



Sequential encoding paradigm reliably captures the individual differences from a simultaneous visual working memory task

Chong Zhao^{1,2} · Edward K. Vogel^{1,2}

Accepted: 21 December 2022 / Published online: 9 January 2023
© The Psychonomic Society, Inc. 2023

Abstract

Converging behavioral and neural evidence have suggested that visual stimuli could be attached to existing visual working memory sequentially in time. However, whether individual differences in sequential visual working memory paradigm are similar to those measured by the classical simultaneous change detection paradigm remain unknown. Here, we first show that sequentially presented visual stimuli exhibit similar working memory capacity bottlenecks as previous research using simultaneously presented items. We further reveal that within the same subject, the accuracy and capacity estimates using sequential and simultaneous paradigm were comparable across four different set sizes. Also, we discover that the individual differences measured by the sequential paradigm were highly correlated to those by the simultaneous paradigm within the same subject across all four set sizes of interest. Finally, we find that in a larger sample of subjects ($n = 200$), the variance and higher-order statistics were similar for sequential and simultaneous paradigms with set size of 6. Collectively, these findings indicate that individual differences measured by the sequential presentation of visual items rely on the similar working memory resources as those by the simultaneous form of presentation.

Keywords Individual differences · Visual working memory · Sequential paradigm

Introduction

Individual differences in visual working memory (VWM) capacity predict a wide range of cognitive abilities, such as general intelligence (Unsworth et al., 2014), long-term memory encoding (Miller et al., 2019), and even mathematical reasoning (Raghubar et al., 2010), to name a few. Individual differences in VWM performance are thought to connect to such a wide range of cognitive performance measures because they have high demands for attentional control (Kane et al., 2004; Kyndt et al., 2012; Martin et al., 2021). One broadly used paradigm to determine these capacity estimates is the change detection paradigm (Luck & Vogel, 1997; Rouder et al., 2011). In these experiments, subjects first encoded an array of simultaneously appearing colored squares. During the test phase, subjects are probed by a color square appearing at one of the original locations and

determine whether the square changed its color or not compared with the encoding phase. However, in this paradigm, multiple objects must be encoded into memory simultaneously which may produce high levels of competition for attentional control processes necessary for successful storage. This may result in only a subset of the stimuli being encoded and maintained in the capacity-limited VWM (Zhang & Luck, 2008). Neural signatures also reflect this limit in VWM capacity. In a simultaneous change detection experiment, the contralateral delay activity (CDA) would increase with load until the VWM capacity limit was reached (Vogel & Machizawa, 2004). Crucially, Vogel and Machizawa (2004) also revealed that the CDA increase from Set Size 2 to Set Size 4 was correlated to the individual differences in VWM capacity. Further, the individual differences observed in both simultaneous change detection performance and CDA amplitudes have been shown to have high reliability and predict performance in a wide variety of cognitive measures (Fukuda et al., 2015; Fukuda & Vogel, 2011). Collectively, this evidence suggests that the simultaneous change detection paradigm is an ideal candidate task for studying individual differences in cognitive ability. However, one significant ambiguity remains with regards to the individual differences measured in this task. Is the variability across subjects primarily driven by the competition for attentional control amongst objects during the simultaneous encoding phase? Or are these differences the driven by demands

✉ Chong Zhao
chongzhao@uchicago.edu

¹ Department of Psychology, University of Chicago,
Chicago, IL 60637, USA

² Institute for Mind and Biology, University of Chicago,
Chicago, IL 60637, USA

that are incurred at later stages of processing (e.g., storage, retrieval, decision)?

There have been several studies that have used variations of the change detection paradigm in which the items were presented sequentially. The primary modification in these studies was that each stimulus was presented one at a time during the encoding phase until the set size was reached. In these sequential tasks, competition at encoding is minimized because each item is presented in isolation. Average performance in these sequential paradigms seems to show similar overall performance the simultaneous change detection paradigm. For example, Woodman, Vogel, and Luck (Woodman et al., 2012) showed that the mean performance was similar in the sequential and the simultaneous versions of the change detection paradigm with Set Sizes 2, 4, and 6. Further, neural evidence also showed that sequentially attaching items into VWM would similarly boost CDA until the time when working memory capacity was full (Berggren & Eimer, 2016; Feldmann-Wüstefeld et al., 2018; Ikkai et al., 2010; Vogel et al., 2005; Wang et al., 2018). Therefore, one possibility was that sequential encoding and simultaneous encoding tested on the same underlying VWM ability construct.

However, none of these prior studies attempted to determine whether the two tasks captured the same individual differences in performance across subjects. That is, the observation of similar mean performance in either the behavior or the neural measure does not necessarily mean that the performance variance captured in one task is the same as what was observed in the other task. For example, it is possible that both tasks are limited by a common storage limit, but that the individual differences in performance are only observed when there is competition at encoding. Such a scenario could plausibly produce similar mean performance without capturing the same individual differences across subjects. In sequential tasks, only one item is attended at a time (Olivers et al., 2011), and thus may be much less dependent upon the attention control ability of the individual during encoding. Alternatively, it is plausible that the individual differences in performance measured in simultaneous tasks are driven by demands for control processes that are not exclusively restricted to competition during encoding. For example, during sequential tasks, attentional control demands may become necessary as the accumulated load increases and nears or exceeds capacity limits or perhaps due to competition amongst items during the retrieval stage.

Our current study aimed to directly test if the individual differences in sequentially storing items into VWM draws upon the same limitations as those observed when simultaneously storing items into VWM. To address this question, we first wanted to establish the parameters of a benchmark sequential change detection task to determine whether a well-powered sequential study would replicate the overall levels of performance observed in previous literature, which had limited numbers of trials and subjects (e.g., Woodman et al., 2012). To test

if the two modes of encoding shared their variance, we then tested if the within-subject capacity measures correlated between sequential and simultaneous encoding paradigms across four different set sizes.

Overview of experiments

We measured the individual differences in sequential change detection performance and its relationship to its simultaneous equivalent. In Experiment 1, we measured the accuracy and K capacity in only the sequential paradigm in Set Size 3, 4, 5, and 6 with large number of trials ($t = 240$). In Experiment 2, we administered the simultaneous and sequential task with the same total set size to subjects to ask if the relation between the two paradigms change with respect to set size. This resulted in four groups of subjects for each set size (3–6) and 24 subjects for each set size. In Experiment 3a, we aimed to determine how the individual differences in high set size (6) between two modes of encoding were correlated with a large sample size ($n = 200$) and trial number ($t = 240$). In Experiment 3b, we aimed to generalize our findings in individual differences between the two paradigms into two different set sizes (4 and 6) within the same subject.

Experiment 1

Method

Subjects

Forty subjects were recruited at the rate of \$9.50 per hour from Prolific, an online platform for subject recruitment. All subjects were 18 to 35 years old, currently lived in the United States, had normal or corrected-to-normal vision, and with no ongoing psychological or neurological disorders. Our screening for Prolific online subjects remained the same as our in-lab sample, such that participants who had below chance (or capacity score below zero) would be eliminated. Additionally, participants who finished less than 80% of the trials, or failed more than 25% of our attentional check questions, would also be eliminated.

Stimuli

All stimuli were colored squares generated in JavaScript using the jsPsych canvas keyboard interface. The colored squares were all in 40×40 pixels size on a 400×400 pixels canvas page. Color squares could appear anywhere within a circular area of the monitor within 30 to 200 pixels from the center of the canvas screen. Each square could appear in one of the nine distinct colors with no repetitions within any trial (RGB

values: red = 255, 0, 0; green = 0, 255, 0; blue = 0, 0, 255; magenta = 255, 0, 255; yellow = 255, 255, 0; cyan = 0, 255, 255; orange = 255, 128, 0; white = 255, 255, 255; black = 0 0 0). Subjects were instructed to fixate at a small black plus (30 px in Arial) at the center of the screen throughout the trial.

Procedures

In pilot studies, we ran several versions of the sequential paradigms and determined that 250 ms was enough for encoding but would not enable extensive rehearsal in a sequential paradigm. Furthermore, our pilot experiments also revealed that 250 ms of interstimulus interval would be enough to break the remaining imagery of the previous stimuli and encourage more of the VWM usage. Therefore, in our finalized experimental design, each trial in the sequential paradigm started with a 500-ms fixation plus at the center of the screen. Then, the subjects viewed three, four, five or six colored squares one at a time each for 250 ms. Following each color square, 250 ms of fixation period was added to ensure that the perceptual continuity of the previous color squares did not bring to the next frame. A retention period of 1,000 ms was displayed following the memoranda, before when one test color square would appear at one of the previous locations of the studied color squares (see Fig. 1a). During half of the trials, the test item would have the same color as the studied color square. In the

other half of trials, the test item would change to one of the new colors that was not sampled during the study phase. In each trial, the subjects were asked to respond change when the test item had a different color as the studied item, or no change, when the test item was the same color as one of the studied items, to the test color square. The serial order of which study item was selected as a test item was randomized within subject. For instance, in a Set Size 3 setting, the probability of the first, second, or third item being test would each be 1/3. To eliminate the confounding effect of test order, half of the subjects finished five blocks of the low set size trials in their odd numbered blocks, and high set size trials in their even numbered blocks. Consequently, the other half of subjects had five blocks of the low set size trials in the even number of blocks and high set size trials in the odd number of blocks.

Due to the concern that online test might be harder for subjects to maintain sustained attention compared with in-lab test, the study was designed to be completed in under an hour. To reduce the time each subject spent in doing the task, our four set sizes of interest were divided into three and five, or four and six items, between two groups of subjects (see Fig. 2). Therefore, 20 subjects finished 240 trials each of Set Sizes 3 and 5 sequential change detection task, and the other 20 subjects finished 240 trials each of Set Sizes 4 and 6 sequential change detection task.

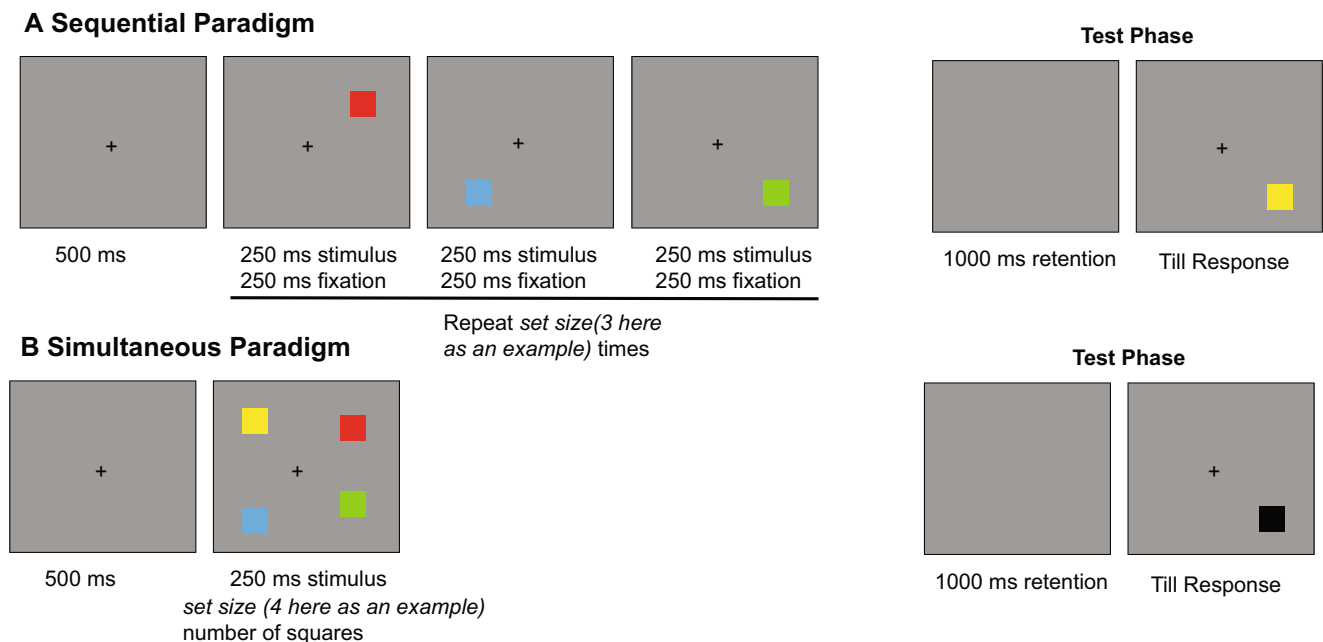


Fig. 1 The change detection paradigm. **A** A sample of the sequential change detection paradigm. The color squares appeared one at a time and repeated until the trial reached its set size. **B** A sample of the

simultaneous change detection paradigm. Colored squares appeared on the screen simultaneously and the number of colored squares was determined by the set size of the trial

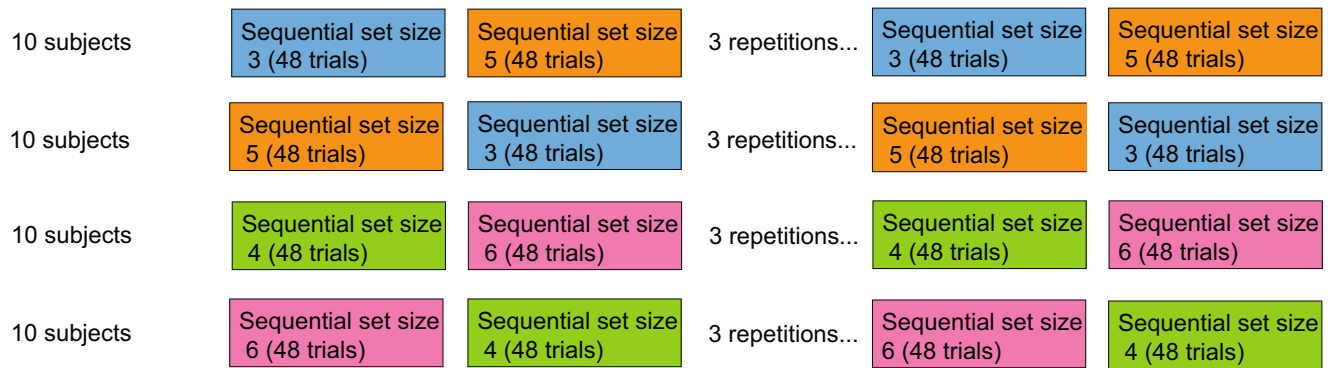


Fig. 2 Sequential Change detection participant assignment in Experiment 1. Four groups of 10 unique subjects were recruited. Half of them finished set size 3 and 5, and the other half finished set size 4 and 6, with block order counterbalanced and 240 trials per set size

Result

Accuracy and K capacity of the sequential paradigm

The accuracy patterns in our sequential change detection task were highly similar to these seen in simultaneous change detection tasks. The accuracy decreased as set size increased. Set Size 3 accuracy was significantly higher than that of Set Size 4, $t(19) = 3.72, p = 6.34 \times 10^{-4}$, Cohen’s $d = 1.26$, and Set Size 4 was significantly higher in accuracy than Set Size 6, $t(19) = 9.63, p = 9.59 \times 10^{-9}$, Cohen’s $d = 1.73$ (see Fig. 3). The only exception was that the accuracy of Set Size 5 was not different from that of Set Size 4, $t(19) = 0.28, p = .77$. All of our conclusions held after we performed Bonferroni correction for multiple comparison, with a new p threshold at .008. We then calculated the K score capacity according to Cowan’s one probe change detection equation, $K = N \times (H-F)$ (Cowan, 2001). The K capacity of Set Size 3 was not different from that of Set Size 4, $t(19) = 1.48, p = .14$, Cohen’s $d = 0.48$, or Set Size 6, $t(19) = 1.34, p = .19$, Cohen’s $d = 0.44$. Similarly, the only exception was that Set Size 5 had a higher K than that of Set Size 3, $t(19) = 6.27, p = 5.06 \times 10^{-6}$, Cohen’s $d = 1.44$, and that of Set Size 4, $t(19) = 2.66, p = .01$, Cohen’s $d = 0.87$. The fact that Set Size 5 had a K higher than Set Size

3 was mainly due to the within-subject paired t test employed in comparing the results. Moreover, in the comparison between Set Size 5 and Set Size 4, the p value would not survive the correction of multiple comparison, since the new p value was .008 here for significance. Therefore, we concluded that as set size increased from 3 to 6 in a sequential paradigm, the pattern of accuracy and K capacity was similar to what has been previously observed in simultaneous change detection paradigms (Xu et al., 2018).

Reliability of the sequential paradigm

One key advantage of the simultaneous paradigm was that it was highly reliable even with a moderate number of trials and subjects (Xu et al., 2018). To ensure the internal reliability of this 240-trial sequential paradigm, we computed the Spearman–Brown corrected split-half reliability using even-odd split of the trials in each set size. The accuracy measure was highly reliable for all four set sizes (Set Size 3: 0.944; Set Size 4: 0.941; Set Size 5: 0.937; Set Size 6: 0.884). Similarly, the K capacity measure was highly reliable across all set sizes of interest in our study (Set Size 3: 0.949; Set Size 4: 0.939; Set Size 5: 0.937; Set Size 6: 0.880). Collectively, these

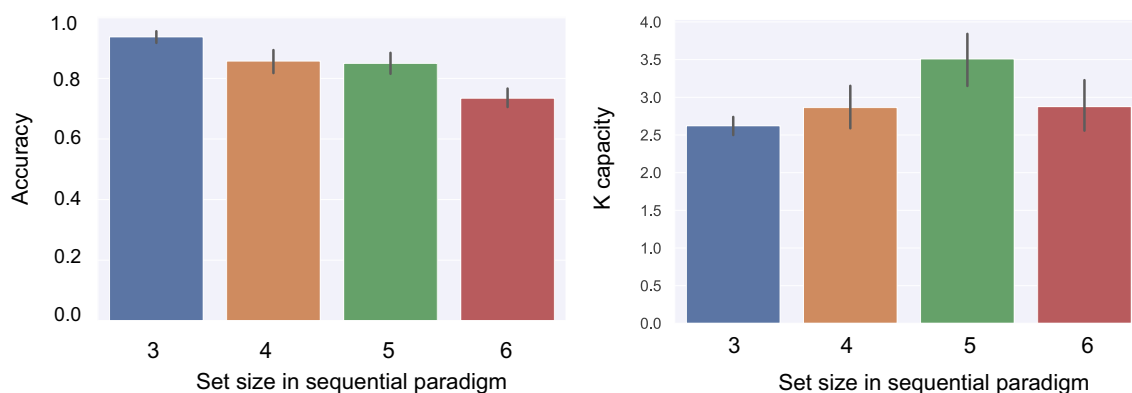


Fig. 3 Sequential Change detection performance in Experiment 1. (Left) The accuracy of sequential change detection with respect to different set sizes. (Right) The K capacity score with respect to different set sizes

evidences suggested that sequential change detection paradigm was highly reliable within each subject.

Discussion

In Experiment 1, we revealed that the sequential paradigm had high reliability across all four set sizes of our interest. Specifically, numbers of trials and subjects needed to reach optimal reliability were comparable to that of the simultaneous change detection paradigm shown in previous research (Xu et al., 2018). Meanwhile, the capacity limit measured by the sequential paradigm was also similar to that in previous work by the simultaneous paradigm (Luck & Vogel, 1997). However, such face similarity itself was not enough to testify if the sequential and the simultaneous paradigm were reflecting the same inherent working memory ability of subjects.

To better quantify the relationship between the simultaneous and sequential paradigm within subject, in Experiment 2, we recruited subjects to complete both versions of the change detection task with the same set size in the same session. This design aimed to directly test if the sequential and simultaneous change detection tasks were relying on similar cognitive resources across various set sizes.

Experiment 2

Method

Subjects

Ninety-six subjects were recruited at the rate of \$9.50 per hour from Prolific. All subjects were 18 to 35 years old, currently lived in the United States, had normal or corrected-to-normal vision, and with no ongoing psychological or neurological disorders. Our screening for Prolific online subjects remained the same as our in-lab sample, such that participants who had below chance (or capacity score below zero) would be eliminated. Additionally, participants who finished less than 80% of the trials, or failed more than 25% of our attentional check questions, would also be eliminated.

Stimuli

The colors and size of the color squares and fixation plus were the same as the ones used in Experiment 1.

Procedures

Each subject was randomly assigned to one of the four set sizes of interest and complete 240 trials of sequential and 240 trials of simultaneous change detection task (see Fig. 4

for subject assignment). The sequential task was the same as the one used in Experiment 1. In the simultaneous paradigm, each trial started with a 500-ms fixation plus at the center of the screen. Then, the subjects viewed three, four, five or six colored squares simultaneously for 250 ms. A retention period of 1,000 ms was displayed following the memoranda, before when one test color square would appear at one of the previous locations of the studied color squares (see Fig. 1b). During half of the trials, the test item would have the same color as one of the studied color squares. In the other half of trials, the test item would change to one of the new colors that was not sampled during the study phase. In each trial, the subjects were asked to respond change when the test item had a different color as the studied item, or no change, when the test item was the same color as one of the studied items, to the test color square. To eliminate the confounding effect of test order, half of the subjects finished five blocks of the simultaneous trials in their odd numbered blocks, and sequential trials in their even numbered blocks. Consequently, the other half of subjects had five blocks of the simultaneous trials in the even number of blocks and sequential trials in the odd number of blocks.

Twenty-four unique subjects each participated in one set size of both the sequential and simultaneous tasks. Our four set sizes of interests thus resulted in a total of 96 subjects in Experiment 2.

Result

Accuracy and K capacity for simultaneous task

The accuracy of the simultaneous task decreased with increasing set size, with Set Size 3 accuracy higher than that of Set Size 4, $t(23) = 6.61$, $p = 3.50 \times 10^{-8}$, Cohen's $d = 2.04$, and Set Size 5 accuracy higher than that of Set Size 6, $t(23) = 2.65$, $p = .01$, Cohen's $d = 0.74$. The only exception was at Set Size 5, where the accuracy was not different from that of Set Size 4, $t(23) = 1.28$, $p = .21$, Cohen's $d = 0.37$. To acquire a more accurate measure by accounting for false alarm rate, Cowan's K was compared with all set sizes, and none of the pairwise comparison reached significance ($ps > .07$). Therefore, we concluded that the capacity measure of Set Sizes 3, 4, 5, and 6 was not different from each another. All of our conclusions held after we performed Bonferroni correction for multiple comparison, with a new p threshold at .008.

Accuracy and K capacity for sequential task

The accuracy of the simultaneous task decreased with increasing set size, with Set Size 3 accuracy higher than that of Set Size 4, $t(23) = 4.70$, $p = 2.36 \times 10^{-5}$, Cohen's $d = 1.44$, and Set Size 5 accuracy higher than that of Set Size 6, $t(23) = 3.25$, $p = 2.17 \times 10^{-3}$, Cohen's $d = 1.10$ (see Fig. 5). The only exception

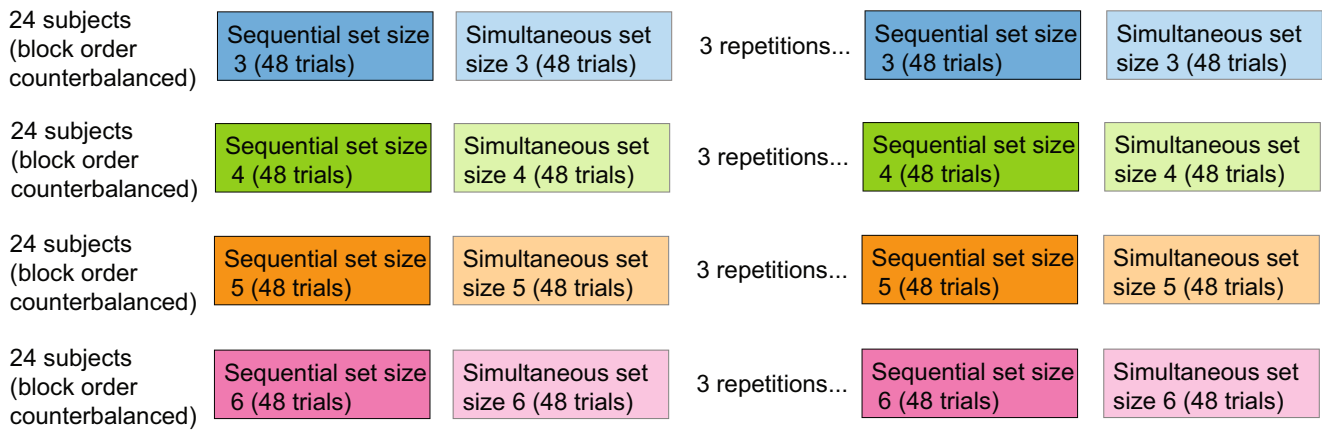


Fig. 4 Sequential and Simultaneous Change detection participant assignment in Experiment 2. Four groups of 24 unique subjects were recruited. Each one group of them finished one of the 4 set sizes of

sequential and simultaneous tasks. The order of simultaneous black first or sequential black first was counterbalanced. Each participant finished 240 simultaneous and 240 sequential trials with the same set size in total

was at Set Size 5, where the accuracy was not different from that of Set Size 4, $t(23) = 0.30, p = .77$, Cohen’s $d = 0.08$. The accuracy results in sequential task highly mirrored that in the simultaneous task. All of our conclusions held after we performed Bonferroni correction for multiple comparison, with a new p threshold at .008. To acquire a more accurate measure by accounting for false-alarm rate, Cowan’s K was compared with all set sizes, and only one of the pairwise comparison reached significance (Set Size 3 versus 5, $p = .007$). Similarly, all of our conclusions held after we performed Bonferroni correction for multiple comparison, with a new p threshold at .008. Therefore, we concluded that in general, the capacity measure of Set Size 3, 4, 5, and 6 was not different from each another.

Meanwhile, the pattern of capacity limit as set size increased was also similar in the sequential paradigm to that in the simultaneous paradigm. However, such qualitative observation is not sufficient to prove that the simultaneous and sequential paradigm rely on the same pool of working memory resources. To better investigate whether the individual differences in working memory capacity were reflected similarly in both paradigms, correlation analyses were performed on the accuracy and capacity measures in each set size of interest. Across all four set sizes, the correlation between the accuracy in sequential paradigm and that in simultaneous paradigm was constantly high, Set Size 3: $r(23) = .690, p = 1.92 \times 10^{-4}$; Set Size 4: $r(23) = .726, p = 5.85 \times 10^{-5}$; Set Size 5: $r(23) = .751, p = 2.39 \times 10^{-5}$; Set Size 6: $r(23) = .745, p = 2.97 \times 10^{-5}$. Similarly, the K capacity correlation between simultaneous and sequential was also constantly significant for all set sizes of our interest, Set Size 3: $r(23) = .717, p = 8.03 \times 10^{-5}$; Set Size 4: $r(23) = .739, p = 3.69 \times 10^{-5}$; Set Size 5: $r(23) = .735, p = 4.33 \times 10^{-5}$; Set Size 6: $r(23) = .758, p = 1.78 \times 10^{-5}$.

Correlation between simultaneous and sequential VWM capacity within subject

As shown above, the capacity estimate of sequential paradigm was similar to that of the simultaneous paradigm.

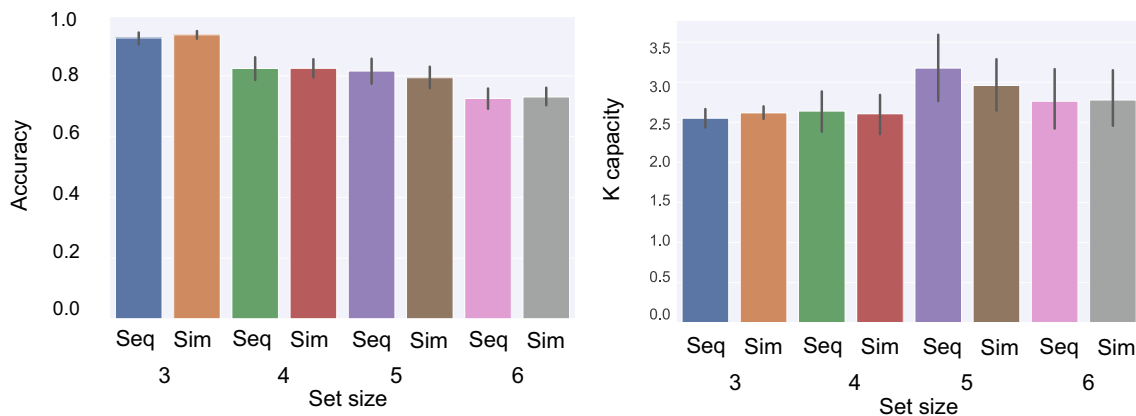


Fig. 5 Simultaneous and Sequential Change detection performance in Experiment 2. (Left) The accuracy of sequential and simultaneous change detection with respect to different set sizes. Each participant completed

both simultaneous and sequential versions of 240 trials each in one of the four set sizes. (Right) The K capacity score of sequential and simultaneous change detection with respect to different set sizes

Discussion

In Experiment 2, we concluded that the simultaneous paradigm and the sequential paradigm reflected similar limits in capacity within subject. The K capacity measures did not change as set sizes increased, resulting in the same bottleneck of capacity as in previous research using merely simultaneous paradigm (Luck & Vogel, 1997). Additionally, the individual differences in the sequential paradigm were highly correlated to those in the simultaneous paradigm. These evidence collectively suggested that the two paradigms, despite of having different modes of input, may largely reflect a similar pool of VWM resources.

One potential limitation of Experiment 2 was that the correlation between simultaneous and sequential VWM paradigm was drawn with plenty of trials, but a relatively small sample. The limitation in sample size may prevent us from observing the full range of individual differences in VWM ability. Therefore, we conducted Experiment 3 below, where Set Size 6, the load with the highest range of individual differences in performance, of both sequential and simultaneous designs were assigned to 200 unique subjects. We aimed to test the level of similarity in the statistical properties of the simultaneous and sequential paradigm. Meanwhile, we were also curious about the reliability of their correlation with a smaller sample or number of trials using the iterative down-sampling approach.

Experiment 3a

Method

Subjects

One hundred subjects were initially recruited at the rate of \$9.50 per hour from Prolific. After a round of reviews, an additional hundred subjects were recruited to address reviewer concerns such that the final sample size was 200. All subjects were eighteen to thirty-five years old, currently lived in the United States, had normal or corrected-to-normal vision, and with no ongoing psychological or neurological disorders. Our screening for Prolific online subjects remained the same as our in-lab sample, such that participants who had below chance (or capacity score below zero) would be eliminated. Additionally, participants who finished less than 80% of the trials, or failed more than 25% of our attentional check questions, would also be eliminated.

Stimuli

The colors and size of the color squares and fixation plus were the same as the ones used in Experiment 1.

Procedures

To maximize the individual differences between subjects, each subject would complete 240 trials of sequential and 240 trials of simultaneous change detection task with Set

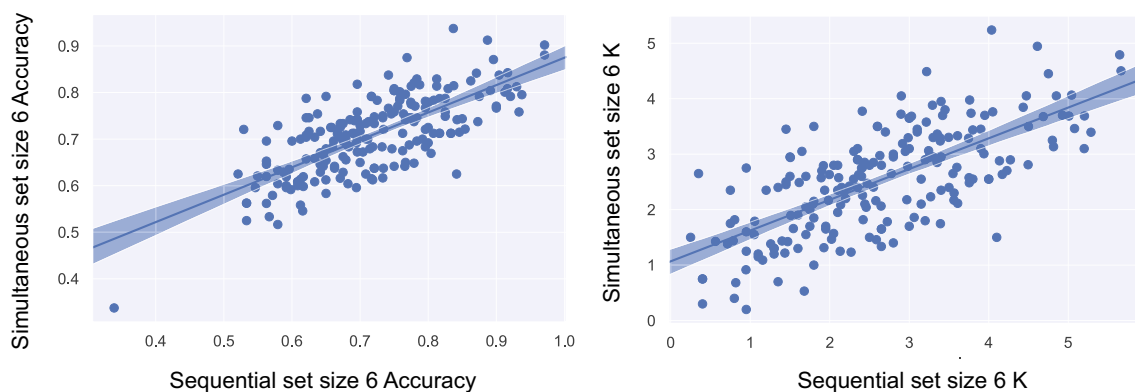


Fig. 6 Correlation between set size 6 simultaneous and sequential change detection performance in Experiment 3a. **Left** The correlation between set size 6 simultaneous and sequential change detection accuracies. **Right**

The correlation between set size 6 simultaneous and sequential change detection K capacities

Table 1 The statistical moments of the sequential and simultaneous K scores in Experiment 3a

K capacity estimates	Mean	Variance	Skewness	Kurtosis
Sequential Set Size 6	2.923	1.467	0.206	-0.601
Simultaneous Set Size 6	2.646	0.965	0.201	-0.405

Size 6. The sequential task was the same as the Set Size 6 one used in Experiments 1 and 2. The simultaneous task was the same as the Set Size 6 one used in Experiment 2.

Result

Relation between simultaneous and sequential VWM capacity within subject

As a first pass, we examined how reliable the simultaneous and sequential paradigm were with 100 subjects and 240 trials. Therefore, we did an even-odd split-half reliability test of the simultaneous and sequential paradigm. Both the simultaneous, $r(199) = .800$, $p = 2.20 \times 10^{-45}$, and sequential, $r(199) = .885$, $p = 7.42 \times 10^{-68}$, paradigm show a significant Pearson correlation between their even-number and odd-number trials. Given the high reliability of both paradigms, we then tested whether the individual differences in sequential paradigm were correlated to these in simultaneous paradigm. We further collected 100 more subjects with 240 trials each of the paradigm. Compiling the two datasets with 200 subjects in total, we discovered that the correlation between the accuracies of the two paradigms was highly significant, $r(199) = .725$, $p = 7.43 \times 10^{-34}$, as well as the K measures in both tasks, $r(199) = .698$, $p = 1.55 \times 10^{-30}$ (see Fig. 6). Lastly, we examined the first four orders of statistical moments of the K capacity scores in sequential and simultaneous paradigms within each subject. The resulting values were numerically similar for the two measures (see Table 1). In particular, similar variance, skewness and kurtosis on K score collectively indicated that the two measures were likely to reflect similar levels of differences across individuals when measuring VWM ability.

Experiment 3b

Method

Subjects

Two hundred subjects were recruited at the rate of \$9.50 per hour from Prolific. All subjects were 18 to 35 years old, currently lived in the United States, had normal or corrected-to-normal vision, and with no ongoing psychological or

neurological disorders. Our screening for Prolific online subjects remained the same as our in-lab sample, such that participants who had below chance (or capacity score below zero) would be eliminated. Additionally, participants who finished less than 80% of the trials, or failed more than 25% of our attentional check questions, would also be eliminated.

Stimuli

The colors and size of the color squares and fixation plus were the same as the ones used in Experiment 1.

Procedures

To maximize the individual differences between subjects, each subject would complete 240 trials of sequential and 240 trials of simultaneous change detection task. Different from Experiment 3a, each subject would complete 120 trials of Set Size 6 and 120 trials of Set Size 4 under each condition. The sequential task was the same as the Set Sizes 4 and 6 ones used in Experiments 1 and 2. The simultaneous task was the same as the Set Sizes 4 and 6 ones used in Experiment 2.

Result

Relation between simultaneous and sequential VWM capacity within subject

Since only half of the trials were used for each condition compared with Experiment 3a, we first examined the split-half reliabilities for sequential and simultaneous paradigms across our two set sizes of interest. Our even-odd split-half reliability test revealed that the simultaneous, Set Size 4: $r(199) = .817$, $p = 3.08 \times 10^{-49}$; Set Size 6: $r(199) = .667$, $p = 4.11 \times 10^{-27}$, and sequential, Set Size 4: $r(199) = .852$, $p = 1.11 \times 10^{-57}$; Set Size 6: $r(199) = .769$, $p = 2.61 \times 10^{-40}$, paradigm show significant Pearson correlations between their even-number and odd-number trials. Given that we had two different set sizes of sequential and simultaneous tasks within the same subject, we first tested whether the individual differences in sequential paradigm were correlated to these in simultaneous paradigm. We replicated the findings in Experiment 3a that visual working memory measured by sequential and simultaneous paradigms were highly correlated with Set Size 6 arrays, $r(199) = .914$, $p = 1.02 \times 10^{-79}$. Furthermore, we also found out that this correlation held with smaller Set Size 4 arrays within the same subjects, $r(199) = .930$, $p = 7.42 \times 10^{-88}$ (see Fig. 7, also see Table 2 for statistical moments). Having both set sizes within the same subject, we would also like to confirm if sequential and simultaneous tasks were measuring highly overlapping individual differences despite of having slightly different set sizes. We first performed a correlational analysis between Set Size 4 sequential and Set Size 6 simultaneous K capacity

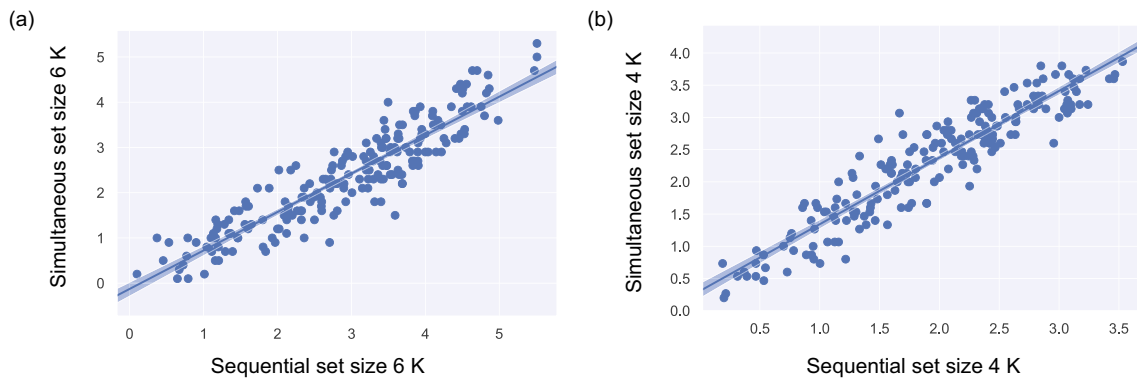


Fig. 7 Correlation between set size 6 and 4 simultaneous and sequential change detection performance in Experiment 3b. **a** The correlation between set size 6 simultaneous and sequential change detection

capacity estimates were highly positive. **b** The correlation between set size 4 simultaneous and sequential change detection capacity estimates were highly positive

measures. The Pearson correlation was highly positive across the 200 subjects, $r(199) = .917, p = 4.73 \times 10^{-81}$. A similar analysis between Set Size 6 sequential and Set Size 4 simultaneous K capacity measures also reveals highly positive correlation, $r(199) = .934, p = 1.61 \times 10^{-90}$ (see Fig. 8). Collectively, we confirmed that both paradigms were measuring similar individual differences in visual working memory capacity, even across slightly different set sizes and with smaller number of trials.

Table 2 The statistical moments of the sequential and simultaneous K scores in Experiment 3b

K capacity estimates	Mean	Variance	Skewness	Kurtosis
Sequential Set Size 4	1.959	0.609	-0.237	-0.681
Simultaneous Set Size 4	2.334	0.754	-0.453	-0.660
Sequential Set Size 6	2.924	1.410	0.139	-0.533
Simultaneous Set Size 6	2.358	1.218	-0.222	-0.711

Discussion

In Experiment 3a and 3b, we discovered that sequential paradigm shared the similar statistical properties, particularly the second or above order moments, with the simultaneous paradigm. Moreover, we replicated our findings in Experiment 2 that both paradigms were reliable, and the individual differences in the two paradigms were highly correlated. Therefore, we concluded that, the sequential change detection task, despite of having a different load at any given moment during encoding, retained the individual differences measured by the simultaneous paradigm.

General discussion

Here, we first showed that sequential change detection paradigm had a comparable capacity limit as in the previous research using simultaneous paradigm. Within the same subject,

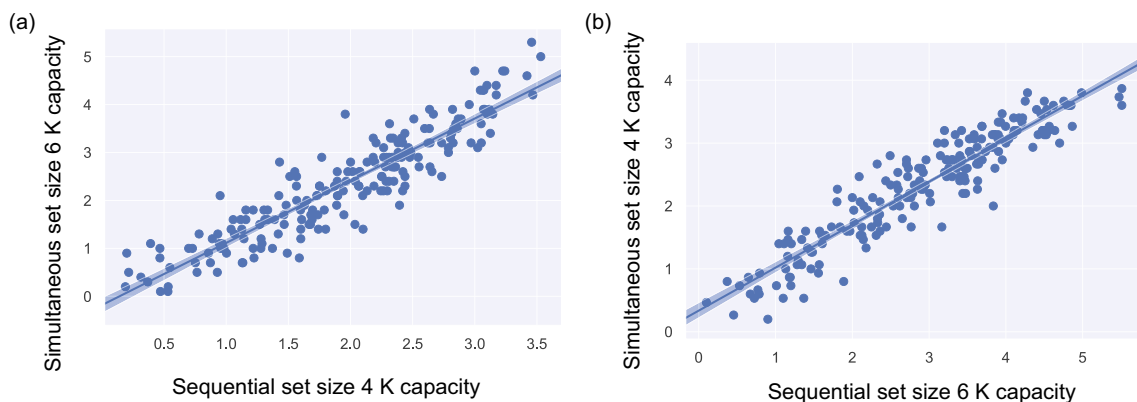


Fig. 8 Set size 4 performance predicted set size 6 performance even with different paradigms in Experiment 3b. **a** The correlation between set size 6 simultaneous and set size 4 sequential change detection capacity

estimates were highly positive. **b** The correlation between set size 4 simultaneous and set size 6 sequential change detection capacity estimates were highly positive

the sequential capacity limit correlated to the simultaneous capacity limit across Set Sizes 3 to 6. Moreover, we discovered that the individual differences in visual working memory performance were highly correlated between the simultaneous and sequential tasks. Collectively, this evidence suggests that sequential and simultaneous versions of the change detection task capture the same individual differences in VWM ability of subjects.

Our findings indicate that the sequential and simultaneous tasks draw upon the same pool of cognitive resources related to performance. That is, even when the competition at encoding is dramatically reduced, essentially the same individual differences can still be observed. These results suggest that the individual differences that have been measured in numerous prior studies using the simultaneous change detection task are not simply the result of encoding competition, but instead may reflect demands for attentional control processes that may also occur at other stages of the task such as maintenance or retrieval (Fukuda et al., 2015; Gazzaley & Nobre, 2012). Moreover, given that the sequential paradigm appears to largely capture the same cognitive ability as the simultaneous paradigm, it may be used to enhance the predictive power of abilities measured by other sequentially designed tasks. For instance, classical long-term memory paradigms tend to present items sequentially, such as recognition memory (Shepard, 1967), associative memory (Raaijmakers & Shiffrin, 1981), or source memory (Shimamura & Squire, 1987). A simultaneous paradigm would introduce target selection bottlenecks during encoding, which could potentially bring in measurement artifacts when being used to predict long-term memory abilities. By contrast, the sequential VWM paradigm always had one item at a time during the encoding. Thus, there may be some circumstances in which a sequential measure of individual differences may be preferable. The current work helps to establish the validity of the sequential change detection task as a measure of these commonly examined individual differences.

Another useful advance of the current work is that it provides a separate measure of performance that largely captures the same individual differences. In latent variable analysis studies, it is necessary to measure two separate tasks that aim to measure the same underlying construct so that a more reliable the underlying construct could be measured. Given its high correlation with the more commonly used simultaneous paradigm, the sequential change detection task would be an ideal candidate as an additional measure of VWM in latent variable experiments. As a first step based on correlational analysis, we did not address both simultaneous and sequential tasks tapped on the exact same construct of visual working memory as the other commonly used working memory tasks. However, our current study would be useful for researchers interested in including multiple measures of visual working memory into their latent variable studies in cognitive abilities.

Empirically, the addition of this reliable sequential VWM measure to existing measurements of working memory could help refine the construct of working memory (Engle et al., 1999). Specifically, as the only sequential measure of visual WM, the sequential change detection paradigm may help to bridge the gaps between the simultaneous task (Luck & Vogel, 1997) and the complex span task (Turner & Engle, 1989), which measures working memory by sequentially interleaving the to-be-recalled word stimuli with simple math questions. Balancing the number of simultaneous and sequential tasks, as well as the number of VWM and span tasks, makes the sequential change detection a promising measure for future individual differences studies of working memory.

Acknowledgements We acknowledge the funding from National Institute of Mental Health (Grant RO1MH087214); Office of Naval Research (Grant N00014-12-1-0972).

References

- Berggren, N., & Eimer, M. (2016). Does contralateral delay activity reflect working memory storage or the current focus of spatial attention within visual working memory? *Journal of Cognitive Neuroscience*, 28(12), 2003–2020. <https://doi.org/10.1162/jocn>
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114. <https://doi.org/10.1017/S0140525X01003922>
- Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309–331. <https://doi.org/10.1037/0096-3445.128.3.309>
- Feldmann-Wüstefeld, T., Vogel, E. K., & Awh, E. (2018). Contralateral delay activity indexes working memory storage, not the current focus of spatial attention. *Journal of Cognitive Neuroscience*, 30(8), 1185–1196. <https://doi.org/10.1162/jocn>
- Fukuda, K., & Vogel, E. K. (2011). Individual differences in recovery time from attentional capture. *Psychological Science*, 22(3), 361–368. <https://doi.org/10.1177/0956797611398493>
- Fukuda, K., Woodman, G. F., & Vogel, E. K. (2015). Individual differences in visual working memory capacity. *Mechanisms of Sensory Working Memory*, 25, 105. <https://doi.org/10.1016/b978-0-12-801371-7.00009-0>
- Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Sciences*, 16(2), 129–135. <https://doi.org/10.1016/j.tics.2011.11.014>
- 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(4), 1963–1968. <https://doi.org/10.1152/jn.00978.2009>
- Kane, M. J., Tuholski, S. W., Hambrick, D. Z., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133(2), 189–217. <https://doi.org/10.1037/0096-3445.133.2.189>
- Kyndt, E., Cascallar, E., & Dochy, F. (2012). Individual differences in working memory capacity and attention, and their relationship with

- students' approaches to learning. *Higher Education*, 64(3), 285–297. <https://doi.org/10.1007/s10734-011-9493-0>
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–284. <https://doi.org/10.1038/36846>
- Martin, J. D., Tsukahara, J. S., Draheim, C., Shipstead, Z., Mashburn, C. A., Vogel, E. K., & Engle, R. W. (2021). The visual arrays task: Visual storage capacity or attention control? *Journal of Experimental Psychology: General*, 150(12), 2525–2551. <https://doi.org/10.1037/xge0001048>
- Miller, A. L., Gross, M. P., & Unsworth, N. (2019). Individual differences in working memory capacity and long-term memory: The influence of intensity of attention to items at encoding as measured by pupil dilation. *Journal of Memory and Language*, 104, 25–42. <https://doi.org/10.1016/j.jml.2018.09.005>
- Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, 15(7), 327–334. <https://doi.org/10.1016/j.tics.2011.05.004>
- Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88(2), 93–134. <https://doi.org/10.1037/0033-295X.88.2.93>
- Raghubar, K. P., Bames, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, 20(2), 110–122. <https://doi.org/10.1016/j.lindif.2009.10.005>
- Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, 18(2), 324–330. <https://doi.org/10.3758/s13423-011-0055-3>
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6(1), 156–163. [https://doi.org/10.1016/S0022-5371\(67\)80067-7](https://doi.org/10.1016/S0022-5371(67)80067-7)
- Shimamura, A. P., & Squire, L. R. (1987). A neuropsychological study of fact memory and source amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3), 464–473. <https://doi.org/10.1037/0278-7393.13.3.464>
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28(2), 127–154. [https://doi.org/10.1016/0749-596X\(89\)90040-5](https://doi.org/10.1016/0749-596X(89)90040-5)
- 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. <https://doi.org/10.1016/j.cogpsych.2014.01.003>
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428(6984), 748–751. <https://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(7067), 500–503. <https://doi.org/10.1038/nature04171>
- Wang, S., Rajsic, J., & Woodman, G. F. (2018). The contralateral delay activity tracks the sequential loading of objects into visual working memory, unlike lateralized alpha oscillations. *Journal of Cognitive Neuroscience*, 31(11), 1689–1698. https://doi.org/10.1162/jocn_a_01446
- Woodman, G. F., Vogel, E. K., & Luck, S. J. (2012). Flexibility in visual working memory: Accurate change detection in the face of irrelevant variations in position. *Visual Cognition*, 20(1), 1–28. <https://doi.org/10.1080/13506285.2011.630694>
- Xu, Z., Adam, K. C. S., Fang, X., & Vogel, E. K. (2018). The reliability and stability of visual working memory capacity. *Behavior Research Methods*, 50(2), 576–588. <https://doi.org/10.3758/s13428-017-0886-6>
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233–235. <https://doi.org/10.1038/nature06860>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.