The Nature of Task Set Representations in Working Memory

Atsushi Kikumoto and Ulrich Mayr

Abstract

■ Selection and preparation of action plans (task sets) is often assumed to occur in working memory (WM). Yet, the absence of consistent evidence that WM capacity and task selection efficiency is correlated raises questions about the functional relationship between these two aspects of executive control. We used the EEG-derived contralateral delay activity (CDA) to index the WM load of task sets. In Experiment 1, we found a CDA set size effect (2 vs. 4 stimulus–response [S-R] rules) for high-WM, but not for low-WM, individuals when S-R sets were novel. In contrast, when only four task sets were presented throughout the experiment, we observed a sustained yet set size-independent use of WM for high-WM participants. Moreover, Experiment 2 showed an increase of the CDA in situations with task conflict, and this effect was larger the more that participants experienced RT conflict effects. Combined, these results indicate that even highly familiar S-R settings are maintained in WM, albeit in a compressed manner, presumably through cues to long-term memory representations. Finally, participants with low-WM capacity represented even familiar tasks in a load-dependent manner, suggesting that the establishment of effective retrieval structures itself is a capacity-limited process. ■

INTRODUCTION

How do we represent plans for upcoming actions and what are the constraints on such representations? A standard answer to this question is that working memory (WM) serves as both the critical holding device and bottleneck when it comes to selecting and maintaining plans for action. After all, the main function of WM is to preserve limited amounts of goal-relevant information in a privileged state until it can be acted on (Unsworth, Fukuda, Awh, & Vogel, 2014; Luck & Vogel, 2013; Woodman, Carlisle, & Reinhart, 2013; Ackerman, Beier, & Boyle, 2005; Engle & Kane, 2003).

Yet, evidence from the task-switching paradigm, a model situation to investigate how we select, change, and prepare action plans, does not provide strong evidence for this view. In this paradigm, participants are usually cued trial-by-trial to apply one of several stimulus–response (S-R) rules to an upcoming stimulus (Mayr, Kuhns, & Rieter, 2012; Kiesel et al., 2010; Monsell, 2003). Selection efficiency is reflected in the size of RT costs when the task set changes relative to the previous trial (local switch costs) and when more than one task is relevant within a block of trials, even in the absence of local changes (global or mixing costs). Both local and global costs are reliable. Yet, neither has shown consistent relationships with measures of WM capacity or with related measures, such as fluid intelligence (e.g., Friedman et al., 2006; Oberauer, Süß, Wilhelm, & Wittman, 2003).

One might be inclined to conclude from such results that standard WM is not involved in the selection of task sets. In fact, Oberauer, Souza, Druey, and Gade (2013) recently proposed that task set selection is handled by a separate WM system that represents "proceduralized" S-R rules. Because procedural WM does not share a common pool of representational resources with the standard, declarative WM system, this conceptualization has no difficulty in explaining the apparent independence between individual differences in standard WM tasks and task selection efficiency.

An alternative possibility is that S-R sets are represented within standard WM in a manner that allows these to slip through that system's capacity limitation. Specifically, to select or prepare S-R rules, it may be sufficient to maintain a reduced representation in WM that can serve as a cue for retrieving the entire set of rules from long-term memory (LTM), once needed. In standard task-switching situations, a limited number of tasks is used repeatedly throughout the experiment, which could allow building up LTM representations of S-R rules (Mayr, Kuhns, & Hubbard, 2014; Waszak, Hommel, & Allport, 2003).

The Contralateral Delay Activity

How can we distinguish between these two competing hypotheses about how WM is used during action control?

University of Oregon

Journal of Cognitive Neuroscience 29:11, pp. 1950–1961
doi:10.1162/jocn_a_01173

WM readily reveals its functionality when capacity limitations are overloaded, such as in terms of performance effects of a secondary WM task or of individual differences in WM capacity. It is much more difficult to demonstrate subcapacity representational functions. However, Vogel and Machizawa (2004; Luria, Balaban, Awh, & Vogel, 2016) introduced an ERP, the contralateral delay activity (CDA), that allows online assessment of the amount of information contained in visual WM. The CDA is observed as a sustained negativity contralateral to the side of visual fields where participants are attending during the delay period. Importantly, the CDA's amplitude is sensitive to the amount of information represented in WM (i.e., set size effects), and it reaches a plateau at individuals' capacity limits (Vogel & Machizawa, 2004). We use here the CDA to examine the representational load of sets of S-R mappings, with the goal of determining the role of WM for the selection of action plans.

EXPERIMENT 1

In our paradigm, we married the change detection procedure, used to elicit the CDA, with a cued task-switching procedure. On each trial, participants were presented two lateralized sets of either two or four color stimuli with a central arrow cuing which sides had to be attended on a given trial. The locations of the cue color stimuli indicated spatially compatible response keys. Thus, the arrangement of color stimuli served as cues of S-R rules that had to be applied to an individual color stimulus (i.e., the probe stimulus) presented after a delay interval (Figure 1).

Half of the participants (n = 36) worked with novel constellations of S-R rules on 50% of the trials (change trials) and repetitions of the previous trial constellation on the remaining trials (repeat trials). We refer to this as the unconstrained condition. Given that it requires frequent updating of WM with new sets of rules, we expected that WM is required in a load-dependent manner (at least on change trials). Thus, we predict here a CDA set size effect with a larger amplitude for sets of four than sets of two S-R rules. Based on previous results, we also expect that CDA set size effects are apparent in individuals with high-WM capacity (as measured in a standard change detection task), but to be small or nonexistent in low-WM individuals (Vogel & Machizawa, 2004).

The other half of participants (n = 36) worked with only two sets of S-R rules per set size across the entire experiment (i.e., four tasks total). This constrained condition resembles the standard task-switching paradigm where typically only a small set of tasks is used throughout an experimental session. Here, participants should be able to represent task sets either in a proceduralized format (according to the two-system model) or in a "compressed" manner (according to the one-system model). If practiced S-R rules are in fact represented in a proceduralized format, then we should not see a sustained CDA, reflecting the fact that standard WM is not utilized (see Figure 2). If, however, practiced task sets can be represented in standard WM in terms of cues to LTM representations, we expect to see a CDA, but no set size effect (Figure 2). Finally, these predictions about the contrast between the unconstrained and constrained task conditions can be appropriately evaluated only in participants who actually exhibit set size effects for the unconstrained condition. Thus, for our main hypothesis, we focus on participants with high-WM capacity, who in past work have exhibited set size effects with novel WM content. We expect no CDA set size effect in the unconstrained condition for low-WM individuals (e.g., Vogel & Machizawa, 2004). However, this group will allow us to examine in which way low-WM capacity constrains task set representations even after additional experience (i.e., through immediate task repetitions in the constrained condition or through frequent repetitions in the unconstrained condition).

Finally, our pilot studies had shown that the hybrid WM/task-switching paradigm produces relatively small, local RT switch costs. This is consistent with findings indicating that, with long intervals between consecutive

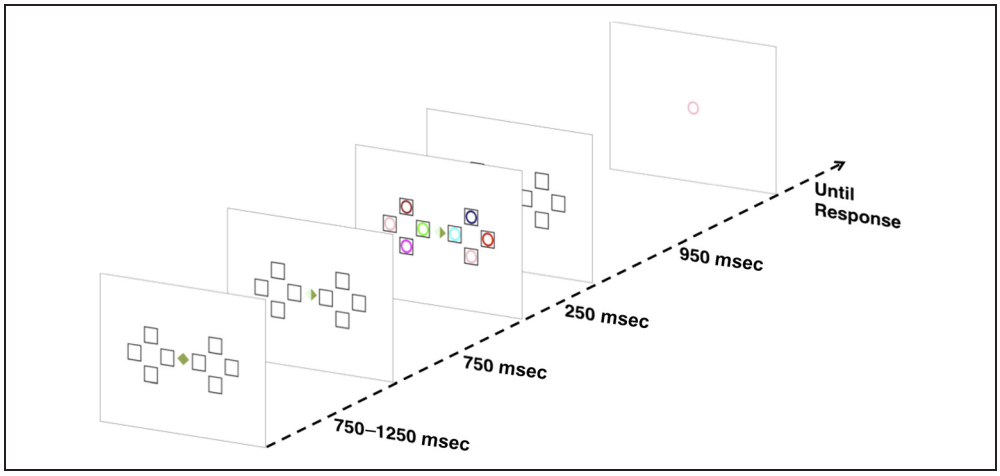

Figure 1. Trial timeline and sample stimulus displays for Experiment 1. Participants attended the cued side (indicated through dark green triangle) to encode color positions as S-R rules to execute a response to the final probe display. For example, in the presented trial, the correct response to the final probe display was to the bottom key, corresponding to the bottom location of the pink color patch in the cued, right-side encoding display. The relevant EEG signal was measured during both the encoding and retention intervals.

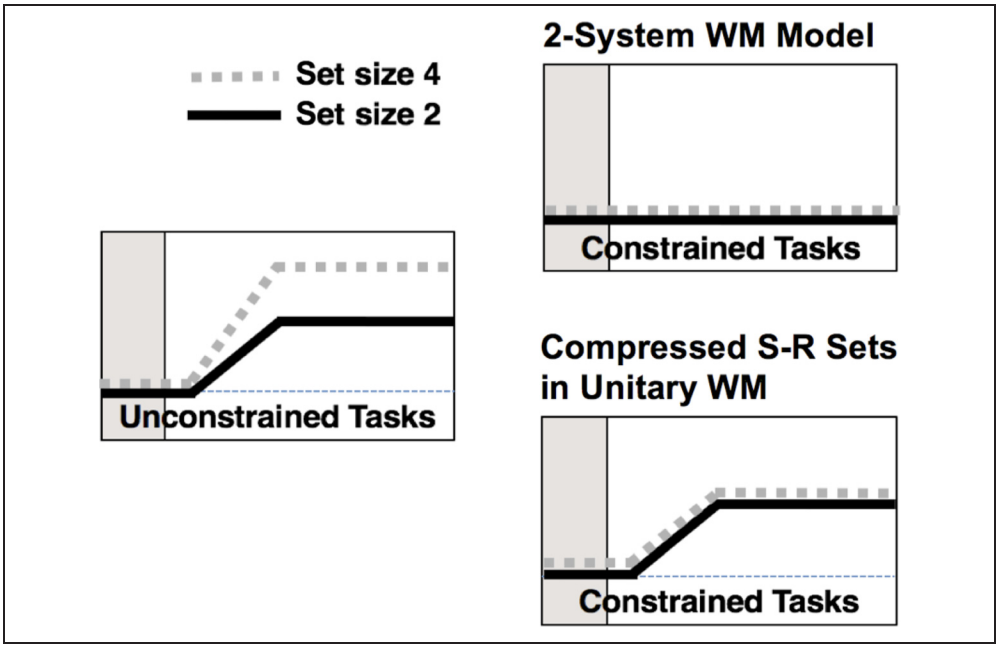

Figure 2. Predicted CDA for the unconstrained and constrained task conditions. For the unconstrained condition, both the two-system model and the unitary model make identical predictions. For the constrained condition, the two-system model predicts the elimination of the CDA—indicating that standard WM is no longer involved. The unitary WM + LTM model predicts that the CDA set size effect is eliminated, but that a sustained CDA remains—indicating that WM remains critical for cue-based preparation and selection of task-relevant information.

trials, participants often do not maintain task sets across trials, thus reducing or eliminating the between-switch and no-switch trials (Bryck & Mayr, 2008). Nevertheless, because this is an important variable in the context of task-switching research, we included an experimental contrast between change and repeat transitions in an exploratory manner.

Methods

Participants

Eighty-one neurologically normal individuals participated in exchange for course credit or monetary payment and after giving informed consent. The study protocol was approved by the Human Subjects Committee at the University of Oregon. Participants were randomly assigned to the unconstrained task condition or constrained task condition. Seven participants were excluded from the final sample because of extensive trial rejection rates of >30%, resulting in too few trials for adequate ERP analysis. The final sample comprised data of 36 participants for each of the two between-subject conditions.

Tasks, Stimuli, and Procedure

For each trial, S-R rules were coded by color patches (i.e., two or four patches, radius = 2°) appearing inside of four square placeholders (width and height = 6.5°) that were organized in terms of a cross. Color patches were randomly selected from 10 possible isoluminant colors (green[0,255,0], red[255,0,0], magenta[255,0,255], cyan [0,255,255], yellow[255,255,0], gray[128,128,128], brown[153,76,0], light pink[255,102,178], midnight blue [0,76,153], orange[255,153,51]). Each placeholder was mapped onto a spatially compatible response key (top, right, bottom, and left key). Within a given block, participants were exposed to set sizes of either two or four color patches (i.e., S-R rules). Set size 2 and set size 4 blocks alternated throughout the session, and the first block set size was counterbalanced across participants. On repeat trials, the same constellation of color patches was used as on the preceding trial. On change trial, half of the color patches were the same as on the previous trial, but on a different location, and the other half were in colors not used on the previous trial. For both the unconstrained and constrained conditions, the probability of a change trial was $p = .5$.

In the unconstrained condition, new task set constellations of color patches were generated for each change trial with the above-mentioned constraints. For the constrained condition, two task set constellations were randomly selected for each set size so that, on change trials, task sets changed back and forth between these two constellations. The experimental session included 22 blocks of 40 trials each; the first two blocks were practice blocks (one block for each set size). In half of the randomly intermixed blocks, S-R sets with set size 2 were presented; in the other half, S-R sets with set size of 4 were used.

Throughout each trial, participants were asked to fixate the central diamond (width = 1°), and the viewing distance was kept approximately 77 cm. Placeholders and color patches were displayed bilaterally. The center of each side's stimulus array was 7.2° to the left or the right of the center of the screen. During an initial, jittered intertrial interval (750–1250 msec), bilateral placeholders and the central diamond were presented, after which either the left or the right side of the diamond turned green for 750 msec to cue participants to attend to either to the left side or the right side of the display. Then, two

or four color patches appeared for 250 msec in both the attended and irrelevant side. The CDA was measured from the onset of color patches to the end of following 950 msec of blank retention period. For the probe display, one of the colors that was part of the encoded task set appeared at the center. Participants were instructed to press on the number pad (8[top], 6[right], 2[bottom], and 4[left]) the key associated with that color's location as quickly and accurately as possible. Participants were asked to switch between left and right hand in the middle of the session, with the starting had counterbalanced across participants. All stimuli were generated using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) for MATLAB (The MathWorks, Natick, MA) on a 17-in. flat cathode ray tube monitor.

EEG Procedures and Analyses

EEG activity was recorded from 22 tin electrodes held in place by an elastic cap (Electro-Cap International, Eaton, OH) using the International 10/20 system. The 10/20 sites F3, Fz, F4, T3, C3, CZ, C4, T4, P3, PZ, P4, T5, T6, O1, and O2 were used along with five nonstandard sites: OL midway between T5 and O1; OR midway between T6 and O2; PO3 midway between P3 and OL; PO4 midway between P4 and OR; and POz midway between PO3 and PO4. The left mastoid was used as reference for all recording sites. Data were re-referenced offline to the average of all scalp electrodes. Electrodes placed ~1 cm to the left and right of the external canthi of each eye recorded the horizontal EOG to measure horizontal saccades. To detect blinks, the vertical EOG was recorded from an electrode placed beneath the left eye and reference to the left mastoid. The EEG and EOG were amplified with an SA Instrumentation amplifier with a bandpass of 0.01–30[80] Hz and were digitized at 250 Hz in LabView 6.1 running on a PC.

The EEG signal was segmented into 1400-msec epochs, each starting 200 msec before the onset of the task cues. The 200-msec interval preceding the onset of the task set served as baseline. We used standard analysis procedures, including exclusion of trials containing blinks, large eye movements (>1°), and blocking of signals (McCollough, Machizawa, & Vogel, 2007; Vogel, Luck, & Shapiro, 1998). We used the EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes and custom scripts for EEG and ERP processing. The CDA was measured by computing the difference between contralateral and ipsilateral activity in reference to the side of visual field where participants were cued to attend. The difference waves were measured and averaged among parietal (P3/P4), posterior parietal (PO3/PO4), and lateral occipital (OL/OR) electrode sites. CDA amplitudes were compared across the main design factors using an ANOVA for the interval from 450 and 1200 msec following the onset of the task cue display. This interval was predetermined based on previous studies showing CDA typically start appearing at 300–450 msec after the onset of items (Luria et al., 2016; Ikkai, McCollough, & Vogel, 2010).

WM Assessment

Individuals' visual WM capacity was measured using a standard change detection paradigm (Vogel & Machizawa, 2004; Luck & Vogel, 1997). In this task, participants were presented, during a 100-msec encoding interval, arrays of four or eight colored squares (0.65° × 0.65°) at random locations. After a delay period of 900 msec, a single probe at the location of one of the objects of the encoding set was presented. Participants had to indicate whether or not the color of the probe matched the color of the original object. The number of trials was 320. The capacity was computed as $K = S \times (H - F)$, where K is the memory capacity, S is the size of the array, H is the observed hit rate, and F is the false alarm rate (Cowan, 2001).

Results

Behavioral Results

RTs from error trials, posterror trials, and trials in which RTs were shorter than 100 msec or longer than 5000 msec were excluded from analyses, which excluded 0.4% of data. Table 1 shows RTs and accuracy as a function of task set context (unconstrained vs. constrained) as a between-subject variable, set size (2 vs. 4), and switch (change vs. repeat). RTs were larger and accuracy was lower for set size 4 versus set size 2, $F(1, 70) = 40.26$, $p < .001$, $\eta^2 = .36$, and $F(1, 70) = 148.34$, $p < .001$,

Table 1. Mean (*SD*) RTs and Accuracy Rates as a Function of Unconstrained Task versus Constrained Task Conditions, Set Size, and Change versus Repeat Trials

Unconstrained Tasks				*Constrained Tasks*			
Set Size 2		*Set Size 4*		*Set Size 2*		*Set Size 4*	
Change	*Repeat*	*Change*	*Repeat*	*Change*	*Repeat*	*Change*	*Repeat*
583 msec (106)	562 msec (96)	750 msec (123)	740 msec (112)	611 msec (116)	588 msec (107)	758 msec (133)	739 msec (126)
6.58% (2.67)	96.58% (3.10)	88.26% (7.47)	91.60% (7.16)	96.20% (2.78)	97.81% (2.16)	86.40% (5.46)	80.97% (5.61)

$\eta^2 = .69$, and for change versus repeat trials, $F(1, 70) = 511.81$, $p < .001$, $\eta^2 = .88$, and $F(1, 70) = 261.21$, $p < .001$, $\eta^2 = .79$. Also, in the constrained task set condition, participants were overall more accurate, $F(1, 70) = 10.84$, $p < .001$, $\eta^2 = .13$, but not significantly faster than in the unconstrained condition, $F(1, 70) = 0.52$, $p = .47$. In addition, performance benefits in the constrained compared with the unconstrained condition were greater for set size 4 than set size 2 for accuracy, $F(1, 70) = 10.84$, $p < .001$, $\eta^2 = .13$, but there was no significant effect for RTs, $F(1, 70) = 0.91$, $p = .34$.

To examine how WM capacity is related to task selection performance, we also correlated overall RTs and accuracy for the unconstrained and constrained task conditions. The correlation between WM capacity and RTs was $r(35) = -.56$, $p < .001$, for the unconstrained condition and was reduced to $r(35) = -.002$, $p = .99$, for the constrained condition. This difference between correlations was highly significant, $z = -2.57$, $p < .01$. For accuracy, the correlation for the unconstrained condition was also higher than for the constrained condition, unconstrained: $r(35) = .63$, $p < .01$, constrained: $r(35) = .42$, $p < .01$, even though here the difference was not significant, $z = 1.01$, $p = .14$. These results suggest that individuals' WM capacity limits performance when novel task sets need to be used, but is less of a constraint when the same task sets are used repeatedly—as is the case in standard task-switching situations. Finally, for both the unconstrained and constrained conditions, trial-to-trial switch costs were generally small and showed no reliable correlations with WM capacity, an aspect we will return to in the Discussion.

CDA for High-WM Individuals

Our primary question refers to how the CDA set size effect, as an indicator of WM load, behaves in the unconstrained and constrained task conditions. Reliable CDA set size effects are usually only observed in individuals with high-WM capacity (Vogel & Machizawa, 2004). Therefore, we used individuals' WM capacity scores to conduct a median split, separately within the unconstrained (median = 2.72) and constrained conditions (median = 2.58). The CDA difference waves across all conditions are presented in Figure 3 and Table 2. As mentioned above, we analyzed CDA effects by averaging across the time interval between 450 and 1200 msec. In an ANOVA, we obtained a reliable interaction between all experimental factors (i.e., WM capacity, unconstrained/constrained, change/repeat, set size), $F(1, 68) = 6.31$, $p = .014$, $\eta^2 = .09$. Beyond our theoretical motivation, this effect also statistically justifies a separate focus on high-WM and low-WM individuals.

As expected, in the group of high-WM individuals we found in the unconstrained condition a reliable set size effect, $F(1, 17) = 18.63$, $p < .001$, $\eta^2 = .52$ (Figure 3). Although this effect seemed somewhat reduced in the repeat condition, the change/repeat by set size interaction was not reliable, $F(1, 17) = 1.83$, $p = .19$, $\eta^2 = .10$. Most importantly, if familiar task sets require less WM capacity than novel task sets, we should expect a reduced set size effect in the constrained task condition. In fact, in this condition there was no reliable set size effect, $F(1, 17) = 0.56$, $p = .46$, $\eta^2 = .03$; the theoretically critical set size by unconstrained/constrained interaction was also reliable, $F(1, 17) = 7.45$, $p = .01$, $\eta^2 = .18$ (Figure 3).

Turning to the question of whether or not participants were relying on WM even in the constrained condition, we need to look at the absolute CDA in that condition. The amplitude of the CDA remained significantly negative for both set size conditions throughout the preparation interval, $F(1, 17) \geq 6.49$, $p = .02$. The CDA for both set sizes in the constrained condition was about the same size as for set size 2 in the unconstrained condition. Thus, our results for high-WM participants are consistent with the hypothesis that, although novel task sets require WM capacity in a load-dependent manner, familiar task sets are still represented in WM (i.e., the sustained CDA in the constrained condition), however, in a load-independent manner (i.e., the eliminated set size effect). This pattern is most consistent with the hypothesis that, even for familiar task sets, WM remains involved in representing cues or pointers to representations of complete sets of S-R associations in LTM.

CDA for Low-WM Individuals

Generally, our predictions for low-WM individuals were less clear than for high-WM individuals. However, based on past work (Vogel & Machizawa, 2004) we did expect either small or nonexistent set size effects in the unconstrained task/change condition, which most closely resembles the typical WM situation. In fact, low-WM participants showed no set size effects here, $F(1, 17) = 0.06$, $p = .82$, $\eta^2 = .003$ (Figure 3). Somewhat surprisingly, we did find a reliable set size effect on repeat trials, $F(1, 17) = 8.73$, $p = .009$, $\eta^2 = .34$; the interaction between set size and the change/repeat factor was also reliable, $F(1, 17) = 4.83$, $p = .04$, $\eta^2 = .22$. In addition, a reliable set size effect also emerged in the constrained/change condition, $F(1, 17) = 5.69$, $p = .029$, $\eta^2 = .25$, albeit with a nonsignificant interaction between the unconstrained/constrained and set size factors, $F(1, 17) = 3.31$, $p = .15$, $\eta^2 = .06$. The small set size effect found in constrained task/change trials was all but eliminated for constrained task/repeat trials, $F(1, 17) = 0.97$, $p = .34$, $\eta^2 = .05$. Thus here, even low-WM participants seemed to have reached a stage of relatively load-independent task set selection/preparation.

Figure 4 summarizes the overall set size effects in each of the experimental conditions. Apparently, task familiarity, whether due to immediate repetitions in the unconstrained condition or due to practice within a session (i.e., in the constrained condition), had opposite effects

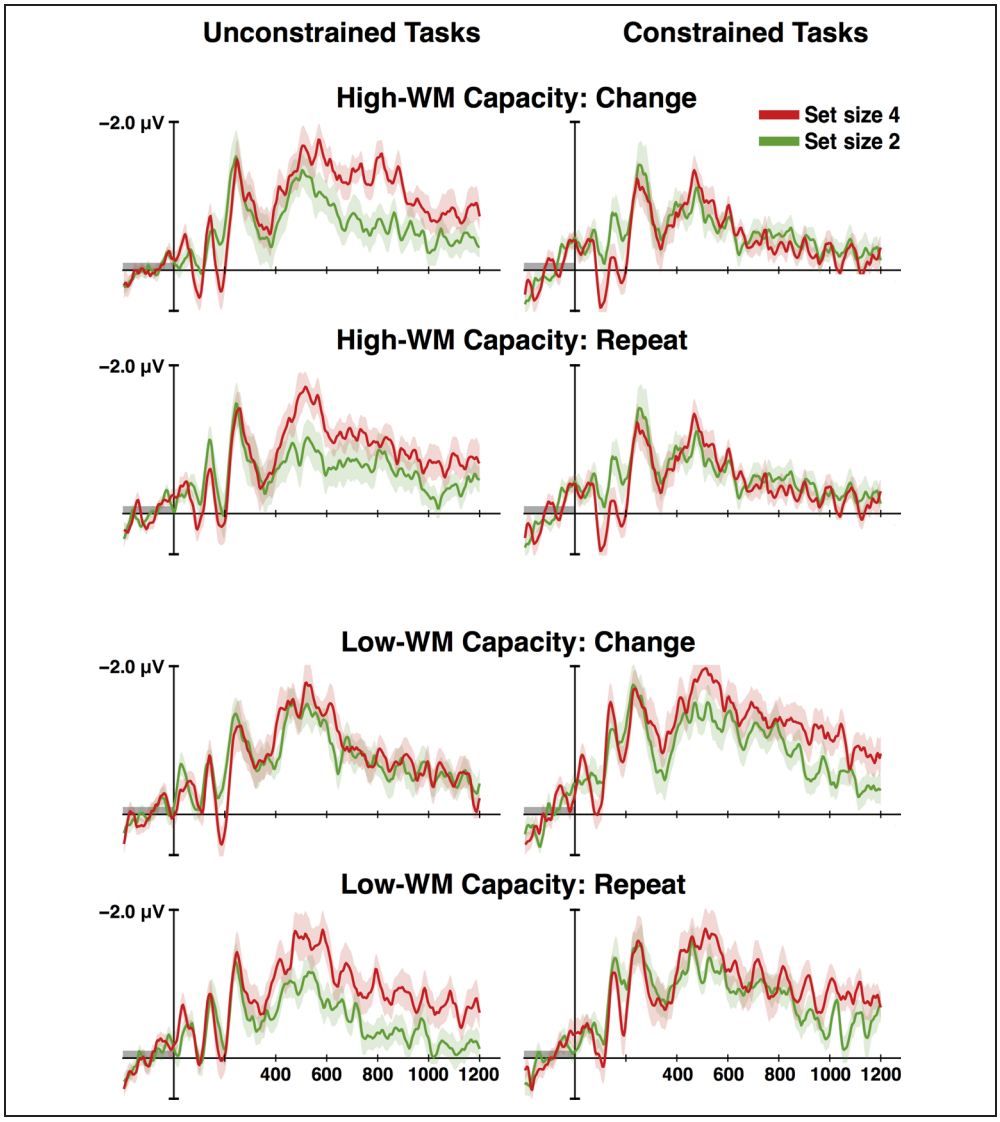

Figure 3. CDA (P3/P4, PO3/PO4, OL/OR) for set sizes 2 and 4, as a function of unconstrained versus constrained task conditions, high- versus low-WM capacity, and change versus repeat trials. Error bars indicate standard errors for each individual set size effect.

depending on WM capacity. In high-WM participants, we found the predicted pattern of a reduction of set size effects (albeit not significantly for immediate repetitions). In contrast, in low-WM participants, set size effects were present only after previous exposure with an S-R mapping, either on repetition trials in the unconstrained condition or on change trials in the constrained task condition. As elaborated further in the Discussion, this

Table 2. Mean (*SD*) CDA Amplitudes (450–1250 msec) as a Function of WM Group, Unconstrained Task versus Constrained Task Conditions, Set Size, and Change versus Repeat Trials

Unconstrained Tasks				*Constrained Tasks*			
Set Size 2		*Set Size 4*		*Set Size 2*		*Set Size 4*	
Change	*Repeat*	*Change*	*Repeat*	*Change*	*Repeat*	*Change*	*Repeat*
High-WM Capacity							
−0.52 µV (0.87)	−0.49 µV (0.91)	−1.02 µV (0.84)	−0.81 µV (0.82)	−0.55 µV (0.69)	−0.43 µV (0.48)	−0.57 µV (0.51)	−0.53 µV (0.57)
Low-WM Capacity							
−0.58 µV (0.66)	−0.36 µV (0.60)	−0.61 µV (0.48)	−0.86 µV (0.87)	−0.66 µV (0.52)	−0.67 µV (0.5)	−0.97 µV (0.54)	−0.86 µV (0.69)

pattern suggests that the absent set size effect for novel task sets (i.e., in the unconstrained/change condition) not necessarily means that low-WM individuals have less actual storage capacity. Rather, they seem to have greater difficulty in encoding novel material into the existing storage space.

EXPERIMENT 2

From the results of Experiment 1 alone, it is not clear to what degree the CDA amplitude for well-practiced tasks actually reflects representations or processes involved in the selection of S-R sets. For example, it is known that the CDA is reduced when multiple stimuli can be grouped into chunks (Balaban & Luria, 2015). Thus, perhaps the variations in the CDA reflect relatively superficial perceptual effects, with no implications for task set control. To provide evidence that the CDA in the constrained task condition actually does index task selection processes, it would be important to show that it is responsive to key variables reflecting task selection demands. One such variable is the degree of task set competition/conflict. Conflict is present (a) when more than one S-R set can be applied within the same context and (b) when the stimulus alone does not allow disambiguating between competing S-R sets.

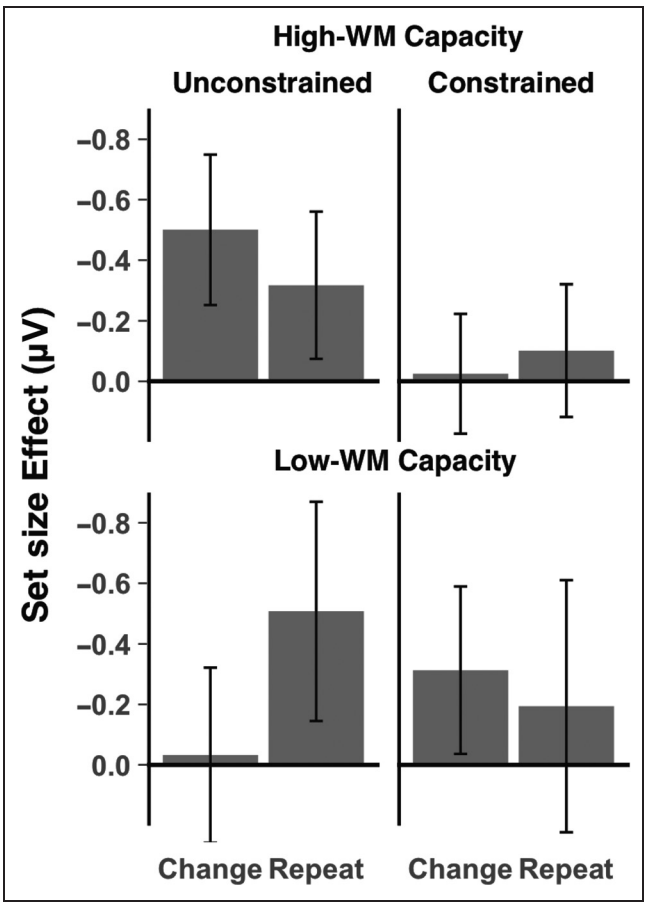

Figure 4. Aggregate set size effect (450–1200 msec) as a function of high- versus low-WM capacity, unconstrained task versus constrained task conditions, and change versus repeat trial. Whereas for high-WM individuals, both trial-to-trial and within-session repetitions reduced the set size effect; for low-WM individuals, both within-trial and within-session repetitions led to an increase in set size effects. Error bars indicate within-subject 95% confidence intervals for each individual set size effect.

We used the same basic paradigm as in Experiment 1, albeit focusing only on the constrained task situation with a constant set size of 3. Two conditions were contrasted in a blocked manner. In mixed-set blocks (i.e., high conflict), one of two S-R sets were randomly cued on a trial-by-trial basis. In the low-conflict, single-set blocks, participants had to work with only one S-R set. In standard task-switching situations, this manipulation leads to RT effects known as mixing or global costs (Rubin & Meiran, 2005; Mayr, 2001). We predicted that the CDA would indicate a larger WM load in the high-competition condition than in the low-competition condition, reflecting the fact that WM representations are in fact used to counteract bottom–up, stimulus-elicited task set conflict.

In principle, S-R cues are not necessary for accurate performance in single-set blocks, and therefore, a small or absent CDA in that condition might simply indicate that participants ignored the S-R cues. Therefore, instead of a central stimulus probe (as in Experiment 1), we presented both a relevant probe on the same side as the relevant cue and an irrelevant probe on the opposite side (as is usually done in studies assessing the CDA in the context of visual WM tasks; Vogel & Machizawa, 2004). Thus, participants had to attend to the S-R cues to know which of the two probes to respond to. In addition, the S-R sets for the task selection condition were constructed such that one of the individual S-R links shared the stimulus across sets whereas the other two stimuli were not shared across sets. This allowed us to measure the degree of response conflict in RTs and accuracy elicited by the probe stimulus, similar to response congruency effects in the standard task-switching paradigm.

Methods

Thirty participants were recruited from the same population as in Experiment 1; two participants were excluded based on the same rejection criteria as in Experiment 1. Stimulus material and procedures were identical to Experiment 1, except for the following changes: In Experiment 2, all participants were exposed to a constrained task condition with a total of three S-R sets, each having a set size of 3. Half of the twenty 40-trial blocks used only a single S-R set (the single-task condition); in the other half of blocks, two S-R sets were presented (as in Experiment 1). The two block types were randomly intermixed. The color patches that served as task cues appeared for 100 msec on both the attended and irrelevant side, followed by a 1100-msec retention interval. Instead of the centralized probe stimulus in Experiment 1, two different probe stimuli appeared in Experiment 2; one on the cued

Table 3. Mean RTs (*SD*), Accuracy Rates (*SD*) for Single-task and Mixed-task Blocks, as well as for Change versus Repeat and No-conflict versus Conflict Trials within the Mixed-task Blocks

	Mixed-task Block			
	Change		*Repeat*	
Single-task Block	*No Conflict*	*Conflict*	*No Conflict*	*Conflict*
715 msec (127)	718 msec (128)	773 msec (146)	709 msec (107)	743 msec (128)
97.42% (2.67)	95.75% (3.21)	91.66% (6.99)	96.83% (3.21)	94.77% (6.28)

side (i.e., relevant probe) and the other at the uncued side (i.e., irrelevant probe). Finally, S-R sets were constructed such that the set used in the single-task block had no overlap in terms of stimuli with any of the two mixed-set blocks. However, for the two mixed-task sets, one of the S-R links overlapped in terms of the stimulus but was associated with different responses (i.e., the high-conflict stimulus).

Results

Behavioral Results

Trials were filtered in the same manner as in Experiment 1. Table 3 shows RTs and accuracy separately for both the single-task and mixed-task conditions, and as a function of the change-repeat and the conflict factor within the mixed-task condition. Inspection of the table indicates that the mixed-task condition with no conflict and task repetitions—which is most similar to the single-task condition—shows very little indication of the selection costs when compared with the single-task condition. Simple t tests indicated nonsignificant differences for RTs and accuracy, $t(27) < 1.0$, $p > .36$. However, mixed-task performance as a function of the switch/repeat and the conflict factor yielded for RTs highly reliable effects for both factors, change/repeat: $F(1, 27) = 11.75$, $p = .002$, $\eta^2 = .30$, conflict: $F(1, 27) = 14.62$, $p = .001$, $\eta^2 = .35$, with no reliable interaction, $F(1, 27) = 2.31$, $p = .14$, $\eta^2 = .08$. For accuracy, both the two main effects, change/repeat: $F(1, 27) = 12.31$, $p = .002$, $\eta^2 = .31$, conflict: $F(1, 27) = 14.24$, $p = .001$, $\eta^2 = .35$, and the interaction was reliable, $F(1, 27) = 10.35$, $p = .003$, $\eta^2 = .28$, indicating particularly large conflict effects for change trials.

We also examined the relationship between WM capacity and response selection conflict. High-WM individuals tended to show smaller RT, $r(27) = -.47$, $p = .01$, and accuracy conflict effects, $r(27) = -.37$, $p = .05$. As in Experiment 1, relationships between change-repeat effects and WM capacity exhibited no consistent pattern for RTs, $r(27) = .19$, $p = .33$, and accuracy, $r(27) = -.27$, $p = .17$. Overall, these results indicate that the mixed-task condition induced response selection conflict between the competing task sets and that the degree of conflict was modulated through WM capacity.

CDA Effects

Figure 5 shows the CDA for the single-task condition, as well as the mixed-task, repeat and change conditions. The CDA was present across all conditions, even in single-task blocks. Compared with the single-task condition ($-.74$ μV, $SD = .78$), the CDA was larger for the two mixed-task conditions (repeat: -1.06 μV, $SD = .93$, change $= -1.21$, $S = .92$), but with little difference between these two conditions. To analyze the condition effects, we specified two orthogonal contrasts, the first comparing single task with the two mixed conditions and the second the change and the repeat condition. Whereas the first contrast was highly reliable, $F(1, 27) = 17.68$, $p < .001$, $\eta^2 = .40$, the second was not, $F(1, 27) = 1.60$, $p = .217$. Thus, as expected, the mixed-task context elicited a stronger CDA, presumably because in this condition a greater amount of WM capacity was used for task selection/preparation than in the single-task condition.[1]

Finally, we examined to what degree the CDA difference between the single-task and mixed conditions (i.e., CDA mixing costs) was related to WM capacity and the conflict effects in the mixed-task condition. We found a nonreliable tendency for individuals with small WM capacity to show larger CDA mixing costs, $r = -.32$, $p = .1$. We also found that the larger the CDA mixing costs, the larger tended to be the accuracy conflict effects, $r = .40$, $p = .03$; the effect for RTs was in the

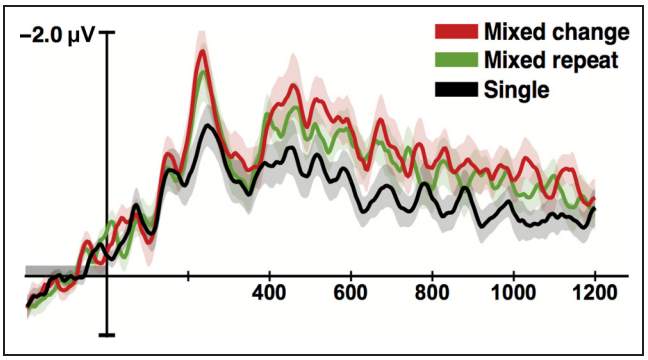

Figure 5. CDA (P3/P4, PO3/PO4, OL/OR) for the single-task condition, as well as for repeat and change trials from the mixed-task condition. Error bars indicate standard errors for each individual set size effect.

same direction, though not reliable, $r = .16$, $p = .43$. As noted above, there was also a reliable inverse relationship between WM capacity and accuracy conflict effects, as well as a nonreliable relationship in the same direction with RT conflict effects. Thus, albeit modest in size, the observed relationships consistently indicate that lower WM capacity and larger conflict effects go along with an increased CDA.

DISCUSSION

We used the CDA, an ERP waveform that gauges the load of information in visual WM, to assess how task sets are represented during task selection. In Experiment 1, when completely new task sets had to be prepared (i.e., in the unconstrained task condition), the CDA amplitude scaled with the number of S-R rules. Consistent with previous results using standard visual WM tasks (Vogel & Machizawa, 2004), this CDA set size effect was very robust for individuals with high, but nonexistent for individuals with low-WM capacity. Moreover, WM capacity also predicted RTs and, to a lesser degree, also accuracy during task selection.

These results confirm that the CDA can be used as an index of representational load of S-R mappings in WM. Next, we investigated how the representational load changed when a small number of familiar tasks were used consistently, much like in a regular task-switching situation. It is possible that here, standard WM may no longer be relevant, for example, because well-practiced task sets are passed on to a procedural memory system (e.g., Oberauer et al., 2013; see Figure 2). In this case, we should expect the CDA as an indicator of standard WM load, to be absent for all set sizes. However, it is also possible that task sets continue to be actively represented in WM, albeit in a condensed, set size-independent manner. In this case, we would expect a robust CDA, however, no longer a set size modulation.

In fact, in terms of behavioral effects we found that, in our constrained task set condition, WM capacity was less of limiting factor than in the unconstrained condition: RTs were no longer predicted by WM capacity; for accuracy, the effect was in the same direction, though not reliable. This pattern is still consistent with either the dual-system or unitary-system account. However, we also found that the number of S-R rules no longer modulated the CDA amplitude for individuals with high-WM capacity but at the same time stayed robustly negative for both set sizes. This pattern supports the unitary model, where practiced S-R sets continue to be represented in standard WM, albeit in a load-independent manner (Figure 2).

Experiment 1 confirmed that the CDA responds to standard manipulations of WM load (i.e., the set size effect and its reduction with practice). In Experiment 2, we wanted to confirm that the CDA is also sensitive to the degree of task set conflict. We found that the CDA was larger in the high-conflict, mixed-task than in the low-conflict, single-task condition. This pattern indicates that the representation of even highly familiar S-R sets requires WM capacity to counteract competition between S-R sets. Furthermore, individuals with larger competition effects tended to have both lower WM capacity and a larger CDA mixing cost.

Interplay between WM and LTM during Task Selection

Combined, our results indicate that even for familiar tasks, S-R rules are not represented in some kind of special, proceduralized format (Oberauer et al., 2013). Rather, they suggest that representations of S-R rules consist of declarative knowledge about how to respond (e.g., at which key location) to a particular environmental event (e.g., a particular stimulus color). However, this leaves the question how WM and LTM work together to reduce the demands on active WM representations through experience—as indexed through the eliminated CDA set size effect for familiar tasks (in high-WM participants).

On an abstract level, Ericsson and Kintsch (1995) have suggested in their "long-term WM" model that LTM knowledge structures can provide rapid access to relevant information so that only appropriate retrieval cues need to be represented in WM. At this point performance is no longer constrained by the fixed capacity of WM. More concretely, several studies using visual WM tasks have demonstrated that statistical regularities in the stimulus material can substantially increase WM capacity (Brady, Konkle, & Alvarez, 2009; Cowan, 2001) and reduce the CDA (e.g., Balaban & Luria, 2016). Applying this to the current situation, statistical regularities may allow participants to rely on actively represented parts of the entire task set to retrieve the remaining information from LTM. To be specific, in our paradigm it is plausible that for familiar sets of color–location pairings, participants activated only one of the individual pairs in preparation for the upcoming stimulus (e.g., Lien, Ruthruff, Remington, & Johnston, 2005). Associative links between the activated color–location pair and the remaining color–location pairs in the task set would be sufficient to complete the task once the stimulus appears. For example, this could be achieved through S-S associations between co-occurring colors. Obviously, with such a retrieval strategy, the WM load during the preparatory phase would be independent of the size of the task set.

On the basis of both the current results and previous reports in the literature, it is clear that experience can reduce WM load, as reflected in the CDA. However, it is also an important question whether or not there are situations in which WM representations become completely obsolete. Interestingly, Carlisle, Arita, Pardo, and Woodman (2011) showed in the context of a visual search task that when only a single search target was used throughout the experimental session, the CDA was completely eliminated. In contrast, in the single-task situation used in

our Experiment 2, a robust CDA remained, albeit smaller than in the mixed-task condition. In principle, this overall pattern could imply that, with sufficient practice and/or sufficiently simple task sets, declarative representations can transform into proceduralized representations that bypass standard WM. However, we believe that a more plausible interpretation of this set of results is that when there is very little conflict (e.g., a single target throughout the experimental session as in Carlisle et al., 2011) participants can ignore the task cue and simply rely on the unambiguous stimulus information to trigger adequate response information from LTM. In contrast, in our Experiment 2, single-task blocks were embedded within mixed-task blocks and task sets consisted of three different S-R links. Likely, these factors increased participants' tendency to rely on task cues to represent upcoming task demands in WM, even in single-task blocks.

WM Capacity Constraints

The pattern of results found for individuals with high-WM capacity was fully consistent with the "unitary WM + LTM" account. In contrast, for low-WM individuals, no reliable CDA set size effects were present for the unconstrained, change trial condition but appeared for either unconstrained, repeat trials or for constrained, change trials (see Figures 3 and 4). This twofold crossover pattern between WM capacity and level of experience suggests that both short-term (i.e., trial to trial) and longer-term (i.e., within a session) experience allowed low-WM individuals to operate in a way that is more similar to how high-WM participants do with novel tasks.

To our knowledge, this is the first study that tracked the effect of experience on the CDA set size effect as a function of WM capacity. The fact that, in low-WM individuals, set size effects emerged with experience confirms that the CDA reflects the degree to which available capacity is used to encode a given set of information (Balaban & Luria, 2016; Vogel, McCollough, & Machizawa, 2005). Repeated exposure seems to allow even low-WM individuals to construct representations that make use of their WM capacity—something high-WM individuals are able to do after only a single exposure to a new S-R set. Past research has suggested that the set size effect reflects in part the ability to represent relevant information, while efficiently filtering irrelevant information (Jost, Bryck, Vogel, & Mayr, 2010; Fukuda & Vogel, 2009; Vogel et al., 2005). In the current context, this could imply that repetitions allow individuals with low-WM capacity to represent S-R rules without potentially interfering information from either previous trials or perhaps from the visually presented placeholders (see Figure 1).

Although experience allowed low-WM individuals to represent task sets in a load-dependent manner, even the high frequency of repetitions in the constrained condition reduced set size effects only on repeat, but not on change trials, indicating continued reliance on active, set size-dependent WM representations. A similar pattern was also observed in Experiment 2, where individuals with low-WM capacity und large conflict effects tended to have a larger CDA. Combined, these results suggest that low-WM capacity not only reduces the ability to operate with novel material but can also constrain the establishment of a robust knowledge base that enables capacity-independent processing.

Although we had strong a priori predictions about the effects of practice on the set size effect in high-WM individuals, the conclusions regarding low-WM individuals are more tentative. Therefore, it will be important to explore in future research how WM capacity limits the ability to construct compact representations that allow individuals to efficiently use their limited capacity.

Limitations

To obtain a robust CDA, we had to use spatially induced S-R rules. Thus, we need to be careful in generalizing our conclusions to other representational formats, such as verbally instructed task sets. Nevertheless, our results provide an existence proof that a unitary WM system is involved in representing at least some types of action plans—in a load-dependent manner early and in a load-independent manner at later stages of practice. Although it is a parsimonious assumption that this principle generalizes to other representational formats, it certainly would need to be tested empirically. Our results are conceptually consistent with work by Dux, Tombu, Harrison, Rogers, Tong, & Marois (2009), who showed in a dual-task context that extensive practice does not divert processing from frontal, presumably WM-related processing sites to nonfrontal, or subcortical sites, as a two-system model might suggest (Oberauer et al., 2013). Instead, practice shortens the duration of frontal, WM-related processes.

Another qualification is that although there were robust correlations between WM capacity and overall performance in the unconstrained condition (corresponding to global task selection costs), we found no reliable correlations with local switch costs. A possible reason for this is that with both the long cue-stimulus interval and the strong emphasis on advanced preparation (which were both necessary to assess the CDA), our situation was not particularly conducive to producing robust local switch costs (see Table 1). Yet, our results are at least indirectly relevant for the question why switch costs and WM capacity seemed only weakly related in past research. We found that high-WM and low-WM individuals used the available WM capacity very differently, depending on the degree of familiarity with the relevant S-R sets, resulting in counteracting experience-dependent effects between these two groups (see Figure 4). Such complexities can make it very difficult to establish consistent relationships between WM and task-switching measures, even though WM is clearly involved in selection/preparation.

Conclusion

To conclude, flexible task selection/preparation is based on an active WM representation in a load-dependent manner for novel tasks. Even when a small number of tasks is repeated throughout the session, WM remains actively involved, at least for people with sufficient WM capacity in a load-independent manner—probably to maintain a retrieval cue that allows efficient response selection once the stimulus appears. Finally, our results indicate that the degree to which people can offload representations from WM to LTM itself is dependent on WM capacity.

Acknowledgments

This work was supported in part by National Instute of Aging grant R01 AG037564-01A1 and an award by the Humboldt Foundation to Ulrich Mayr.

Reprint requests should be sent to Ulrich Mayr, Department of Psychology, University of Oregon, Eugene, OR 97403, or via e-mail: mayr@uoregon.edu.

Note

1. In addition to the CDA results, we also observed that the amplitude of the N2pc (a negative, posterior component arising around 200 msec after stimulus onset) was significantly larger for the mixed-task than the single-task context, $F(1,27) = 6.96$, $p = .014$, $\eta^2 = .26$. As this pattern can be interpreted in terms of greater demands on attentional selection in the mixed-task condition (Luck & Hillyard, 1994), it is generally consistent with the CDA-related results. Note that, in Experiment 1, N2pc effects were likely masked by the 250 msec cue offset (100 msec in Experiment 2).

REFERENCES

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131,* 30–60.

Balaban, H., & Luria, R. (2015). Integration of distinct objects in visual working memory depends on strong objecthood cues even for different-dimension conjunctions. *Cerebral Cortex, 26,* 2093–2104.

Balaban, H., & Luria, R. (2016). Object representations in visual working memory change according to the task context. *Cortex, 81,* 1–13.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138,* 487–502.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision, 10,* 433–436.

Bryck, R., & Mayr, U. (2008). Task selection cost asymmetry without task switching. *Psychonomic Bulletin and Review, 15,* 128–134.

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *Journal of Neuroscience, 31,* 9315–9322.

Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences, 24,* 154–176.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134,* 9–21.

Dux, P. E., Tombu, M. N., Harrison, S., Rogers, B. P., Tong, F., & Marois, R. (2009). Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. *Neuron, 63,* 127–138.

Engle, R. W., & Kane, M. J. (2003). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation, 44,* 145–199.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211–245.

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

Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *Journal of Neuroscience, 29,* 8726–8733.

Ikkai, A., McCollough, A. W., & Vogel, E. K. (2010). Contralateral delay activity provides a neural measure of the number of representations in visual working memory. *Journal of Neurophysiology, 103,* 1963–1968.

Jost, K., Bryck, R. L., Vogel, E. K., & Mayr, U. (2010). Are old adults just like low working memory young adults? Filtering efficiency and age differences in visual working memory. *Cerebral Cortex, 21,* 1147–1154.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., et al. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136,* 849–874.

Lien, M., Ruthruff, E., Remington, R. W., & Johnston, J. C. (2005). On the limits of advance preparation for a task switch: Do people prepare all the task some of the time or some of the task all the time? *Journal of Experimental Psychology: Human Perception and Performance, 31,* 299–315.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8.*

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31,* 291–308.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17,* 391–400.

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience & Biobehavioral Reviews, 62,* 100–108.

Mayr, U. (2001). Age differences in the selection of mental sets: The role of inhibition, stimulus ambiguity, and response-set overlap. *Psychology and Aging, 16,* 96.

Mayr, U., Kuhns, D., & Hubbard, J. (2014). Long-term memory and the control of attentional control. *Cognitive Psychology, 72,* 1–26.

Mayr, U., Kuhns, D., & Rieter, M. (2012). Eye movements reveal dynamics of task control. *Journal of Experimental Psychology: General, 142,* 489–509.

McCollough, A. W., Machizawa, M. G., & Vogel, E. K. (2007). Electrophysiological measures of maintaining representations in visual working memory. *Cortex, 43,* 77–94.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7,* 134–140.

Oberauer, K., Souza, A. S., Druey, M. D., & Gade, M. (2013). Analogous mechanisms of selection and updating in declarative and procedural working memory: Experiments

and a computational model. *Cognitive Psychology, 66,* 157–211.

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

Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision, 10,* 437–442.

Rubin, O., & Meiran, N. (2005). On the origins of the task mixing cost in the cuing task-switching paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1477–1491.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology, 71,* 1–26.

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1656.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428,* 748–751.

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

Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology, 46,* 361–413.

Woodman, G. F., Carlisle, N. B., & Reinhart, R. M. (2013). Where do we store the memory representations that guide attention? *Journal of Vision, 13,* 1.