Journal of Experimental Psychology:
Human Perception and Performance
2018, Vol. 44, No. 7, 1038–1053

0096-1523/18/$12.00 http://dx.doi.org/10.1037/xhp0000513

Individual Differences in Working Memory Capacity and Filtering

Matthew K. Robison, Ashley L. Miller, and Nash Unsworth
University of Oregon

In three experiments we examined individual differences in working memory (WM) and their relationship with filtering—the selective encoding and maintenance of relevant information in the presence of irrelevant information. While some accounts argue that filtering is an important element of individual differences in WM (McNab & Klingberg, 2008; Robison & Unsworth, 2017b; Unsworth & Robison, 2016; Vogel, McCollough, & Machizawa, 2005), recent investigations have challenged this view (Mall, Morey, Wolff, & Lehnert, 2014; Shipstead, Lindsey, Marshall, & Engle, 2014). In all three experiments, we measured WM span with three complex span tasks and then had participants complete a visual WM task with a filtering component. In Experiment 1, participants were instructed to remember the orientation of relevant items (red rectangles) and ignore irrelevant items (blue rectangles). In Experiment 2, the color of relevant items changed randomly on a trial-by-trial basis. In Experiment 3, we presented a constant number of items. On half of the trials, participants were told which color item would be tested before each trial. On the other half of the trials, participants received no such cue. In situations where filtering was especially required, WM span accounted for a significant portion of variance in filtering trials beyond shared variance between filtering and nonfiltering trials. We argue that filtering is one of several control processes that gives rise to individual differences in WM, but that the relationship is constrained by the degree to which filtering is demanded by the task.

Public Significance Statement

At any given moment in our lives, we are bombarded with sensory information from a variety of sources. Our cognitive system is limited in its ability to attend to, encode, and remember information in a meaningful way. Therefore, we must selectively attend to, encode, and remember information that is most relevant to our current goals, an ability commonly referred to as filtering. The ability to do so has been proposed as an important element of a high-functioning cognitive system. However, this theoretical viewpoint has been challenged recently. We bring new evidence to the debate by further clarifying the situations in which filtering is especially required, as the relationship appears moderated by context. We demonstrate that in dynamic situations that overload working memory, high-capacity individuals more effectively filter irrelevant information, and this can partially explain why such individuals generally have higher-functioning cognitive systems.

Keywords: working memory capacity, visual working memory, attention control, filtering

A considerable debate remains about the nature of the relationship between WM capacity and control over access to WM. Whereas some have argued that an important element of individual differences in WM is the ability to selectively encode and maintain relevant information and ignore irrelevant information (i.e., filtering; McNab & Klingberg, 2008; Unsworth & Robison, 2016; Vogel et al., 2005), recent investigations have challenged this idea (e.g., Mall et al., 2014). Given the nature of this debate, the present investigation attempts to resolve some outstanding issues regarding the nature of individual differences in WM. Specifically, we measured WM with three complex span tasks and a visual WM task to examine the WM-filtering relationship. A secondary goal of the present investigation is to demonstrate that there a number of different ways to measure filtering, and we address some of the strengths and weaknesses of these various approaches.

A number of different attention control-related abilities are necessary for the successful execution of a visual WM task. These control abilities include the consistent deployment of attention to the task while resisting mind wandering and other attentional diversions (Adam, Mance, Fukuda, & Vogel, 2015; Adam & Vogel, 2017; Mrazek et al., 2012; Unsworth & Robison, 2016), allocating attention to items during maintenance intervals to ensure items stay in an easily accessible state (Unsworth, Fukuda, Awh, & Vogel, 2015; Unsworth & Robison, 2015; Vogel & Machizawa, 2004), selecting relevant information within WM (Robison & Unsworth, 2017b), and dealing with large amounts of information that exceed one's WM capacity (Fukuda, Woodman, & Vogel,

This article was published Online First April 23, 2018.

Matthew K. Robison, Ashley L. Miller, and Nash Unsworth, Department of Psychology, University of Oregon.

This research was supported by Office of Naval Research Grant N00014-15-1–2790 and National Science Foundation Grant 1632327 during data analysis and writing of this article.

Correspondence concerning this article should be addressed to Matthew K. Robison, Department of Psychology, University of Oregon, 1227, Eugene, OR 97403. E-mail: mkr@uoregon.edu

2015). Filtering—the ability to selectively encode and maintain task-relevant information in the presence of irrelevant information—is the focus of the present investigation. Evidence suggests that filtering is an important individual difference. For example, in the task used by Vogel and Machizawa (2004) individuals were briefly cued to a relevant hemifield of the screen (right or left), and items appeared on both sides of the screen. Individuals had to selectively encode and maintain information on the relevant side of the screen only. Contralateral delay activity (CDA) is a difference in activity between the contralateral hemisphere (i.e., activity in response to relevant information) and ipsilateral hemisphere (i.e., activity in response to irrelevant information). So in addition to measuring maintenance of relevant information, differences in CDA index the ability to ignore irrelevant information (Vogel & Machizawa, 2004). Furthermore, when both relevant and irrelevant information is presented in the relevant hemifield, low-capacity individuals' CDA reflects unnecessary storage of irrelevant information, whereas high-capacity individuals' CDA reflects near-exclusive storage of relevant information (Vogel et al., 2005). Follow-up investigations have shown that low-capacity individuals are more susceptible to and take longer to recover from attentional capture during visual WM tasks (Fukuda & Vogel, 2009; Fukuda & Vogel, 2011). Furthermore, Gaspar, Christie, Prime, Jolicœur, and McDonald (2016) found that the amplitude and timing of the distractor positivity, an EEG waveform that theoretically reflects suppression of irrelevant information, significantly correlated with estimates of visual WM capacity.

Additional evidence for the filtering account comes from functional MRI. McNab and Klingberg (2008) discovered two regions that showed different patterns of activity in preparation for filtering trials compared with nonfiltering trials: bilateral posterior middle frontal gyrus and the left basal ganglia. Outside the scanner, McNab and Klingberg (2008) measured individual differences in visual WM capacity. Importantly, activity in these two areas preceding filtering trials significantly correlated with visual WM capacity. Other work has demonstrated that filtering significantly predicts capacity estimates (Unsworth & Robison, 2016). Notably, filtering was uncorrelated with mind wandering, indicating sustained attention to the task and filtering are distinct forms of control. Finally, we have demonstrated that the use of spatial and categorical precues is related to independent measures of WM (complex span; Robison & Unsworth, 2017b). Taken together, this collection of evidence suggests that one important difference that gives rise to variability in WM is the ability to use attention control to filter out irrelevant information from WM (Awh & Vogel, 2008; Awh, Vogel, & Oh, 2006; Lee et al., 2010; McNab & Klingberg, 2008; Robison & Unsworth, 2017b; Unsworth & Robison, 2016; Vogel et al., 2005).

Recently, Mall et al. (2014) challenged the filtering account with an individual differences examination of WM capacity and selective attention. One issue with prior investigations is that filtering is often measured within the same task (or type of task) during which WM capacity is measured (Fukuda & Vogel, 2009; Fukuda & Vogel, 2011; Unsworth & Robison, 2016). To overcome that limitation, Mall et al. (2014) included two complex span tasks (operation span and symmetry span) as independent measures of WM. Mall et al. (2014) also noted that the stimulus presentation and retention interval durations are typically quite short (around 200 ms and 900 ms, respectively). These short durations prevent individuals from making saccades (often intentionally, as eye movements can make EEG analyses impossible). Longer stimulus presentation allows for individuals to fixate on multiple items during the exposure duration, thus potentially encoding more items than during shorter exposures. So Mall et al. (2014) lengthened both encoding (1,200 ms) and maintenance periods (3,000 ms) to allow for eye movements. Additionally, Mall et al. (2014) manipulated the relevance of items in three different ways. In one type of trial, all items had an equal probability of being tested (i.e., full-set trials). In the second type of trial, only items belonging to one particular category (circles or triangles; half-set trials) were tested. In the third type of trial, one category of items was tested two thirds of the time (ratio-set trials). Trial types were blocked and participants completed two blocks of each trial. Importantly, participants were explicitly instructed before half-set and ratio-set blocks that one category of items would be tested all (in half-set) or most (in ratio-set) of the time. Mall et al. (2014) made two predictions based on the filtering account: (a) low-capacity individuals should fixate more often on irrelevant items/locations during encoding and maintenance periods, and (a) low-capacity individuals should demonstrate better memory for infrequently tested information, assuming they encode and maintain this information more often than their high-capacity counterparts. Consistent with the idea that inefficient filtering leads to poor visual WM performance, individuals who performed better on the visual WM task spent less time looking at irrelevant items during both the encoding and maintenance intervals in the half-set condition (when those items were never tested). But individuals with lower complex span scores spent less time fixating on irrelevant items during both encoding and maintenance intervals in half-set blocks. Additionally, individuals with lower complex span scores showed worse memory for infrequently tested information. Mall et al. (2014) argued that both of these findings contradict the filtering account. Further, they highlighted the importance of including independent measures of WM.

Shipstead et al.'s (2014) findings also appear to challenge the filtering account. In their study, Shipstead and colleagues gave participants four different visual WM tasks. Two of those tasks required filtering, and two of them did not. Using a latent variable approach, Shipstead et al. (2014) formed a visual WM latent factor from all four tasks. They then formed a filtering latent variable from the residual shared variance between the two filtering tasks. Shipstead and colleagues also included two complex span tasks (operation span and symmetry span). The filtering latent variable and a complex span latent variable did not correlate. However, the filtering latent variable did correlate with an attention control latent variable representing the shared variance between antisaccade, Stroop, and flanker tasks. Thus it is apparent that there are some discrepancies in the literature surrounding the nature of individual differences in WM and their relationship with filtering. The primary goal of the present study is to resolve some of these discrepancies.

A secondary goal of the present investigation is to highlight the diversity of methods with which filtering can be measured, as well as some issues with those methods. In the four studies reviewed above, each measured filtering in a different way. Vogel et al. (2005) measured filtering as the difference in CDA between distractor-present and distractor-absent trials. Mall et al. (2014) measured filtering in two ways: (a) the proportion of time spent fixating on irrelevant items (or locations) during encoding and maintenance intervals, and (b) memory for infrequently tested information. Shipstead et al. (2014) measured filtering as the residual shared variance between filtering tasks after controlling for shared variance among filtering and nonfiltering tasks. Finally, filtering can be measured as the difference in capacity estimates between filtering and nonfiltering trials (Lee et al., 2010; Unsworth & Robison, 2016). In the present study filtering can be measured two different ways, given the experimental design. The first is as a filtering cost, which can be measured as the difference in performance between distractor-present and distractor-absent trials (or a ratio of performance between distractor-present and distractor-absent trials). The second is using regression to account for shared variance between distractor-present and distractor-absent trials and examining residual shared variance between complex span performance and distractor-present trials.

We chose to measure individual differences in WM capacity with three complex span tasks for two reasons. First, performance on complex span tasks require a constellation of abilities including the temporary storage of goal-relevant information, simultaneous processing of irrelevant/distracting information, and retrieval of goal-relevant information that is lost from active maintenance. Therefore, they require both storage and control, which filtering tasks also presumably require. Second, by using multiple complex span tasks that combine verbal and spatial memoranda, a factor composite score represents a rather domain general ability to store information and exert control. Shared variance between complex span WM and performance on the filtering task should reflect a domain-general ability to selectively encode and maintain goal-relevant information, sometimes in the face of distracting information.

The regression technique rests on two major assumptions: (a) the shared variance between distractor-present (i.e., filtering) and distractor-absent (i.e., nonfiltering) trials represents any shared influences of visual WM capacity and control processes that are necessary for encoding and maintaining information in visual WM more generally, and (b) any residual shared variance between distractor-present trials and complex span performance (which we will refer to as WM span), reflects a unique element of individual differences in WM that represents the selective encoding and maintenance of relevant information in the presence of irrelevant information. If there is no significant amount of shared variance between WM span and filtering trial performance after controlling for shared variance between filtering trials and nonfiltering trials, this would suggest that individual differences in WM span are primarily predictive of filtering because of shared variance between WM span and visual WM more generally. In other words, filtering abilities are not an element of individual differences in WM. However, if WM span and filtering trials share a significant amount of variance after controlling for shared variance between filtering and nonfiltering trials, this would suggest that individual differences in WM comprise a form of control that is unique to filtering trials. The logic of this regression technique is our primary means of testing the nature of individual differences in WM span and their relationship with filtering.

Experiment 1

The goal of Experiment 1 was to examine how individual differences in WM predicted filtering abilities within a visual WM task. During the visual WM task, participants were told that one category of items would always be tested, and one category of items was included as distractors. If WM span accounts for a significant portion of variance in filtering trials beyond the shared variance between filtering trials and nonfiltering trials, this would suggest that filtering is an element of individual differences in WM. However, if WM span does not account for a significant unique portion of variance in filtering trial performance, this would suggest filtering is not an important determinant of individual differences in WM.

Method

Participants and procedure. A sample of 158 participants from the human subjects pool at the University of Oregon completed the study in exchange for partial course credit. Participants completed three complex span tasks and a visual change detection task. The four tasks comprised about 45 min of a 120-min session, during which participants completed other measures that were irrelevant to the current study. These results have been reported elsewhere (Robison & Unsworth, 2017a, in press). We used the end of the academic term as our stopping rule for data collection. The experimental protocol was approved by the Institutional Review Board of the University of Oregon.

Tasks.

Operation span. Participants solved a series of math operations while trying to remember a set of unrelated letters (Unsworth, Heitz, Schrock, & Engle, 2005). Participants were required to solve a math operation, and after solving the operation, they were presented with a letter for 1 s. Immediately after the letter was presented the next operation was presented. At recall, participants were asked to recall letters from the current set in the correct order by clicking on the appropriate letters. For all of the span measures, items were scored correct if the item was recalled correctly from the current list in the correct serial position. Participants were given practice on the operations and letter recall tasks only, as well as two practice lists of the complex, combined task. List length varied randomly from three to seven items, and there were two lists of each length for a total possible score of 50. The score was the total number of correctly recalled items in the correct serial position.

Symmetry span. Participants recalled sequences of red squares within a matrix while performing a symmetry-judgment task (Unsworth, Redick, Heitz, Broadway, & Engle, 2009). In the symmetry-judgment task, participants were shown an 8×8 matrix with some squares filled in black. Participants decided whether the design was symmetrical about its vertical axis. The pattern was symmetrical half of the time. Immediately after determining whether the pattern was symmetrical, participants were presented with a 4×4 matrix with one of the cells filled in red for 650 ms. At recall, participants recalled the sequence of red-square locations by clicking on the cells of an empty matrix. Participants were given practice on the symmetry-judgment and square recall tasks as well as two practice lists of the combined task. List length varied randomly from two to five items, and there were two lists of each length for a total possible score of 28. We used the same scoring procedure as we used in the operation span task.

Reading span. While trying to remember an unrelated set of letters, participants were required to read a sentence and indicated whether or not it made sense (Unsworth et al., 2009). Half of the sentences made sense, while the other half did not. Nonsense sentences were created by changing one word in an otherwise normal sentence. After participants gave their response, they were presented with a letter for 1 s. At recall, participants were asked to recall letters from the current set in the correct order by clicking on the appropriate letters. Participants were given practice on the sentence judgment task and the letter recall task, as well as two practice lists of the combined task. List length varied randomly from three to seven items, and there were two lists of each length for a total possible score of 50. We used the same scoring procedure as we used in the operation span and symmetry span tasks.

Filtering. Participants tried to remember the orientations of colored rectangles (Vogel et al., 2005). We informed participants that they would be briefly presented with a pattern of red and blue rectangles, and that they should pay attention to the red rectangles and ignore the blue rectangles. Each trial began with a screen that said, "Remember, press left for same, right for different." The F and J keys on the keyboard were labeled "S" and "D" for same and different, and participants placed their left and right index fingers on these keys. Participants initiated each trial by pressing the spacebar. Following a 500-ms blank gray screen, a fixation cross appeared for 1,000 ms. After another 100-ms blank screen, an array of blue and red rectangles appeared and remained on-screen for 250 ms. Each rectangle could be angled in one of four directions: vertical, horizontal, 45° to the right, or 45° to the left. After a 900-ms blank retention interval, the items reappeared on the screen. One of the red rectangles had a white dot on it. Participants indicated whether this item was the same orientation or a different orientation than in the first presentation by pressing the key labeled S or D for same or different. The test array remained on-screen until the participant made a response. The tested item changed orientation on 50% of trials, and untested items never changed orientation. A graphical depiction of the task is shown in Figure 1. Arrays could contain two or four targets and zero, two, or four distractors, resulting in six trial types. Participants completed six practice trials after which they were encouraged to ask the experimenter any questions, if necessary. They then completed 144 experimental trials. Trial types were randomly intermixed. Participants completed 24 trials of each trial type, with two exceptions. Due to a programming error, there were no four-target/zero-distractor trials in which the target item changed orientation. Further, participants completed 36 trials in which there were two targets and two distractors, 24 of which were trials in which the orientation of the target changed. This programming error was corrected in Experiment 2. The dependent variable was proportion correct for each trial type.

Results and Discussion

Descriptive statistics and correlations among measures are listed in Table 1. Most measures showed acceptable values for skewness and kurtosis. The two-target/two-distractor trials showed negative kurtosis due to relatively high performance on those trials and a subsequent ceiling effect. To compute a single WM span score for each participant, we submitted operation span, symmetry span, and reading span scores into a principle axis factoring and saved factor scores for each participant. Loadings on this factor score were .83, .46, and .70 for operation span, symmetry span, and reading span, respectively. This factor score was used in all subsequent analysis involving WM span.

Our first analysis focused on performance on the filtering task. We submitted accuracy to a 2 (targets: 2, 4) × 3 (distractors: 0, 2, 4) repeated-measures analysis of variance (ANOVA). The ANOVA revealed a significant main effect of the number of targets ($F(1, 157) = 217.09$, $p < .001$, partial $\eta^2 = .58$), such that performance on two-target trials was better than performance on four-target trials. We also observed a significant main effect of

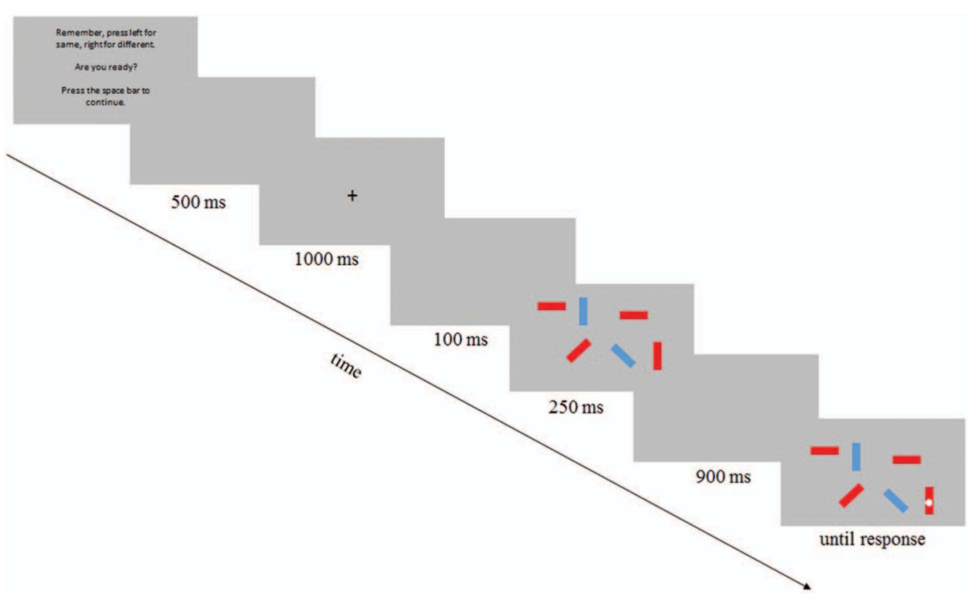

Figure 1. Example trial for filtering task in Experiment 1. See the online article for the color version of this figure.

Table 1
Descriptive Statistics and Correlations for Experiment 1

Measure	1	2	3	4	5	6	7	8	9
1. Operation span									
2. Symmetry span	.39								
3. Reading span	.61	.38							
4. 2 targets/0 distractors	.17	.17	.14						
5. 2 targets/2 distractors	.12	.25	.12	.70					
6. 2 targets/4 distractors	.08	.13	.10	.57	.75				
7. 4 targets/0 distractors	.14	.20	.02	.60	.66	.54			
8. 4 targets/2 distractors	.20	.24	.16	.53	.68	.61	.62		
9. 4 targets/4 distractors	.06	.16	.01	.43	.46	.52	.42	.44	
Mean	38.77	19.30	37.12	.92	.88	.88	.80	.78	.82
SD	8.72	5.16	9.20	.10	.11	.11	.14	.13	.17
Skew	−1.34	−.75	−.88	−2.24	−1.17	−1.16	−.99	−.36	−1.24
Kurtosis	2.77	.56	.62	7.63	.84	1.08	.83	−.55	2.12
Reliability	.71	.65	.70	.70	.77	.93	.93	.59	.48

Note. $N = 158$. *SD* = standard deviation. Reliabilities for operation span, symmetry span, and reading span were computed as Cronbach's α on each set size. Reliability for the trial types in the filtering task were computed using a Spearman-Brown split-half coefficient. Correlations ≥.16 are significant at $p < .05$.

number of distractors ($F(2, 314) = 8.52$, $p < .001$, partial $\eta^2 = .05$), such that the effect of distractors was different at each set size (see Table 1). Finally, we observed a significant Target × Distractor interaction ($F(2, 314) = 12.30$, $p < .001$, partial $\eta^2 = .07$). To estimate the effect of distractor presence, we subtracted accuracy on distractor-present trials from accuracy on distractor-absent trials (Lee et al., 2010; Unsworth & Robison, 2016). On average, accuracy dropped by .02 (*SD* = .08) when distractors were present, a small but significant effect, $t(157) = 3.06$, $p = .003$.

Our next set of analyses focused on how individual differences in WM span accounted for variance in filtering. First, we examined the correlation between the filtering cost described above and WM span. This correlation was not significant, $r = .01$, $p = .94$. In our next analysis, we entered WM span as a covariate into the ANOVA described above. Although there was a main effect of WM span ($F(1, 157) = 4.93$, $p = .03$, partial $\eta^2 = .03$), such that individuals with greater WM spans performed better overall, $r = .18$, $p = .03$, WM span did not interact with the effect of targets, distractors, or the Target × Distractor interaction (all *F*s < 2), suggesting that none of the effects significantly changed as a function of WM span.

Next, we used regression to estimate shared variance between filtering and nonfiltering trials. Presumably, the shared variance between distractor-absent (nonfiltering) and distractor-present (filtering) trials represents any demands on storage capacity and control processes that are common to all trial types. After controlling for this shared variance, any residual shared variance between WM span and distractor-present trials should be due to WM-related control abilities that are unique to these trials. For the regression analysis, we averaged performance on distractor-absent and distractor-present trials across set sizes. We then entered distractor-absent trial accuracy and WM span into a multiple regression as predictors of distractor-present trials. As can be seen in Table 2, only distractor-absent trials accounted for a significant amount of unique variance in distractor-present trials.[1] So in this instance, performance was primarily driven by capacity and control processes shared across all trial types. Although WM span significantly correlated with accuracy on distractor-present trials, $r = .17$, $p = .04$, this covariance was driven by WM span's shared variance with both trial types.

There were several elements of the filtering task in Experiment 1 that could have constrained the relationship between WM span and filtering. First, the relevant items were always red. So over time, most participants may have been able to sufficiently ignore the irrelevant blue items. Indeed, the average filtering costs at each set size were quite small. The main effect of distractors was significant, but because it was so small, there was little between-subjects variance with which WM span could systematically covary. As mentioned earlier, the magnitude of the filtering effect was quite small. In order to detect such a small effect, there must necessarily be little interindividual variation in the magnitude of the effect. Subsequently examining correlations between such an effect and other measures becomes difficult because of the lack of interindividual variability. Therefore, in order to increase the magnitude of the filtering effect, increase the amount of interindividual variability in the filtering effect, or both, we adjusted the filtering task to make the identity of targets change randomly on a trial-to-trial basis.

Experiment 2

In Experiment 1, WM span did not account for a significant portion of variance in filtering trial accuracy after controlling for shared variance between filtering and nonfiltering trials. Therefore, the primary driver of the relationship between WM span and

[1] There are two additional ways to perform this analysis, both of which reach the same conclusion. One way is to regress WM span on distractor-absent trials and distractor-present trials, and examine the unique effect of distractor-present trials on WM span. Another way is to examine the partial correlation between WM span and distractor-present trials, controlling for distractor-absent trials. In all three experiments, these alternative analyses yielded similar patterns to the regressions we report (see the Appendix). Further, using d' as the dependent variable for the filtering task, rather than raw accuracy, yields identical patterns of results in all three experiments. For Experiment 1, the partial correlation was not significant, $r = .07$, $p = .37$.

Table 2
Hierarchical Regression Analysis for Experiment 1

Predictor	R^2	β	SE_β	t	p
Step 1	.55				
Distractor-absent		.74	.03	13.82	<.001
Step 2	.55				
Distractor-absent		.73	.03	13.48	<.001
WM span		.05	.05	.90	.37

Note. $N = 158$. β = standardized regression coefficient; SE_β = standard error of standardized regression coefficient; WM = working memory.

the filtering task was WM span's relationship with capacity and control differences that were shared between filtering and nonfiltering trials. However, target trials consistently and predictably differed from distractors in a salient feature (color). As Mall et al. (2014) note, a consistent target-distractor relationship may mask individual differences. Indeed, the magnitude of the filtering effect in Experiment 1 was quite small, and there was little interindividual variability in the magnitude of the effect. Furthermore, Jost and Mayr (2016) examined the effect of target/distractor consistency and observed significant effects of trial-to-trial changes in item relevance. Specifically, Jost and Mayr (2016) had participants complete a filtering task (Vogel et al., 2005) with two types of trial blocks. In "pure" blocks of trials, targets were always one color (red or blue). In "mixed" blocks, target color changed randomly from trial to trial. Filtering was significantly worse in mixed blocks compared with pure blocks, suggesting it is more difficult for participants to adjust their filter settings on a trial-by-trial basis (Jost & Mayr, 2016). Therefore, the flexible adjustment of filtering on a moment-to-moment basis may be an important element of the filtering-WM relationship.

To test this possibility, the task used in Experiment 2 required participants to update which items were relevant and irrelevant on a trial-by-trial basis. On 50% of trials, the tested item was red. On the other 50%, the tested item was blue. Before each trial, participants were told which color would be tested, and trial types were randomly intermixed. Therefore, they could not utilize the same filtering approach on every trial. If this task requires WM-related control processes that are not shared between filtering and nonfiltering trials, WM span should account for a significant amount of variance in filtering trial accuracy, even after controlling for shared variance between filtering and nonfiltering trials. However, if WM span does not account for a significant unique portion of unique variance in filtering trial accuracy, as in Experiment 1, this would provide further evidence than WM span's covariation with filtering abilities is primarily driven by its shared variance with general WM-related capacity and control differences that are required in both filtering and nonfiltering contexts.

Method

Participants and procedure. A sample of 137 participants from the University of Oregon human subjects pool completed the study in exchange for partial course credit. Participants completed three complex span tasks and a visual change detection task. The four tasks comprised about 45 min of a 90-min session during which participants completed other tasks that were irrelevant to the present study, and the results have been reported elsewhere (Robison & Unsworth, 2017a). The complex span tasks were completed during the first 30 min of the session, and the visual change detection task was the last task in the session. We used the end of the academic term as our stopping rule for data collection. The experimental protocol was approved by the Institutional Review Board of the University of Oregon. No participants in Experiment 2 had taken part in Experiment 1.

Tasks.

Operation span. See Experiment 1.

Symmetry span. See Experiment 1.

Reading span. See Experiment 1.

Filtering. The task was similar to the filtering task in Experiment 1 with one crucial difference. Rather than the same color items being tested on every trial, the task cued participants to the relevant color at the beginning of every trial (Shipstead et al., 2014). Each trial began with a screen saying, "Remember press left for same, right for different. Are you ready? Press the spacebar to begin." After the participant pressed the spacebar, a 500-ms blank gray screen appeared, followed by a 1,000-ms fixation screen on which a black cross was centered on a gray background. After another 50-ms blank screen, the word "RED" or "BLUE" appeared and remained on-screen for 200 ms. This screen informed the participant which color the tested item would be. After a 100-ms blank screen, the target array appeared and remained on-screen for 250 ms. The arrays contained red and blue angled rectangles at one of four orientations: horizontal, vertical, 45° to the right, and 45° to the left. Arrays could contain two or four targets and zero, two, or four distractors. Targets could either be red or blue, resulting in 12 different trial types. If the targets were red, the distractors were blue, and vice versa. After a 900-ms blank retention interval, the array reappeared and remained on-screen until the participant made a response. The participants' task was to indicate whether the tested item was the same orientation or a different orientation than in the initial array. A white dot on one target item indicated the tested item. The tested item was always drawn from the subset of items indicated by the relevant color (i.e., distractor items were never tested). Participants made their response by pressing one of two keys labeled S and D for same and different (the F and J keys on the keyboard). The orientation of the tested item changed on 50% of trials. The orientation of untested items never changed. A graphical depiction of the task is shown in Figure 2. Participants completed six practice trials, after which they were encouraged to ask any questions if necessary. They then completed 120 experimental trials (10 of each trial type).

Results and Discussion

Descriptive statistics and correlations for all measures are shown in Table 3. Most measures showed acceptable values for skewness and kurtosis. The kurtosis values for the two-target/zero-distractor and two-target/two-distractor trials were a bit high, but accuracy on these trials was highest and the kurtosis was probably due to a ceiling effect. Just as in Experiment 1, we computed a WM span score for each participant by submitting operation span, symmetry span, and reading span scores to principle axis factoring. Loadings on this factor were .94, .44, and .57 for operation span, symmetry span, and reading span, respectively. This factor score was used in all subsequent analyses involving WM span.

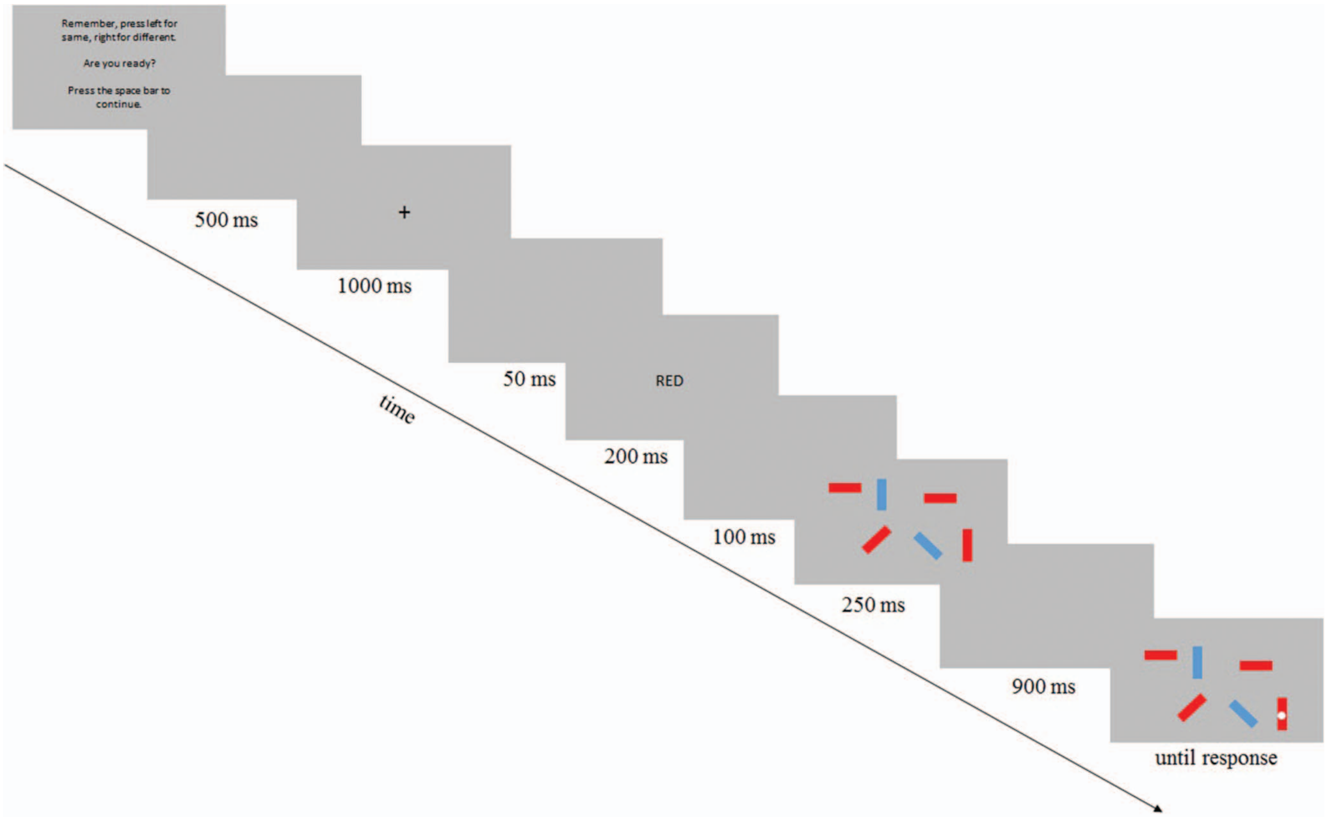

Figure 2. Example trial for filtering task for Experiment 2. See the online article for the color version of this figure.

Our first analysis was a repeated-measures ANOVA on accuracy with targets (2, 4) and distractors (0, 2, 4) as within-subjects factors. The ANOVA revealed a main effect of targets ($F(1, 136) = 220.66$, $p < .001$, partial $\eta^2 = .62$), such that accuracy was higher for two-target trials compared with four-target trials. The ANOVA also revealed a main effect of distractors ($F(2, 272) = 34.63$, $p < .001$, partial $\eta^2 = .20$), such that accuracy decreased with an increasing number of distractors. The Target × Distractor interaction did not reach significance ($F(2, 272) = 2.13$, $p = .12$, partial $\eta^2 = .02$). Follow-up Bonferroni-corrected comparisons revealed that accuracies on all trial types were significantly different from one another (all *p*s $< .01$). On average, accuracy dropped by .04 (*SD* = .07). Therefore, the magnitude of the filtering effect nearly doubled from Experiment 1, $t(293) = 2.91$,

Table 3
Descriptive Statistics and Correlations for Experiment 2

Measure	1	2	3	4	5	6	7	8	9
1. Operation span									
2. Symmetry span	.42								
3. Reading span	.54	.24							
4. 2 targets/0 distractors	.34	.23	.21						
5. 2 targets/2 distractors	.36	.27	.29	.71					
6. 2 targets/4 distractors	.24	.23	.20	.74	.73				
7. 4 targets/0 distractors	.26	.32	.24	.59	.66	.63			
8. 4 targets/2 distractors	.39	.35	.34	.63	.63	.68	.64		
9. 4 targets/4 distractors	.23	.38	.25	.49	.51	.63	.58	.68	
Mean	38.21	18.93	36.72	.90	.88	.86	.81	.77	.74
SD	8.68	5.17	9.10	.12	.13	.13	.14	.14	.14
Skew	−1.22	−.50	−.93	−2.11	−1.91	−1.37	−.80	−.70	−.35
Kurtosis	2.21	−.37	1.22	4.81	4.25	2.02	.42	−.002	−.43
Reliability	.70	.54	.72	.76	.76	.70	.65	.54	.53

Note. $N = 137$. *SD* = standard deviation. Cronbach's α on set sizes was used estimate reliability for the complex span tasks. Spearman-Brown split-half coefficients were used to estimate reliability for each trial type for the filtering task. All correlations are significant at $p < .05$.

$p = .004$. Thus, making the target color random on a trial-by-trial had the desired effect of increasing the filtering effect.

Our next analysis examined how individual differences in WM span predicted filtering abilities. First, we entered WM span as a covariate into the ANOVA described above. Similar to Experiment 1, the analysis of covariance (ANCOVA) revealed a significant main effect of WM span ($F(1, 135) = 23.43$, $p < .001$, partial $\eta^2 = .15$), such that participants with greater WM span performed better overall, $r = .39$, $p < .001$. But in this case, WM span significantly interacted with the effect of distractors ($F(2, 270) = 4.54$, $p = .01$, partial $\eta^2 = .03$). The drop in accuracy between distractor-present and distractor-absent trials was slightly but not significantly smaller for high-WM participants, $r = -.11$, $p = .20$. Similar to Experiment 1, we averaged across set size within distractor-present and distractor-absent trials. Then, we entered WM span and distractor-absent trial accuracy as predictors into a stepwise regression on distractor-present trials. As can be seen in Table 4, distractor-absent trial accuracy accounted for the majority (68%) of variance on distractor-present trials. However, WM span accounted for a small but significant unique proportion of variance (1%).[2] Thus, there is some residual variance explained by a WM-related form of control that contributes uniquely to performance on trials that require filtering, especially when what needs to be filtered changes on a trial-by-trial basis. We also computed a filtering cost by subtracting average accuracy on distractor-present trials from average accuracy on distractor-absent trials. Higher filtering costs reflect a greater effect of distractor presence on performance. On average, accuracy dropped by .04 ($SD = .07$) when distractors were present. This measure and WM span did not significantly correlate, $r = -.06$, $p = .47$. So, had we used this measure of filtering to see how WM span predicted filtering costs rather than the regression, we would have come to a different conclusion. This is an issue we return to later.

It is possible that the effect if WM span was due to trial-to-trial carry-over effects of previous filtering settings (Jost & Mayr, 2016). In other words, low-WM span participants may have had more difficulty switching from remembering red items to remembering blue items (and vice versa) than high-WM span participants. If this is the case, we should see an interaction between WM span and switching. To examine this, we submitted accuracy to a repeated-measures ANCOVA with within-subjects factors of targets (2, 4), distractors (0, 2, 4), and trial type (switch trial, no-switch trial) and WM span as a covariate. This analysis revealed a main effect of trial type ($F(1, 135) = 13.88$, $p < .001$, partial $\eta^2 = .09$), such that accuracy was lower when the relevant color changed relative to the preceding trial (switch: $M = .82$, $SD = .11$; no switch: $M = .84$, $SD = .11$), and a Distractor × Trial Type interaction ($F(2, 270) = 5.42$, $p = .005$, partial $\eta^2 = .04$), such that the effect of distractors was larger on switch trials (switch: $M = .06$, $SD = .09$; no switch: $M = .04$, $SD = .10$). However, WM span did not significantly interact with trial type ($F < 1$) or with the Trial Type × Distractor interaction ($F < 1$). So although we did observe effects of switching between filter settings replicating Jost and Mayr (2016), these effects did not interact with WM span, and they did not explain the relative filtering benefit for high-WM participants.

Table 4
Hierarchical Regression Analysis for Experiment 2

Step	R^2	β	SE_β	t	p
Step 1	.68				
Distractor-absent		.83	.02	17.04	<.001
Step 2	.69				
Distractor-absent		.79	.03	15.44	<.001
WM span		.11	.05	2.13	.04

Note. $N = 137$. β = standardized regression coefficient; SE_β = standard error of standardized regression coefficient; WM = working memory.

In Experiment 1, WM span did not account for a significant portion of variance in filtering trials after controlling for shared variance between filtering and nonfiltering trials. But as mentioned, the relevant color never changed. In Experiment 2, participants had to update which items were relevant and irrelevant on a trial-by-trial basis, which presumably requires greater moment-to-moment control over access to WM. In Experiment 2, we did observe a small but significant effect of WM span on filtering trial performance, even after controlling for shared variance between filtering and nonfiltering trials. Therefore, performance on filtering trials was determined by shared capacity and control differences between filtering and nonfiltering trials and a WM-related control difference that allowed for the selective encoding and maintenance of relevant information when item relevance was dynamic.

Experiment 3

Experiment 3 had several goals. First, we wanted to replicate the findings of Experiment 2 that WM span accounted for unique variance in filtering after controlling for shared variance between filtering and nonfiltering trials. Admittedly, the effect of WM span was small in Experiment 2. Therefore, we wanted to verify this finding with an additional experiment. Second, we wanted to place a premium on the use of the filtering cue. To do so, we fixed the set size to be six items—three blue items and three red items. On 50% of trials, participants received a cue telling them which color the tested item would be. The other 50% of trials provided no such cue. Thus, on some trials, participants had to maintain all six items, which is presumably above capacity for all participants. On others, they could filter out the irrelevant items and reduce the effective set size to three items, which is presumably at or near most participants' visual WM capacity. In Experiment 2, set sizes were often at or below visual WM capacity limits. In such situations, some participants (especially high-WM participants) may strategically choose to ignore the filtering cue, as this places an unnecessary additional burden on WM. Instead, they may attempt to remember all items. Pushing the set size to six items will presumably make this strategy impossible. Thus, we are further incentivizing participants, especially high-WM participants, to make effective use of the filtering cue. This manipulation had two intentions: (a) place a premium on the use of the filtering cue, and (b) increase the magnitude of the filtering effect and thus create more interindividual variability.

[2] The partial correlation between WM span distractor-present trials (controlling for distractor-absent trials) was also significant, $r = .17$, $p = .048$.

Method

Participants and procedure. A sample of 158 participants from the human subjects pool at the University of Oregon completed the study in exchange for partial course credit. Participants completed three complex span tasks and a visual change detection task. The four tasks comprised 45 min of a 90-min session during which participants completed three measures of long-term memory which were irrelevant to the current study. We used the end of an academic term as our stopping rule for data collection. The experimental protocol was approved by the Institutional Review Board of the University of Oregon. No participants in Experiment 3 had taken part in Experiments 1 or 2.

Tasks.

Operation span. See Experiment 1.

Symmetry span. See Experiment 1.

Reading span. See Experiment 1.

Filtering. This task was similar to that used in Experiment 2 with several key differences. Filtering was examined by keeping set size constant and providing a color cue on 50% of trials. The other 50% of trials provided no color cue. Specifically, three red items and three blue items appeared on every trial. Each trial began with a screen that said, "Remember, press left (S) for same, right (D) for different. Press the spacebar to start the trial." After the participant pressed the space bar, a 1,200-ms fixation screen appeared. In the cued trials, the fixation screen was followed by a 200-ms blank gray screen, then a 250-ms cue screen that showed the relevant color in capitalized letters and colored font (i.e., RED in red font or BLUE in blue font). Then, after another 250-ms blank screen, the sample array appeared and remained on-screen for 300 ms. Items could have one of four orientations: horizontal, vertical, 45° to the right, or 45° to the left. On neutral trials, the sample array was preceded by a 450-ms blank screen. After a blank 1,000-ms retention interval, the test array appeared. On cued trials, the tested item was always one of the items in the relevant color. On neutral trials, any item could be tested. A white dot on one item indicated the tested item. Participants responded as to whether the tested item was the same orientation or a different orientation as its first appearance by pressing the key marked S for same or D for different (the F and J keys on the keyboard). The tested item changed orientation on 50% of trials. The orientation of untested items never changed orientation. The test array remained on-screen until the participant made a response. A graphical depiction of each trial type is shown in Figure 3.

Results and Discussion

Descriptive statistics and correlations are listed in Table 5. All measures showed acceptable skew and kurtosis. We computed a WM span factor score for each participant using principle axis factoring on operation span, symmetry span, and reading span scores. The loadings on this factor were .83, .46, and .70 for operation span, symmetry span, and reading span, respectively. This score was used for all subsequent analysis involving WM span.

We first analyzed accuracy as a function of trial type. Accuracy was significantly higher on cued trials compared with neutral trials (paired samples $t(157) = 21.72$, $p < .001$). To compute a filtering score for each participant, we subtracted accuracy on neutral trials from accuracy on cued trials. On average, performance improved by a proportion of .18 ($SD = .10$) when a filtering cue was provided. We next used a repeated-measures ANCOVA on accuracy with trial type (neutral, cued) as a within-subjects factor and WM span as a covariate to whether this effect changed as a function of one's WM span. The ANCOVA revealed a main effect of WM span ($F(1, 156) = 16.25$, $p < .001$, partial $\eta^2 = .09$), such that individuals with greater WM span exhibited higher accuracy overall, $r = .31$, $p < .001$. The WM Span × Trial Type interaction was not quite significant ($F(1, 156) = 2.87$, $p = .09$, partial $\eta^2 = .02$), as the difference between neutral and cued trials did not significantly correlate with WM span, $r = .13$, $p = .09$. But similar to Experiments 1 and 2, we used regression to separate shared and unique variance in cued trials attributable to WM span and neutral trials. The results of the regression are shown in Table 6. Similar to Experiment 2, WM span had a significant amount of shared variance with cued trials (3%), even after accounting for shared variance between cued and neutral trial accuracy.[3] Therefore, there was a WM-related control difference that predicted filtering over and above the capacity and control differences shared between cued and neutral trials. We argue this difference is the selective encoding and maintenance of relevant items on cued trials.

Measuring filtering. One important facet of the current set of experiments is the use of regression techniques to examine the shared and unique influences of visual WM capacity and WM span on filtering abilities. Our logic rested on two assumptions: (1) when entered into a simultaneous regression predicting accuracy on trials that require filtering, covariance between filtering and nonfiltering trials represents any common WM-related differences that lead to variation on visual WM tasks regardless of filtering requirements, and (2) residual covariance between WM span and filtering trials represents a WM-related control difference that reflects the selective encoding and maintenance of relevant information in the presence of irrelevant information. This approach differs from a rather typical approach in which filtering abilities are estimated as a difference between distractor-absent and distractor-present trials (e.g., Lee et al., 2010; Unsworth & Robison, 2016). A secondary goal of the present study was to demonstrate that this approach may not yield the same conclusions as the regression approach.

The present set of experiments demonstrated that filtering can be measured in a number of different ways: subtracting accuracy on two-target/two-distractor trials from accuracy on two-target/zero-distractor trials, subtracting accuracy on two-target/four-distractor trials from accuracy on two-target/zero-distractor trials, subtracting accuracy on four-target/two-distractor trials from four-target/zero-distractor trials, subtracting accuracy on four-target/four-distractor trials from four-target/zero-distractor trials (Experiments 1 and 2), and subtracting noncued trial accuracy from cued trial accuracy (Experiment 3). We also divided accuracies on various trial types to obtain a ratio, rather than a difference score. Let us focus specifically on six filtering scores: (1) the difference between two-target/zero-distractor trials and two-target/two-distractor trials, (2) the difference between four-target/zero-distractor trials and four-target/two-distractor trials, (3) the ratio of accuracy on two-target/zero-distractor trials to two-target/zero-

[3] The partial correlation between WM span and cued trials (controlling for neutral trials) was significant, $r = .22$, $p = .005$.

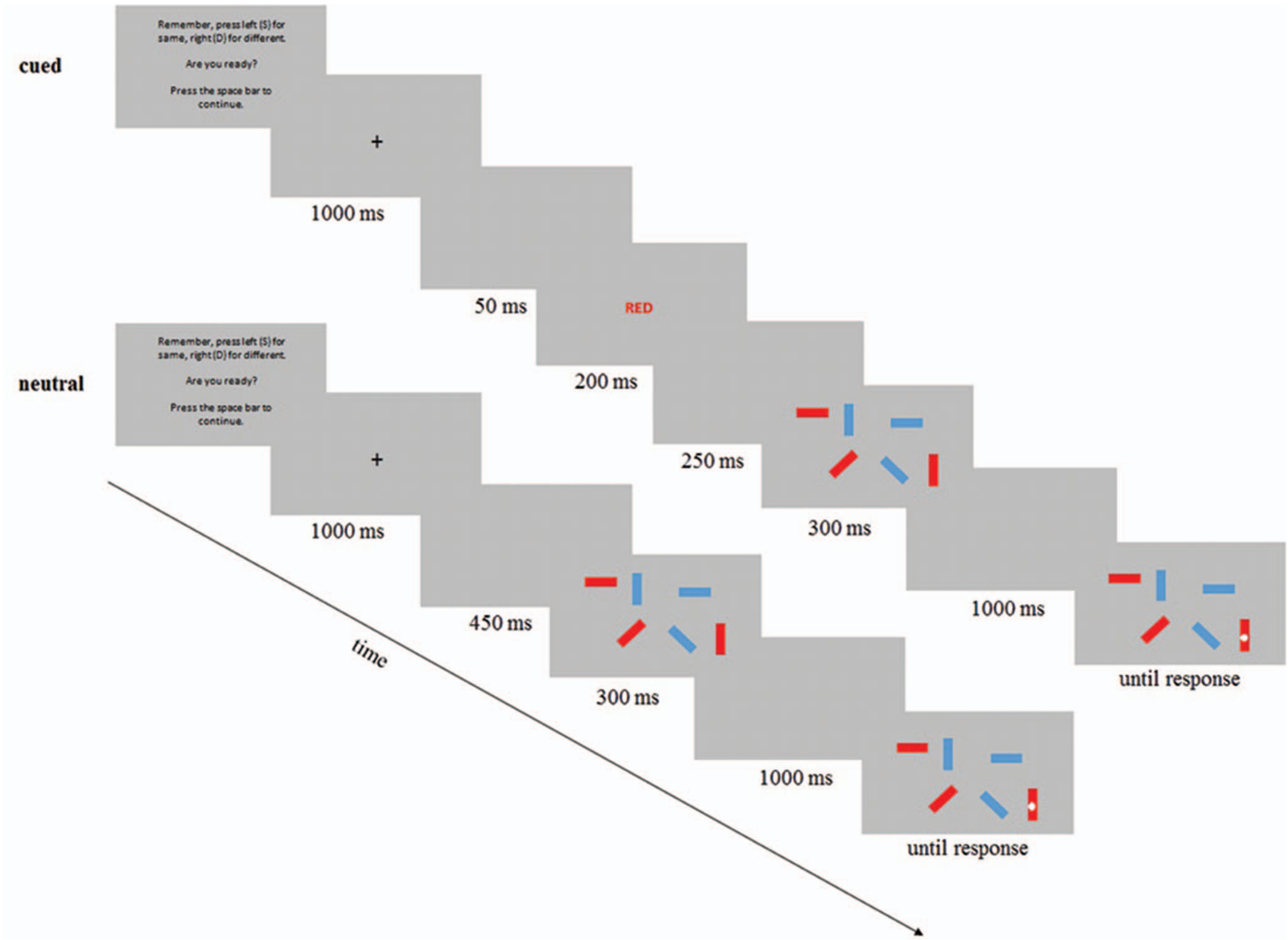

Figure 3. Example of cued and neutral trials in Experiment 3. See the online article for the color version of this figure.

distractor trials, (4) the ratio of accuracy on four-target/zero-distractor trials to four-target/zero-distractor trials, (5) the difference between cued trial accuracy and neutral trial accuracy, and (6) the ratio of cued trial accuracy to neutral trial accuracy. Table 7 shows these estimates, their correlations with WM span, and estimates of reliability. We estimated reliability using Irwin's (1966) formula for difference scores: $r_\Delta = [(r_1 + r_2/2) - r_{12}]/(1 - r_{12})$.

The correlations and reliabilities in Table 7 reveal several things worth noting. First, the reliability estimates are all quite low. Therefore, it may be difficult to use these filtering scores as estimates of a reliable individual difference, meaning a failure to observe a correlation with some other independent measure (e.g., WM span) could result from the unreliability of the filtering measure itself.

Table 5
Descriptive Statistics and Correlations for Experiment 3

Measure	1	2	3	4	5
1. Operation span					
2. Symmetry span	.38				
3. Reading span	.58	.32			
4. Neutral	.18	.29	.18		
5. Cued	.27	.33	.23	.58	
Mean	37.42	19.56	37.84	.65	.83
SD	7.92	5.00	7.76	.10	.12
Skew	−.88	−.61	−.94	−.21	−1.05
Kurtosis	.89	.13	1.46	−.23	1.07
Reliability	.61	.59	.66	.76	.52

Note. $N = 158$. *SD* = standard deviation. Reliability was estimated using Cronbach's α on set sizes for the three complex span tasks. Spearman-Brown split-half coefficient was used to estimate reliability for the neutral and cued trials of the filtering task. All correlations are significant at $p < .05$.

Table 6
Hierarchical Regression Analysis for Experiment 3

Step	R^2	β	SE_β	*t*	*p*
Step 1	.34				
Neutral		.58	.05	8.96	<.001
Step 2	.37				
Neutral		.54	.05	8.27	<.001
WM span		.19	.06	2.89	.004

Note. $N = 158$. β = standardized regression coefficient; SE_β = standard error of standardized regression coefficient; WM = working memory.

Table 7
Estimates of Filtering Using Difference Scores and Proportional Scores

Measure	Mean (*SD*)	WM span *r*	Reliability
Experiment 1 (N = 157)			
Two-target difference	.04 (.08)	−.01	.12
Two-target ratio	.96 (.10)	.01	.26
Four-target difference	.02 (.12)	−.11	.37
Four-target ratio	.99 (.17)	.12	.19
Experiment 2 (N = 137)			
Two-target difference	.02 (.10)	−.08	.17
Two-target ratio	.98 (.12)	.09	.06
Four-target difference	.04 (.12)	−.17*	−.13
Four-target ratio	.96 (.16)	.18*	.01
Experiment 3 (N = 158)			
Cued trial difference	.18 (.10)	.13	.38
Cued trial ratio	.80 (.12)	−.10	.11

Note. *SD* = standard deviation. Working memory (WM) span r = correlation between estimate and WM span factor. Negative reliability for one measure reflected lower intra-measure consistency than inter-measure correlation.
* $p < .05$.

Another issue worth noting is that filtering is often measured within the same task as other estimates of visual WM capacity. For example, Vogel et al. (2005) measured filtering as the difference in CDA between filtering and nonfiltering trials and measured differences in visual WM capacity on the same task. In another example, Lee et al. (2010) used a filtering task in which participants tried to remember the orientations of colored rectangles in one hemifield, which is quite similar to the task used in the current set of experiments. Lee et al. (2010) gave three set sizes: two targets/zero distractors, two targets/two distractors, and four targets/zero distractors. They estimated filtering by subtracting performance (capacity estimates) on two-target/two-distractor trials from performance on two-target/zero-distractor trials. They then correlated this difference with performance on four-target/zero-distractor trials. We have also used this estimation method (Unsworth & Robison, 2016) by computing a difference in capacity estimates between distractor-absent and distractor-present trials, and we observed a significant correlation with capacity estimates. If we did this in the current set of experiments, we would observe a significant correlation in Experiment 1, $r = -.17$, $p = .03$. In Experiment 2 the correlation was not significant, $r = -.15$, $p = .07$. These findings are not consistent with our conclusions using regression. So we agree with Mall et al. (2014) on the importance of having independent measures to examine the joint and unique influence of abilities like WM capacity and filtering.

General Discussion

Across three experiments, we examined the relationship between WM span and the ability to selectively encode and maintain relevant information in the presence of irrelevant information (i.e., filtering) during a visual WM task. To measure individual differences in WM, we gave participants three complex span tasks. Participants then completed a visual WM task with a filtering component. In Experiment 1, participants were told at the beginning of the task that only one category (red items) would be tested, and they should do their best to ignore the irrelevant (blue) items. The independent measure of WM (a factor score derived from the three complex span tasks) did not account for a significant portion of variance in filtering trials over and above shared variance between filtering and nonfiltering trials. So in that instance, WM span did not independently predict filtering abilities. However, as Mall et al. (2014) note, the predictability of targets based on a salient dimension of the items may have limited our ability to observe such a relationship. As the target-distractor relationship is continually reinforced by the task, the presence of distractors may become less impactful. Further, Jost and Mayr (2016) observed significant effects on filtering when the relevant color was random rather than fixed within blocks of trials. So in Experiment 2, we made the color of targets and distractors unpredictable. The to-be-remembered color changed on a trial-by-trial basis. In this instance, WM span did indeed share a significant portion of variance with filtering trials beyond the shared variance between filtering and nonfiltering trials. Therefore, when the relevance of items was unpredictable, there was a WM-related control difference that accounted for variance in filtering trials. Admittedly, this effect was small, only accounting for an additional 1% of variance after controlling for shared variance between filtering and nonfiltering trials. So to replicate this effect, we ran a third experiment. In Experiment 3, the relevance of items again changed on a trial-by-trial basis, but we stabilized set size such that all trials included six items (three targets and three distractors). On half of the trials, participants were cued to a relevant subset by color. On the other half of the trials, participants received no such cue, and thus had to encode and attempt to retain all items. In this case, WM span accounted for a significant portion of variance (3%) on cued trials after controlling for the shared variance between cued and uncued trials. So similar to Experiment 2, there was a WM-related difference in trials that allowed for filtering beyond the WM-related control and capacity differences that are shared across filtering and nonfiltering situations.

The nature of the relationship between WM span and filtering is a subject of considerable debate, and this debate inspired the present set of experiments. Prior research has argued that individual differences in WM are strongly predictive of filtering abilities. Indeed, the ability to control access of information to WM is often offered as one reason for individual differences in WM (Awh et al., 2006; Cowan & Morey, 2006; McNab & Klingberg, 2008; Robison & Unsworth, 2017b; Unsworth & Robison, 2016; Vogel et al., 2005). However, this account is not supported by several recent studies (Mall et al., 2014; Shipstead et al., 2014). Further, various studies have conceptualized WM in a number of different ways. We and others (Mall et al., 2014; Robison & Unsworth, 2017b; Shipstead et al., 2014) have measured WM with complex span tasks, which require the temporary storage and retrieval of relevant information while processing irrelevant information. Others (e.g., McNab & Klingberg, 2008; Vogel et al., 2005) have used visual change-detection tasks to measure WM. One of the goals of the present study, similar to prior work (e.g., Mall et al., 2014; Shipstead et al., 2014), was to measure individual differences in WM in a domain-general way, and to see how these difference relate to specific abilities.

Collectively, our results indicate that at least one component of individual differences in WM is the selective encoding and maintenance of relevant information. So these results are not consistent with the conclusions of Mall et al. (2014). Unfortunately, our specific set of results is difficult to directly compare with Mall et al. (2014) for several reasons. First, we did not record eye movements, so we cannot address how often high- and low-WM individuals fixated on irrelevant items (or item locations) during encoding and maintenance periods. Our stimulus durations and retention intervals were also shorter than those used by Mall et al. (2014). But our task was designed to more closely replicate those used by Vogel et al. (2005) and Shipstead et al. (2014). Second, we never tested uncued/irrelevant items, so we could not test memory for irrelevant information. However, we wanted to ensure all participants would use the filtering cue. Invalid cueing may lead some participants to ignore the cue and try to encode and maintain all items (e.g., Berryhill, Richmond, Shay, & Olson, 2012; Gözenman, Tanoue, Metoyer, & Berryhill, 2014; Matsukura, Luck, & Vecera, 2007; Williams & Woodman, 2012). Although our findings are at odds with Mall et al. (2014), it is hard to directly compare results across these two studies. But we do think Mall et al.'s (2014) approach—using independent measures of WM, analyzing eye movements, and testing irrelevant items—is a valid method that will need to be merged with that used in the present study in future work.

At first glance, our results are also at odds with those of Shipstead et al. (2014). However, a reanalysis of Shipstead et al. indicates that our results actually replicate those data relatively well. Shipstead et al. (2014) used two visual WM tasks that did not require a filtering component and two that did require filtering. Two of the tasks were very similar to those used in Experiments 2 and 3 of the present study, with two differences. Shipstead et al. (2014). used larger set sizes, and rather than indicating if one particular item changed orientation/color, participants had to report if *any* item in the array had changed orientation/color. In one task, participants were given either five or seven blue and red rectangles (nonfiltering). In the other task, participants were cued to the color of the to-be-tested item before each trial (filtering). Using the same regression technique as the present study, we found that a large portion of variance in the filtering orientation task was shared with the nonfiltering version (35%). But complex span accounted for an additional 3.5% of variance in filtering trials after controlling this shared variance. This replicates Experiments 2 and 3 in which WM span accounted for a small but significant portion of filtering trial variance independently of shared variance between filtering and nonfiltering trials. Further, if we run the same analysis on the other two visual WM tasks used by Shipstead et al. (2014), we observe the same pattern. On one task, participants were given patterns of colored squares and asked to indicate if a tested item changed color. On the filtering version of this task, participants were cued to one side of the screen prior to the trial, and items appeared on both sides of the screen. Participants were then asked to determine whether any item on the relevant side of the screen changed color. If we regress nonfiltering trial performance and WM span on filtering trial performance, we again find that a large portion of variance in filtering performance is shared with nonfiltering performance (25%). However, WM span accounts for a small but significant portion of variance (6.8%) in filtering trials, over and above the shared variance between filtering and nonfiltering trials. Our results are actually consistent with those of Shipstead et al. (2014). Further, Shipstead et al. examined shared variance in filtering across contexts using latent variable, whereas we examined filtering in one specific context in the present study. That is another reason why our findings might originally seem contradictory.

Overall, we argue that individual differences in WM are indeed predictive of filtering abilities, but this relationship carries an important caveat. The present results combined with prior work (Robison & Unsworth, 2017b; Shipstead et al., 2014) suggest that in order to observe a significant relationship between WM and filtering, the task must put a premium on the filtering requirement. This can be accomplished in at least two ways. First, filtering demands must change on a trial-by-trial basis. The fact that a significant effect of WM span was observed in Experiment 2 and in Shipstead et al. (2014) but not in Experiment 1 supports this idea. When the identity of targets and distractors is static across the task, the target/distractor relationship is continuously reinforced and no WM-related filtering differences exist. When the target/distractor relationship is not continuously reinforced, but rather requires a moment-to-moment updating of item relevance, WM-related filtering differences emerge. This is consistent with the general finding that WM-related differences emerge only when participants must continuously maintain task goals. When task goals are reinforced by prepotent responses or prior task experience, WM-related differences often disappear (Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2003). For example, Kane and Engle (2003) showed that high- and low-WM participants do not differ when presented with a high proportion of incongruent Stroop trials. When the task goal (i.e., report the color of the font of the word) was continually reinforced by the task, low-WM individuals did not have difficulty maintaining access to this goal. However, when incongruent Stroop trials were less frequent (25% of trials), low-WM individuals showed significantly greater Stroop interference. When the task did not continually reinforce the task goal, low-WM individuals frequently lost access to the goal, resulting in slower response times and more errors. Therefore, in order to observe such a relationship between WM and filtering, the filtering task must require participants to continually update and maintain the task goal (i.e., the identity of relevant items).

A second way to place a premium on filtering is by increasing set size. If set sizes are at or below capacity limits, participants may view the filtering cue as an unnecessary additional burden and instead try to remember all items. In a prior study, we embedded precues and retrocues into visual WM tasks to indicate relevant items either before or after encoding periods (Robison & Unsworth, 2017b). When set size was relatively small, precue and retrocue effects were also small, and the relationship between cueing effects and WM span was rather weak. Participants were nearly at ceiling, and interindividual variability was low. When we used a task similar to that in Experiment 3 (i.e., larger set size), the magnitude of the cueing effects increased, interindividual variability increased, and the relationship with WM span strengthened. Shipstead et al.'s (2014) orientation task contained five or seven targets and five or seven distractors. In our comparable task (Experiment 2), we had maximums of four targets and four distractors, but set sizes were sometimes as low as two items. The reanalysis of Shipstead et al. (2014) revealed a larger relationship between WM span and filtering than we did. Further, Experiment 3 of the present study had a larger set size than most trial types in Experiments 1 and 2, and we observed our largest filtering effects in Experiment 3. Therefore, any task manipulation that puts a premium on filtering (e.g., set size, dynamic item relevance) may give rise to greater systematic variation across participants which covaries with individual differences in WM. Mall et al. (2014) lengthened the durations of encoding periods and maintenance intervals. The manipulation of such durations may represent a third task factor that systematically affects the WM-filtering relationship. However, because Mall et al. (2014) is the only study to examine individual differences with such a manipulation, this finding begs future research.

We should further note that we do not argue that filtering is the primary reason why WM differences arise. Rather, filtering is one of several control processes that differs across individuals. Other control processes include the consistent deployment of attention toward the task and the resistance of attention deviations away from the task (e.g., mind wandering; Adam et al., 2015; Unsworth & Robison, 2016), active and selective maintenance of relevant information over delay intervals (Robison & Unsworth, 2017b; Unsworth & Robison, 2015; Vogel & Machizawa, 2004), and dealing with large amounts of information at one time (Fukuda et al., 2015).

We also acknowledge that the majority of variation in these abilities is shared across filtering and nonfiltering contexts, as nonfiltering trial performance accounted for the vast majority of variance in filtering trial performance in all three experiments. And admittedly, the residual portion of variance in filtering trials shared with WM span was small. However, a small effect does not necessarily mean it is not meaningful. For example, one of the most interesting aspects of WM is its ability to predict mind-wandering tendencies, both in and out of the laboratory. But this effect is quite small. Across a number of latent variable analyses, the latent correlation between WM and mind-wandering tendencies is usually between −.20 and −.30 (Kane et al., 2016; McVay & Kane, 2009; 2012a, 2012b; Robison, Gath, & Unsworth, 2017; Robison & Unsworth, 2015; Robison & Unsworth, 2017c; Unsworth & McMillan, 2013; Unsworth & McMillan, 2014). This suggests that WM and mind-wandering tendencies share about 4–9% of their variance. After accounting for shared variance between mind wandering and other constructs (e.g., attention control, motivation, etc.) the residual relationship is even smaller. This is just one example of when a relationship can be small but theoretically meaningful.

A secondary goal of the present study was to highlight some of the issues in the measurement of filtering. Mall et al. (2014) recognized several issues including the static nature of the target/distractor relationship and inherent dependencies between filtering and visual WM capacity estimates when measured during the same task. We agree that these are two important issues that affect WM-filtering relations. We also highlight the diversity of ways in which filtering can be measured. In the studies reviewed above, we noted several methods including the difference in CDA between filtering and nonfiltering trials (Vogel et al., 2005), the difference in performance between filtering and nonfiltering trials (Lee et al., 2010; Unsworth & Robison, 2016), fixations on irrelevant items/locations during encoding and maintenance (Mall et al., 2014), memory for irrelevant information (Mall et al., 2014), and the separation of shared variance between filtering and nonfiltering trials from shared variance between independent measures of WM and filtering (Shipstead et al., 2014). In the present study, we use an approach similar to Shipstead et al.'s (2014) latent variable investigation. We also demonstrated that using other methods of measuring filtering (e.g., computing a filtering cost as a difference/ratio) would have led to different patterns of results. Further, these filtering costs may not be reliable enough within participants to be used in between-subjects or correlational analyses. This is a problem that has been encountered by cognitive psychologists for decades. If an experimental effect is small (as is sometimes the case with filtering), there must be little interindividual variability in order to detect the effect. When subsequently trying to examine individual differences in such an effect, the lack of interindividual variability can be problematic. In an illustrative example, Salthouse, Siedlecki, and Krueger (2006) examined individual differences in memory control. Although the dependent measures (recall performance on various item lists) were all internally reliable, they were also highly correlated with one another. When Salthouse et al. (2006) computed difference scores as measures of individual differences, the scores all had low reliability estimates. Referring back to the equation for estimating the reliability of a difference score (Irwin, 1966), two factors affect the reliability of a difference between two measures: (a) the reliability of the individual measures, and (b) the correlation between the two measures. Similar to Salthouse et al. (2006), we observed acceptable reliability for individual measures (i.e., trial types), but these measures were also highly correlated with one another. Thus, when we estimate the reliability of a difference in accuracy between two trial types (i.e., a filtering cost), the reliability estimate is quite low. Ways of dealing with the unreliability of difference scores as measures of individual differences include examining part and partial correlations (Horn, 1963) and confirmatory factor analysis (Donaldson, 1983). The logic of our regression analyses closely follows these two techniques, which account for common and unique variance among sets of variables.

Conclusion

One manifestation of individual differences in WM is the ability to selectively encode and maintain relevant information in the presence of irrelevant information (i.e., filtering). However, this relationship may only be observed when the task places a premium on filtering in order to maximize performance. When individuals must update the relevance of incoming information on a moment-to-moment basis, and when set sizes exceed WM capacity limits, individual differences in WM are predictive of filtering abilities. Although the majority of variation in visual WM tasks is shared across filtering and nonfiltering contexts, filtering represents a distinct WM-related form of control that manifests in contexts when it is especially required.

References

Adam, K. C. S., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The contribution of attentional lapses to individual differences in visual working memory capacity. *Journal of Cognitive Neuroscience, 27,* 1601–1616. http://dx.doi.org/10.1162/jocn_a_00811

Adam, K. C. S., & Vogel, E. K. (2017). Confident failures: Lapses of working memory reveal a metacognitive blind spot. *Attention, Percep-*

tion, & Psychophysics, 79, 1506–1523. http://dx.doi.org/10.3758/s13414-017-1331-8

Awh, E., & Vogel, E. K. (2008). The bouncer in the brain. *Nature Neuroscience, 11,* 5–6. http://dx.doi.org/10.1038/nn0108-5

Awh, E., Vogel, E. K., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience, 139,* 201–208. http://dx.doi.org/10.1016/j.neuroscience.2005.08.023

Berryhill, M. E., Richmond, L. L., Shay, C. S., & Olson, I. R. (2012). Shifting attention among working memory representations: Testing cue type, awareness, and strategic control. *The Quarterly Journal of Experimental Psychology, 65,* 426–438. http://dx.doi.org/10.1080/17470218.2011.604786

Cowan, N., & Morey, C. C. (2006). Visual working memory depends on attentional filtering. *Trends in Cognitive Sciences, 10,* 139–141. http://dx.doi.org/10.1016/j.tics.2006.02.001

Donaldson, G. (1983). Confirmatory factor analysis models of information processing stages: An alternative to difference scores. *Psychological Bulletin, 94,* 143–151. http://dx.doi.org/10.1037/0033-2909.94.1.143

Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *The Journal of Neuroscience, 29,* 8726–8733. http://dx.doi.org/10.1523/JNEUROSCI.2145-09.2009

Fukuda, K., & Vogel, E. K. (2011). Individual differences in recovery time from attentional capture. *Psychological Science, 22,* 361–368. http://dx.doi.org/10.1177/0956797611398493

Fukuda, K., Woodman, G. F., & Vogel, E. K. (2015). Individual differences in visual working memory capacity: Contributions of attentional control to storage. In P. Jolicœur, C. Lefebvre, & J. Martinez-Trujillo (Eds.), *Mechanisms of sensory working memory: Attention and performance XXV* (pp. 105–119). Philadelphia, PA: Academic Press. http://dx.doi.org/10.1016/B978-0-12-801371-7.00009-0

Gaspar, J. M., Christie, G. J., Prime, D. J., Jolicœur, P., & McDonald, J. J. (2016). Inability to suppress salient distractors predicts low visual working memory capacity. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 3693–3698. http://dx.doi.org/10.1073/pnas.1523471113

Gözenman, F., Tanoue, R. T., Metoyer, T., & Berryhill, M. E. (2014). Invalid retro-cues can eliminate the retro-cue benefit: Evidence for a hybridized account. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 1748–1754. http://dx.doi.org/10.1037/a0037474

Horn, J. L. (1963). Equations representing combinations of components in scoring psychological variables. *Acta Psychologica, 21,* 184–217.

Irwin, D. O. (1966). Reliability of the Wechsler Intelligence Scale for Children. *Journal of Educational Measurement, 3,* 287–292. http://dx.doi.org/10.1111/j.1745-3984.1966.tb00891.x

Jost, K., & Mayr, U. (2016). Switching between filter settings reduces the efficient utilization of visual working memory. *Cognitive, Affective & Behavioral Neuroscience, 16,* 207–218. http://dx.doi.org/10.3758/s13415-015-0380-5

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130,* 169–183. http://dx.doi.org/10.1037/0096-3445.130.2.169

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132,* 47–70. http://dx.doi.org/10.1037/0096-3445.132.1.47

Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of Experimental Psychology: General, 145,* 1017–1048. http://dx.doi.org/10.1037/xge0000184

Lee, E.-Y., Cowan, N., Vogel, E. K., Rolan, T., Valle-Inclán, F., & Hackley, S. A. (2010). Visual working memory deficits in patients with Parkinson's disease are due to both reduced storage capacity and impaired ability to filter out irrelevant information. *Brain: A Journal of Neurology, 133,* 2677–2689. http://dx.doi.org/10.1093/brain/awq197

Mall, J. T., Morey, C. C., Wolff, M. J., & Lehnert, F. (2014). Visual selective attention is equally functional for individuals with low and high working memory capacity: Evidence from accuracy and eye movements. *Attention, Perception, & Psychophysics, 76,* 1998–2014. http://dx.doi.org/10.3758/s13414-013-0610-2

Matsukura, M., Luck, S. J., & Vecera, S. P. (2007). Attention effects during visual short-term memory maintenance: Protection or prioritization? *Perception & Psychophysics, 69,* 1422–1434. http://dx.doi.org/10.3758/BF03192957

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience, 11,* 103–107. http://dx.doi.org/10.1038/nn2024

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 196–204.

McVay, J. C., & Kane, M. J. (2012a). Drifting from slow to "D'oh!": Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 525–549. http://dx.doi.org/10.1037/a0025896

McVay, J. C., & Kane, M. J. (2012b). 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. http://dx.doi.org/10.1037/a0025250

Mrazek, M. D., Smallwood, J., Franklin, M. S., Chin, J. M., Baird, B., & Schooler, J. W. (2012). The role of mind-wandering in measurements of general aptitude. *Journal of Experimental Psychology: General, 141,* 788–798. http://dx.doi.org/10.1037/a0027968

Robison, M. K., Gath, K. I., & Unsworth, N. (2017). The neurotic wandering mind: An individual differences investigation of neuroticism, mind-wandering, and executive control. *The Quarterly Journal of Experimental Psychology, 70,* 649–663. http://dx.doi.org/10.1080/17470218.2016.1145706

Robison, M. K., & Unsworth, N. (2015). Working memory capacity offers resistance to mind-wandering and external distraction in a context-dependent manner. *Applied Cognitive Psychology, 29,* 680–690. http://dx.doi.org/10.1002/acp.3150

Robison, M. K., & Unsworth, N. (2017a). Individual differences in working memory capacity predict learned control over attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 43,* 1912–1924.

Robison, M. K., & Unsworth, N. (2017b). Variation in the use of cues to guide visual working memory. *Attention, Perception, & Psychophysics, 79,* 1652–1665. http://dx.doi.org/10.3758/s13414-017-1335-4

Robison, M. K., & Unsworth, N. (2017c). Working memory capacity and mind-wandering during low-demand cognitive tasks. *Consciousness and Cognition: An International Journal, 52,* 47–54. http://dx.doi.org/10.1016/j.concog.2017.04.012

Robison, M. K., & Unsworth, N. (in press). Cognitive and contextual correlates of spontaneous and deliberate mind-wandering. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Salthouse, T. A., Siedlecki, K. L., & Krueger, L. E. (2006). An individual differences analysis of memory control. *Journal of Memory and Language, 55,* 102–125. http://dx.doi.org/10.1016/j.jml.2006.03.006

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Mem-*

ory and Language, 72, 116–141. http://dx.doi.org/10.1016/j.jml.2014.01.004

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2015). Working memory delay activity predicts individual differences in cognitive abilities. *Journal of Cognitive Neuroscience, 27,* 853–865. http://dx.doi.org/10.1162/jocn_a_00765

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498–505. http://dx.doi.org/10.3758/BF03192720

Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: Examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 832–842. http://dx.doi.org/10.1037/a0029669

Unsworth, N., & McMillan, B. D. (2014). Similarities and differences between mind-wandering and external distraction: A latent variable analysis of lapses of attention and their relation to cognitive abilities. *Acta Psychologica, 150,* 14–25. http://dx.doi.org/10.1016/j.actpsy.2014.04.001

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17,* 635–654. http://dx.doi.org/10.1080/09658210902998047

Unsworth, N., & Robison, M. K. (2015). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychonomic Bulletin & Review, 22,* 757–765. http://dx.doi.org/10.3758/s13423-014-0747-6

Unsworth, N., & Robison, M. K. (2016). The influence of lapses of attention on working memory capacity. *Memory & Cognition, 44,* 188–196. http://dx.doi.org/10.3758/s13421-015-0560-0

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428,* 748–751. http://dx.doi.org/10.1038/nature02447

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. http://dx.doi.org/10.1038/nature04171

Williams, M., & Woodman, G. F. (2012). Directed forgetting and directed remembering in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1206–1220. http://dx.doi.org/10.1037/a0027389

Appendix

Additional analyses

Table A1
Alternative Regression Analysis for Experiment 1

Step	R^2	β	SE_β	t	p
Step 1	.03				
Distractor-absent		.16	.08	2.03	.044
Step 2	.03				
Distractor-absent		.08	.12	.69	.49
Distractor-present		.11	.12	.90	.37

Note. $N = 158$. Dependent variable: working memory span. β = standardized beta coefficient; SE_β = standard error of standardized beta coefficient.

Table A2
Alternative Regression Analysis for Experiment 2

Step	R^2	β	SE_β	t	p
Step 1	.12				
Distractor-absent		.35	.07	4.28	<.001
Step 2	.15				
Distractor-absent		.10	.14	.69	.49
Distractor-present		.30	.14	2.13	.04

Note. $N = 137$. Dependent variable: working memory span. β = standardized beta coefficient; SE_β = standard error of standardized beta coefficient.

(*Appendix continues*)

Table A3
Alternative Regression Analysis for Experiment 3

Step	R^2	β	SE_β	t	p
Step 1	.05				
Neutral		.23	.07	2.93	<.001
Step 2	.10				
Neutral		.07	.09	.75	.45
Precue		.27	.09	2.89	.004

Note. $N = 158$. Dependent variable: working memory span. β = standardized beta coefficient; SE_β = standard error of standardized beta coefficient.

Received May 4, 2017
Revision received October 25, 2017
Accepted October 28, 2017 ■