye-Ehing, D., Oberauer, K., & Wilhelm, O. (2019). The effect of stimulus-response

compatibility on the association of fluid intelligence and working memory with choice

reaction times. *Journal of Cognition, 2*, Article 14, 1–19. doi: 10.5334/joc.66

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Der Berliner Intelligenzstruktur-Test (BIS-Test;*

Form 4). Test manual. Göttingen: Hogrefe.

Jeffreys, H. (1961). *Theory of probability.* Oxford, UK: Oxford University Press.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Elsevier.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2015). Working memory capacity and fluid

intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle

(2005). *Psychological Bulletin, 131*, 66–71. doi: 10.1037/0033-2909.131.1.66

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models

with small degrees of freedom. *Sociological Methods & Research, 44,* 486–507. doi:

10.1177/0049124114543236

Kievit, R., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., de Mooij, S., Moutoussis, M.,

Goodyer, I., Bullmore, E., Jones, P., Fonagy, P., the Neuroscience in Psychiatry Network

(NSPN) Consortium, Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive

neuroscience using latent change score models: A tutorial and applications. *Developmental*

Cognitive Neuroscience, 33, 99–117. doi:10.1016/j.dcn.2017.11.007

Kühn, S., Schmiedek, F., Schott, B., Ratcliff, R., Heinze, H.-J., Düzel, E., Lindenberger, U., &

Lövdén, M. (2011). Brain areas linked to individual differences in perceptual decision-

making in younger and older adults before and after 100 days of practice. *Journal of Cognitive Neuroscience, 23,* 2127–2138. doi: 10.1162/jocn.2010.21564

Laming, D. R. J. (1968). *Information theory of choice-reaction times*. Academic Press.

Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst-performance" analysis of individual differences. *Intelligence, 14*, 309–325. doi: 10.1016/0160-2896(90)90021-K

Lerche, V., von Krause, M., Voss, A., Frischkorn, G., Schubert, A.-L., & Hagemann, D. (in press). Diffusion modeling and intelligence: Drift rates show both domain-general and domain-specific relations with intelligence. *Journal of Experimental Psychology: General, 149,* 2207–2249. doi: 10.1037/xge0000774

Li, S.-C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., & Lindenberger, U. (2008). Working memory plasticity in old age: Practice gain, transfer, and maintenance. *Psychology and Aging*, *23,* 731–742. doi: 10.1037/a0014343

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4*, 192-211. doi: 10.1037/1082-989X.4.2.192

Löffler, C., Frischkorn, G. T., Rummel, J., Hagemann, D., & Schubert, A.-L. (2022). Do attentional lapses account for the worst performance rule? *Journal of Intelligence, 10*, 2. doi: 10.3390/jintelligence10010002

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577–605. doi:10.1146/annurev.psych.60.110707.163612

McArdle, J. J., & Nesselroade, J. R. (1994). *Using multivariate data to structure developmental change.* Lawrence Erlbaum Associates, Inc.

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in

reading comprehension? On the influence of mind wandering and executive attention.

Journal of Experimental Psychology: General, 141, 302–320. doi: 10.1037/a0025250

Mulder, J., Williams, D. R., Gu, X., Tomarken, A., Böing-Messing, F., Olsson-Collentine, A.,

Meyerink-Bosman, M., Menke, J., van Aert, R., Fox, J.-P., Rosseel, Y., Wagenmakers, E.-J.,

& van Lissa, C. (2021). BFpack: Flexible Bayes Factor Testing of Scientific Theories in R.

Journal of Statistical Software, 100, 1–63. doi:10.18637/jss.v100.i18

Murray, J. D., Jaramillo, J., & Wang, X.-J. (2017). Working memory and decision-making in a

fronto-parietal circuit model. *Journal of Neuroscience, 37,* 12167–12186. doi:

10.1523/JNEUROSCI.0343-17.2017

Noack, H., Lövdén, M., & Schmiedek, F. (2014). On the validity and generality of transfer effects

in cognitive training research. *Psychological Research, 78*, 773–798. doi: 10.1007/s00426-

014-0564-6

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working

memory capacity and reasoning ability. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J.

Towse (Eds.), *Variation in working memory* (pp. 49–75). Oxford: Oxford University Press.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for

Statistical Computing, Vienna, Austria. https://www.R-project.org/.

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

10.1037/0033-295X.85.2.59

Ratcliff, R. (2008). The EZ diffusion method: Too EZ? *Psychonomic Bulletin and Review*, *15*,

1218-1228. doi: 10.3758/PBR.15.6.1218

Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model.

Psychological Review, *120*, 281-292. doi: 10.1037/a0030775

Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice
diffusion model. *Decision*, *2*, 237-279. doi: 10.1037/dec0000030

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice
decision tasks. *Neural Computation*, *20*, 873-922. doi: 10.1162/neco.2008.12-06-420

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter
identification. *Journal of Experimental Psychology: Human Perception and Performance,
26*, 127–140. doi: 10.1037/0096-1523.26.1.127

Ratcliff, R., Schmiedek, F., & McKoon, G. (2008). A diffusion model explanation of the worst
performance rule for reaction time and IQ. *Intelligence*, *36*, 10-17.
doi:10.1016/j.intell.2006.12.002

Ratcliff, R., Smith, P.L., Brown, S.D., & McKoon, G. (2016). Diffusion decision model: Current
issues and history. *Trends in Cognitive Science*, *20*, 260-281. doi: 10.1016/j.tics.2016.01.007

Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging, practice, and perceptual tasks: A diffusion
model analysis. *Psychology and Aging*, *21*, 353–371. doi: 10.1037/0882-7974.21.2.353

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice
tasks. *Cognitive Psychology*, *60*, 127–157. doi:10.1016/j.cogpsych.2009.09.001

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative
memory. *Journal of Experimental Psychology: General*, *140*, 464–487.
doi:10.1037/a0023810

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to
dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin
and Review*, *9*, 438-481. doi: 10.3758/BF03196302

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., ... &
Engle, R. W. (2013). No evidence of intelligence improvement after working memory

training: A randomized, placebo-controlled study. *Journal of Experimental Psychology:*

General, *142*, 359–379. doi: 10.1037/a0029082

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical*

Software, 48, 1-36. doi: 10.18637/jss.v048.i02

Schmiedek, F. (2022, August 8). Practice-Related Changes in Perceptual Evidence Accumulation

Correlate with Changes in Working Memory. Retrieved from osf.io/5qx7e

Schmiedek, F., Bauer, C., Lövdén, M., Brose, A., & Lindenberger, U. (2010). Cognitive enrichment

in old age: Web-based training programs. *Journal of Gerontopsychology and Geriatric*

Psychiatry, *23*, 59–67. doi:10.1024/1662-9647/a000013

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex

span versus updating tasks of working memory: The gap is not that deep. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 35, 1089–1096. doi:

10.1037/a0015730

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training

enhance broad cognitive abilities in adulthood: Findings from the COGITO study. *Frontiers*

in Aging Neuroscience, *2*, 1–10. doi:10.3389/fnagi.2010.00027

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2014). A task is a task is a task: Putting complex

span, n-back, and other working memory paradigms in psychometric context. *Frontiers in*

Psychology, 5, Article 1475. doi: 10.3389/fpsyg.2014.01475

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2020). Training working memory for 100 days:

The COGITO Study. In J. M. Novick, M. F. Bunting, M. R. Dougherty, & R. W. Engle

(Eds.), *Cognitive and Working Memory Training: Perspectives from Psychology,*

Neuroscience, and Human Development (pp. 40–57). Oxford University Press.

Schmitz, F. & Wilhelm, O. (2016). Modeling mental speed: Decomposing response time

distributions in elementary cognitive tasks and correlations with working memory capacity

and fluid intelligence. *Journal of Intelligence, 4*, 13. doi: 10.3390/jintelligence4040013

Schubert, A.-L. (2019). A meta-analysis of the worst performance rule. *Intelligence, 73*, 88–100.

doi: 10.1016/j.intell.2019.02.003

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual

signal detection. *Psychological Review*, *116*, 283-317. doi: 10.1037/a0015156

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model

for response time and accuracy. *Psychonomic Bulletin & Review*, 14, 3–22. doi:

10.3758/BF03194023

Wang, X.-J. (2008). Decision making in recurrent neuronal circuits. *Neuron, 60*, 215–234. doi:

10.1016/j.neuron.2008.09.034

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how

can we measure it? *Frontiers in Psychology, 4,* Article 433. doi: 10.3389/fpsyg.2013.00433

Table 1

Descriptive Information for Performance on and Diffusion Model Parameter Estimates of the CRTs

Task	Masking time	Descriptives			Diffusion Model Parameter Estimates							χ^2	$n(sig.\ \chi^2)$
		$M\ (RT)$ (ms)	$SD\ (RT)$ (ms)	Acc	ν	a	T_{er}	η	s_z	s_t	p_o		
Letters	12ms	592 (114) / 426 (104) $t = 13.446$*	221 (74) / 125 (60) $t = 12.022$*	.640 (.087) / .671 (.089) $t = 2.727$*	0.100 (0.069)/ 0.162 (0.101) $t = 5.830$*	0.110 (0.025)/ 0.080 (0.022) $t = 9.901$*	0.370 (0.046)/ 0.315 (0.048) $t = 9.684$*	0.170 (0.082)/ 0.188 (0.110) $t = 1.472$	0.056 (0.021)/ 0.044 (0.020) $t = 4.545$*	0.146 (0.050)/ 0.110 (0.038) $t = 6.686$*	0.006 (0.011)/ 0.006 (0.015) $t = 0.159$	28.6 33.1	11 15
	24ms	578 (99) / 416 (90) $t = 14.614$*	211 (70) / 121 (52) $t = 11.600$*	.740 (.103) / .748 (.109) $t = 0.587$ n.s.	0.183 (0.101)/ 0.250 (0.133) $t = 4.984$*		0.372 (0.042)/ 0.314 (0.044) $t = 10.857$*						
	47ms	557 (81) / 404 (75) $t = 16.573$*	178 (66) / 106 (50) $t = 9.942$*	.861 (.090) / .829 (.111) $t = 2.907$*	0.315 (0.130)/ 0.361 (0.168) $t = 2.725$*		0.386 (0.039)/ 0.316 (0.045) $t = 12.909$*						
	94ms	536 (64) / 403 (67) $t = 17.174$*	150 (59) / 97 (42) $t = 8.330$*	.927 (.057) / .877 (.109) $t = 5.114$*	0.425 (0.144)/ 0.462 (0.178) $t = 2.035$*		0.396 (0.039)/ 0.324 (0.048) $t = 12.811$*						
Digits	12ms	721 (137) / 442 (86) $t = 20.251$*	302 (98) / 128 (65) $t = 15.608$*	.718 (.124) / .752 (.118) $t = 2.230$*	0.162 (0.099)/ 0.269 (0.137) $t = 7.586$*	0.141 (0.031)/ 0.090 (0.023) $t = 14.531$*	0.397 (0.047)/ 0.321 (0.039) $t = 14.153$*	0.199 (0.076)/ 0.238 (0.111) $t = 3.267$*	0.062 (0.031)/ 0.051 (0.027) $t = 2.588$*	0.150 (0.047)/ 0.113 (0.042) $t = 7.680$*	0.010 (0.020)/ 0.008 (0.020) $t = 0.723$	26.0 34.9	6 22
	24ms	656 (113) / 426 (79) $t = 20.470$*	264 (90) / 123 (71) $t = 13.321$*	.849 (.095) / .841 (.109) $t = 0.677$ n.s.	0.289 (0.112)/ 0.409 (0.174) $t = 7.550$*		0.391 (0.044)/ 0.318 (0.037) $t = 14.811$						
	47ms	604 (92) / 413 (70) $t = 20.405$*	218 (80) / 106 (61) $t = 12.514$*	.923 (.058) / .886 (.093) $t = 4.743$*	0.405 (0.125)/ 0.500 (0.175) $t = 5.746$*		0.393 (0.041)/ 0.321 (0.036) $t = 15.196$*						
	94ms	588 (80) / 416 (66) $t = 20.719$*	185 (71) / 99 (56) $t = 11.206$*	.953 (.043) / .900 (.089) $t = 5.954$*	0.465 (0.138)/ 0.535 (0.179) $t = 3.738$*		0.405 (0.040)/ 0.329 (0.039) $t = 16.791$*						
Figures	12ms	546 (107) / 366 (62) $t = 17.301$*	213 (73) / 101 (41) $t = 15.182$*	.631 (.102) / .805 (.123) $t = 14.959$*	0.101 (0.086)/ 0.330 (0.170) $t = 15.895$*	0.107 (0.024)/ 0.077 (0.018) $t = 10.704$*	0.336 (0.044)/ 0.279 (0.033) $t = 12.385$*	0.196 (0.094)/ 0.159 (0.092) $t = 2.883$*	0.047 (0.028)/ 0.047 (0.022) $t = 0.101$	0.141 (0.046)/ 0.096 (0.032) $t = 9.975$*	0.008 (0.018)/ 0.006 (0.012) $t = 1.036$	26.9 33.2	6 13
	24ms	529 (90) / 357 (60) $t = 19.060$*	192 (61) / 91 (38) $t = 15.226$*	.719 (.116) / .863 (.110) $t = 13.190$*	0.182 (0.122)/ 0.436 (0.1) $t = 16.549$*		0.341 (0.043)/ 0.280 (0.034) $t = 13.655$*						
	47ms	502 (70) / 349 (54) $t = 22.063$*	158 (49) / 84 (36) $t = 14.523$*	.886 (.090) / .908 (.093) $t = 2.714$*	0.381 (0.162)/ 0.538 (0.185) $t = 8.963$*		0.354 (0.041)/ 0.280 (0.034) $t = 17.533$*						
	94ms	483 (57) / 349 (50) $t = 22.663$*	125 (42) / 81 (35) $t = 9.122$*	.954 (.044) / .923 (.074) $t = 5.188$*	0.515 (0.148)/ 0.573 (0.163) $t = 3.214$*		0.366 (0.045)/ 0.282 (0.035) $t = 19.280$*						

Note. Values are for pretest/posttest. Values in brackets are the between-person SDs; Acc = accuracy; v = drift rate; a = boundary separation; T_{er} = nondecision time; η = SD in drift across trials; s_z = range of the distribution of starting point (z); s_t = range of the distribution of nondecision times; p_0 = proportion of contaminants. Because the two response options produced symmetrical data, the correct responses for the two options were combined and the same for errors, and the starting point z was set to $a/2$. Note that drift rates and nondecision times were estimated with separate parameters for the different masking time conditions (but the same parameter for the two response options). All other diffusion model parameters were constrained to be the same across masking time conditions. Paired t tests (absolute values shown) for diffusion model parameters (comparing pretest and posttest values) are based on $df = 100$. χ^2 values describe overall model fit ($df = 31$; critical value for alpha level of .05 is 44.99). $n(sig. \chi^2)$ = number of participants with significant model misfit according to χ^2 test.

* $p < .05$.

Table 2

Correlations of Drift Rate Estimates for the Three CRTs at Different Masking Time Conditions Among Each Other and With WM Tasks

Masking time	Pretest												Posttest											
	Letters				Digits				Figures				Letters				Digits				Figures			
	12ms	24ms	47ms	94ms	12ms	24ms	47ms	94ms	12ms	24ms	47ms	94ms	12ms	24ms	47ms	94ms	12ms	24ms	47ms	94ms	12ms	24ms	47ms	94ms
Pretest																								
Letters																								
24ms	.875																							
47ms	.610	.827																						
94ms	.488	.690	.875																					
Digits																								
12ms	.673	.637	.464	.335																				
24ms	.561	.610	.590	.524	.804																			
47ms	.393	.471	.571	.588	.579	.862																		
94ms	.320	.394	.499	.562	.443	.764	.910																	
Figures																								
12ms	.441	.452	.367	.260	.516	.446	.270	.182																
24ms	.463	.508	.453	.369	.500	.467	.343	.265	.940															
47ms	.384	.498	.557	.517	.382	.422	.385	.344	.742	.859														
94ms	.237	.345	.489	.492	.155	.256	.333	.329	.447	.556	.779													
Posttest																								
Letters																								
12ms	.259	.294	.288	.156	.225	.271	.231	.144	.309	.301	.257	.242												
24ms	.286	.369	.351	.215	.258	.291	.208	.104	.380	.382	.313	.252	.873											
47ms	.273	.343	.380	.297	.301	.329	.257	.144	.342	.348	.361	.274	.725	.845										
94ms	.348	.397	.442	.374	.335	.381	.333	.194	.346	.349	.362	.300	.624	.759	.925									
Digits																								
12ms	.272	.368	.372	.274	.322	.377	.331	.303	.335	.344	.325	.207	.591	.574	.570	.470								
24ms	.265	.359	.378	.295	.376	.434	.368	.305	.349	.361	.373	.237	.547	.575	.648	.595	.897							
47ms	.290	.399	.427	.409	.388	.457	.415	.340	.291	.331	.406	.285	.429	.510	.660	.641	.735	.882						
94ms	.301	.374	.431	.396	.343	.430	.402	.318	.219	.268	.362	.254	.378	.430	.598	.599	.615	.787	.915					
Figures																								
12ms	.308	.366	.325	.245	.364	.443	.344	.266	.523	.558	.538	.351	.526	.537	.530	.503	.607	.666	.629	.567				
24ms	.273	.360	.378	.296	.352	.489	.400	.314	.489	.523	.546	.371	.475	.491	.516	.505	.593	.657	.626	.566	.921			
47ms	.236	.302	.363	.294	.308	.404	.380	.299	.392	.440	.491	.368	.409	.443	.523	.552	.482	.599	.614	.587	.788	.861		
94ms	.280	.331	.390	.364	.314	.427	.393	.322	.344	.389	.463	.327	.345	.408	.550	.586	.427	.587	.658	.675	.719	.805	.903	

Pretest
WM Practiced

Alpha Span	.225	.235	.261	.292	.228	.378	.364	.247	.188	.237	.207	.085	.203	.236	.288	.326	.281	.343	.378	.384	.244	.295	.255	.278
Spatial N-Back	.063	.174	.233	.263	.123	.227	.193	.136	.253	.259	.257	.206	.076	.226	.283	.320	.219	.228	.296	.228	.283	.279	.267	.244
Numerical MU	.001	.125	.282	.293	.113	.241	.242	.177	.206	.240	.247	.257	.249	.317	.317	.291	.297	.284	.294	.251	.244	.249	.221	.179
WM Transfer																								
Animal Span	.159	.216	.265	.222	.211	.314	.310	.178	.261	.259	.253	.097	.223	.231	.323	.326	.284	.369	.391	.384	.302	.323	.298	.293
Numerical N-back	.162	.228	.281	.299	.231	.361	.332	.253	.169	.230	.238	.241	.083	.170	.201	.265	.210	.227	.285	.270	.262	.265	.209	.223
Spatial MU	.027	.127	.250	.269	.091	.229	.253	.189	.160	.121	.094	.097	.309	.380	.481	.453	.344	.388	.380	.295	.266	.296	.266	.271
Posttest																								
WM Practiced																								
Alpha Span	.164	.226	.312	.301	.248	.417	.392	.255	.308	.337	.318	.239	.307	.350	.415	.451	.360	.407	.430	.364	.420	.516	.452	.447
Spatial N-Back	.125	.197	.283	.304	.168	.296	.316	.216	.261	.245	.249	.299	.276	.353	.448	.481	.315	.402	.501	.440	.377	.404	.392	.419
Numerical MU	.080	.216	.322	.379	.073	.232	.263	.197	.205	.217	.295	.326	.203	.290	.375	.393	.282	.340	.390	.329	.291	.370	.318	.330
WM Transfer																								
Animal Span	.149	.203	.276	.270	.161	.308	.361	.269	.204	.218	.206	.134	.332	.317	.341	.384	.357	.394	.411	.348	.376	.420	.441	.410
Numerical N-back	.168	.252	.365	.350	.192	.328	.308	.242	.219	.230	.229	.267	.159	.280	.372	.422	.243	.312	.415	.378	.322	.382	.371	.421
Spatial MU	-.025	.096	.227	.306	-.035	.109	.190	.167	.120	.066	.121	.217	.207	.271	.351	.359	.244	.277	.300	.274	.232	.256	.249	.268

Note. The critical value for correlations being significant ($N = 101$; $\alpha = .05$; two-sided testing) is $\pm.196$.

Table 3

Model Fit for LCSMs of Common Factors for Drift Rates of the Three CRTs

Masking time	Model Fit				Comparison with configural model
	χ^2[df]	CFI	RMSEA	SRMR	$\Delta\chi^2$ [df]
12 ms	4.26 [7]	1.00	.00	.06	3.73 [4]
24 ms	16.30 [7]	.96	.12	.09	9.23 [4]
47 ms	21.20 [7]	.94	.14	.08	17.33 [4]*
94 ms	10.28 [7]	.98	.08	.05	8.12 [4]
all[a]	22.25 [7]	.94	.15	.09	16.22 [4]*

Note. Strong measurement invariance; correlated residuals for same tasks at pretest and

posttest, as well as for the Letters and Digits task at pretest and at posttest.

[a] Drift rate estimates for the different masking time conditions averaged before analysis.

* $p < .05$.

Table 4

Standardized Factor Loadings of Drift Rates in Measurement Models of LCSMs

Masking time	Letters		Digits		Figures	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
12ms	.381*	.499*	.448*	.597*	1.047*[a]	1.025*[a]
24ms	.360*	.452*	.515*	.568*	.911*	1.091*[a]
47ms	.541*	.619*	.552*	.652*	.763*	.901*
94ms	.664*	.659*	.727*	.790*	.628*	.836*
all	.462*	.553*	.563*	.662*	.841*	1.005*

Note. Strong measurement invariance; correlated residuals for same tasks at pretest and

posttest, as well as for the Letters and Digits task at pretest and at posttest.

* $p < .05$.

[a] Standardized factor loadings > 1.00 indicate that the residual variance is estimated slightly

negative. In these cases, it is likely that the true variance of the indicator tasks can almost

completely be explained by the common factor, so that the true residual variance is very

small, and error of estimation may lead to a negative variance estimate.

Table 5

Correlations of Latent Common Factors for Practiced WM Tasks and Drift Rates at Pretest

and Posttest

Masking time	Pretest		Posttest		Difference of Correlations
	r (95%-CI)	χ^2[df][b]	r (95%-CI)	χ^2[df][b]	$\Delta\chi^2$ [df][c]
12 ms	.303 (.072; .534)	6.54 [1]*	.437 (.232; .643)	15.52 [1]*	1.12 [1]
24 ms	.458 (.183; .734)	10.03 [1]*	.504 (.301; .701)	21.06 [1]*	0.11 [1]
47 ms	.566 (.328; .803)	15.87 [1]*	.630 (.445; .816)	26.32 [1]*	0.25 [1]
94 ms	.517 (.273; .760)	13.15 [1]*	.602 (.431; .774)	28.05 [1]*	0.40 [1]
all[a]	.522 (.272; .773)	13.36 [1]*	.566 (.376; .756)	24.03 [1]*	0.11 [1]

Note. Strong measurement invariance; correlated residuals for same tasks at pretest and

posttest, as well as for the Letters and Digits task at pretest and at posttest.

[a] Drift rate estimates for the different masking time conditions averaged before analysis.

[b] χ^2 difference of likelihood ratio test comparison with a model in which the correlation is

constrained to zero.

[c] χ^2 difference of likelihood ratio test comparison with a model in which the pre- and posttest

correlations are constrained to be equal.

* $p < .05$.

Table 6

Model Fit and Estimated Change-Change Correlations for Correlated LCSMs of Common

Factors for Practiced WM Tasks and Drift Rates

Masking time	Model Fit				r (95%-CI)	$\Delta\chi^2$ [df][b]
	χ^2[df]	CFI	RMSEA	SRMR		
12 ms	61.85 [48]	.97	.05	.08	.261 (-.028; .550)	3.08 [1]
24 ms	83.54 [48]	.94	.09	.11	.495 (.173; .816)	8.99 [1]*
47 ms	79.91 [48]	.94	.08	.09	.379 (-.007; .766)	3.55 [1]
94 ms	72.68 [48]	.95	.07	.07	.249 (-.095; .593)	1.97 [1]
all[a]	83.46 [48]	.94	.09	.10	.371 (.046; .696)	4.79 [1]*

Note. Strong measurement invariance; correlated residuals for same tasks at pretest and

posttest, as well as for the Letters and Digits task at pretest and at posttest.

[a] Drift rate estimates for the different masking time conditions averaged before analysis.

[b] χ^2 difference of likelihood ratio test comparison with a model in which the correlation is

constrained to zero.

* $p < .05$.

Table 7

Correlations of Latent Common Factors for WM Near Transfer Tasks and Drift Rates at

Pretest and Posttest

Masking time	Pretest		Posttest		Difference of Correlations
	r (95%-CI)	$\Delta\chi^2[df]^b$	r (95%-CI)	$\Delta\chi^2[df]^b$	$\Delta\chi^2 [df]^c$
12 ms	.330 (.072; .588)	6.41 [1]*	.521 (.282; .760)	17.00 [1]*	1.71 [1]
24 ms	.529 (.209; .848)	9.92 [1]*	.597 (.351; .843)	21.14 [1]*	0.17 [1]
47 ms	.659 (.412; .907)	18.31 [1]*	.807 (.605; 1.009)	33.59 [1]*	1.05 [1]
94 ms	.551 (.292; .810)	13.06 [1]*	.747 (.561; .932)	35.70 [1]*	NA[d]
all[a]	.592 (.316; .868)	13.89 [1]*	.714 (.500; .928)	28.98 [1]*	0.64 [1]

Note. Strong measurement invariance; correlated residuals for same tasks at pretest and

posttest, as well as for the Letters and Digits task at pretest and at posttest.

[a] Drift rate estimates for the different masking time conditions averaged before analysis.

[b] χ^2 difference of likelihood ratio test comparison with a model in which the correlation is

constrained to zero.

[c] χ^2 difference of likelihood ratio test comparison with a model in which the pre- and posttest

correlations are constrained to be equal.

[d] Constrained model could not be estimated.

* $p < .05$.

Table 8

Model Fit and Estimated Change-Change Correlations for LCSMs of Common Factors for

WM Near Transfer Tasks and Drift Rates

Masking time	Model Fit				r (95%-CI)	$\Delta\chi^2$ [df][b]
	χ^2[df]	CFI	RMSEA	SRMR		
12 ms	46.24 [48]	1.00	.00	.07	.127 (-.347; .601)	0.29 [1]
24 ms	67.62 [48]	.96	.06	.10	.452 (-.135; 1.039)	2.81 [1]
47 ms	74.80 [48]	.94	.07	.08	.499 (-.141; 1.140)	2.59 [1]
94 ms	53.96 [48]	.99	.04	.06	.141 (-.407; .690)	0.26 [1]
all[a]	70.52 [48]	.95	.07	.09	.378 (-.179; .936)	1.97 [1]

Note. Strong measurement invariance; correlated residuals for same tasks at pretest and

posttest, as well as for the Letters and Digits task at pretest and at posttest.

[a] Drift rate estimates for the different masking time conditions averaged before analysis.

[b] χ^2 difference of likelihood ratio test comparison with a model in which the correlation is

constrained to zero.

Figure 1

Digits

Odd　　　　　　　　Even

Letters

Consonants　　　　　　Vowels

Figures
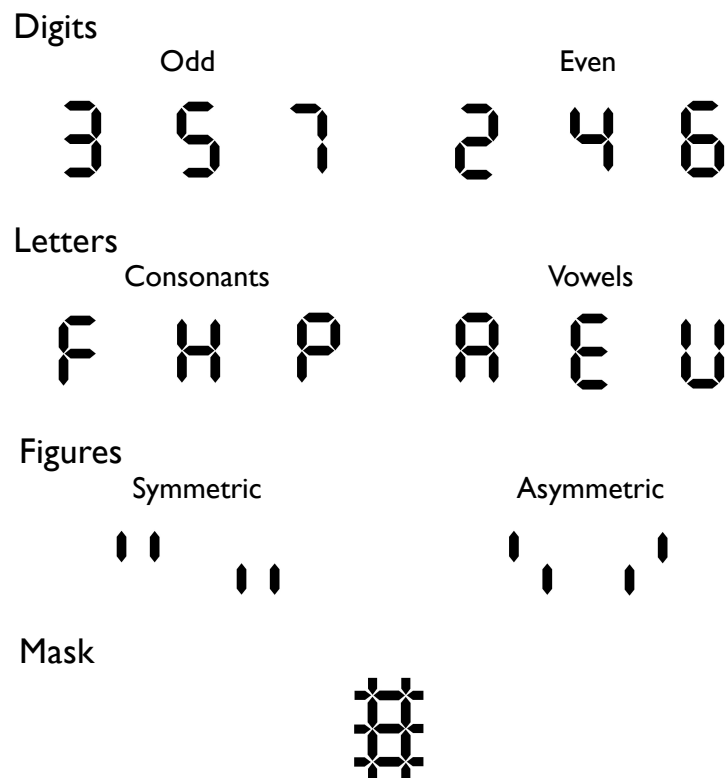
Symmetric　　　　　　Asymmetric

Mask

Figure 1. Stimuli sets (digits, letters, and figures) and masking stimulus used in the simple two-choice decision tasks.
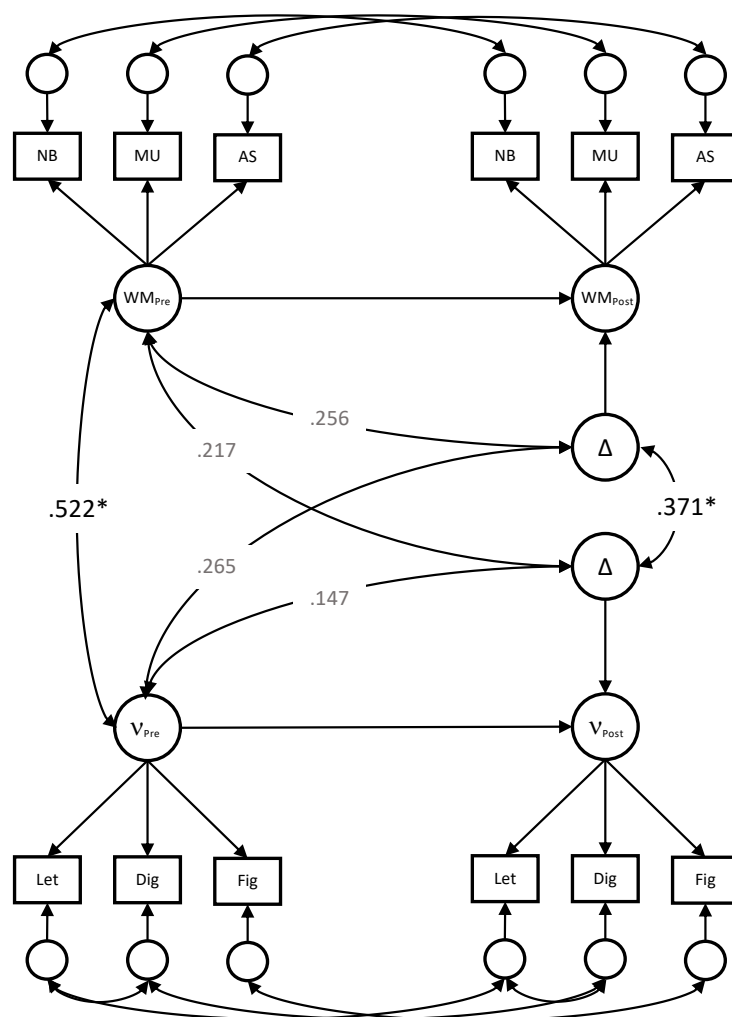
Figure 2

Figure 2. Bivariate latent change score model of drift rates (ν) and practiced working

memory (WM) tasks. Rectangles denote observed variables (composite scores of different

presentation/masking times). Circles denote latent variables (residual terms and common

factors). One-headed arrows denote factor loadings/regression parameters. Two-headed

arrows denote correlations. Let = letters; Dig = digits; Fig = figures; NB = N-back; MU =

Memory updating; AS = Alpha span; Pre = Pretest; Post = Posttest; Δ = latent change score.

* = $p < .05$. Parameter estimates are shown for the model with drift rates averaged across

masking time conditions. Non-significant factor correlations are shown in grey.