CrossMark

Depth benefits now loading: Visual working memory capacity and benefits in 3-D

Dawn M. Sarno · Joanna E. Lewis · Mark B. Neider [1]

Abstract

The present studies explored how performance in multidimensional displays varies as a function of visual working memory load, item distribution across depths, and individual capacity differences. In Experiment 1, the benefit of depth information (one depth vs. two depths) was examined across seven set sizes within a change-detection paradigm. Multiple depth planes engendered performance benefits with five items, but elicited performance decrements with three items. These effects were associated with working memory capacity, such that benefits were only observed when the working memory load exceeded an individual's max capacity. Experiment 2 evaluated how the distribution of items in depth aids working memory performance. Equal distribution of items across depths produced higher accuracy compared with when the target was isolated in depth. Lastly, Experiment 3 explored how differences in working memory capacity affect an individual's ability to use depth information to improve their performance. The results indicate that both low-capacity and high-capacity individuals can benefit from depth information, but this may vary as a function of working memory load. Overall, the results indicate that multidimensional displays can improve performance with sufficient working memory load, possibly through some sort of depth tag.

Keywords Visual working memory · Visual short-term memory · 3-D · Capacity

Our interactions with the real world rely heavily on our visual working memory (VWM) abilities. Whether you are dialing a phone number or waiting for a pause at a busy traffic intersection, VWM performance is critical to the successful completion of most tasks. Most research in VWM has explored performance in primarily two-dimensional environments. From this research, we have learned a great deal of how we retain visual information in a short-term memory buffer, which has consistently been shown to have a capacity of approximately three to five items, regardless of the complexity of the items (Awh, Barton, & Vogel, 2007; Luck & Vogel, 1997). The number of items held in VWM varies due to individual differences in capacity and the strategies used to retain items. For instance, high-capacity individuals have been found to be better at excluding irrelevant information in a display compared with low-capacity individuals (Vogel, McCollough, & Machizawa, 2005). Such selectivity allows high-capacity individuals to reduce their VWM load, and by extension

performance. Other strategies of reducing memory load include combining bits of information into meaningful chunks (Miller, 1956). By chunking information, individuals are better able to retain more information at any given time (a common example is how we chunk phone numbers into three-digit groups). Perceptual grouping has also been found to improve VWM performance, where individuals are able to retain more items in VWM if they are grouped by factors such as similarity, proximity, and connectedness (Peterson, & Berryhill, 2013; van Lamsweerde & Beck, 2015; Woodman, Vecera, & Luck, 2003; Xu, 2006). Additionally, recent research has suggested that chunking or grouping items in a display by the depth planes within which they reside may improve performance through some sort of depth tag (Qian, Li, Wang, Liu, & Lei, 2017).

There is evidence demonstrating that humans use depth information to improve their performance efficiency on a variety of tasks. Multiple object tracking studies indicate that participants ignore irrelevant information in depth planes distinct from the target (Haladjian, Montemayor, & Pylyshyn, 2008). Similarly, the visual search literature not only suggests that the mere percept of depth improve performance but also posits the existence of a depth-aware attentional spotlight

✉ Mark B. Neider
  mark.neider@ucf.edu

[1] Department of Psychology, University of Central Florida, 4111 Pictor Lane, Psychology Bldg. 99 Ste. 320, Orlando, FL 32816-1390, USA

(Atchley, Kramer, Andersen, & Theeuwes, 1997; Enns & Rensink, 1990). Despite various demonstrations of the beneficial affordances provided by depth information, limited research has explored VWM in multidimensional environments. Xu and Nakayama (2007) showed that more items could be held in visual short-term memory when they were presented in two depth planes compared with one. Others determined that visual short-term memory may be independent of depth information (Reeves & Lei, 2014), or, at the very least, only beneficial for closer in-depth items compared with farther items (Qian et al., 2017). The latter studies suggest that depth information does not aid VWM performance.

It is possible, however, that benefits associated with depth information in VWM may be subtle in nature, arising only when needed. Consistent with this idea, visual attention research has demonstrated that the benefit of depth information depends on the perceptual load of the task. Downing and Pinker (1985) discovered that there were greater attentional costs when the target item did not appear in the cued depth plane and determined that attention can indeed be oriented in depth. Ghirardelli and Folk (1996) reported contradictory results and determined that the attentional spotlight was blind to depth information. These disparate findings were resolved when Atchley et al. (1997) revealed that previously reported differences were associated with the use of distractors. Downing and Pinker (1985) included distractors in their displays, whereas Ghirardelli and Folk (1996) did not. These results indicate that attention can be allocated in depth and that the use of depth information may vary because of task characteristics, such as perceptual load.

Extending from these findings in the visual attention domain, it seems plausible that inconsistent findings for VWM associated with depth may be due to differences in the VWM load of the task. We are not making the argument that perceptual load and working memory load rely on the same mechanisms, but rather that it is reasonable that as both are related to general cognitive load, they may both affect the utilization of depth information. Prior studies exploring VWM in 3-D used relatively small set sizes and did not examine how individual differences in VWM capacity might affect the utilization of depth information (Qian et al., 2017; Reeves & Lei, 2014; Xu & Nakayama, 2007). It is conceivable that depth information is only used when it is in fact useful, such as when VWM processes are sufficiently taxed, and that prior failures to document a depth benefit associated with VWM may have lacked set sizes large enough to elicit sufficient memory load. Furthermore, individual differences in working memory capacity may further complicate the interpretation of depth effects. More specifically, it is possible that only high capacity individuals can use depth information to improve their performance (Vogel et al., 2005), and depth benefits wash out when averaging across both high-capacity and low-capacity individuals' accuracies. By systematically examining how VWM load affects the utilization of depth information, we may be able to resolve the disparate findings in the literature and identify if and when depth information is in fact useful.

The present studies explore how the utilization of depth information varies as a function of VWM load and individual capacity. Based on the previous literature, we hypothesized that only when the VWM load of the display exceeded capacity would we see the benefit of depth information. All experiments used a change-detection paradigm (i.e., determine if a colored cube changed colors) to manipulate VWM load demands. Experiment 1 compared the benefit of multiple depths using a single depth plane condition and a two depth plane condition. In the single depth plane condition, the target and the distractors were in the same depth. In the two depth plane condition, the target was in a depth plane by itself, with the distractors in a different depth. The results from Experiment 1 indicated if multiple depth planes improved performance. Experiment 2 was conducted to see if depth benefits may be determined by the distribution of items across depth planes. The target was either separated from the distractors by depth or there was an even distribution of items across depth planes. Lastly, Experiment 3 compared the single depth plane (Experiment 1) to the even distribution condition (Experiment 2). Importantly, Experiment 3 used an independent measure of capacity to compare performance for high-capacity and low-capacity individuals to determine if equivalent depth benefits are observed regardless of capacity differences.

Experiment 1

Method

Participants A total of 27 (M_{age} = 18.48 years, 18 female) participants were recruited from the University of Central Florida in exchange for course credit. All participants had normal or corrected-to-normal vision.
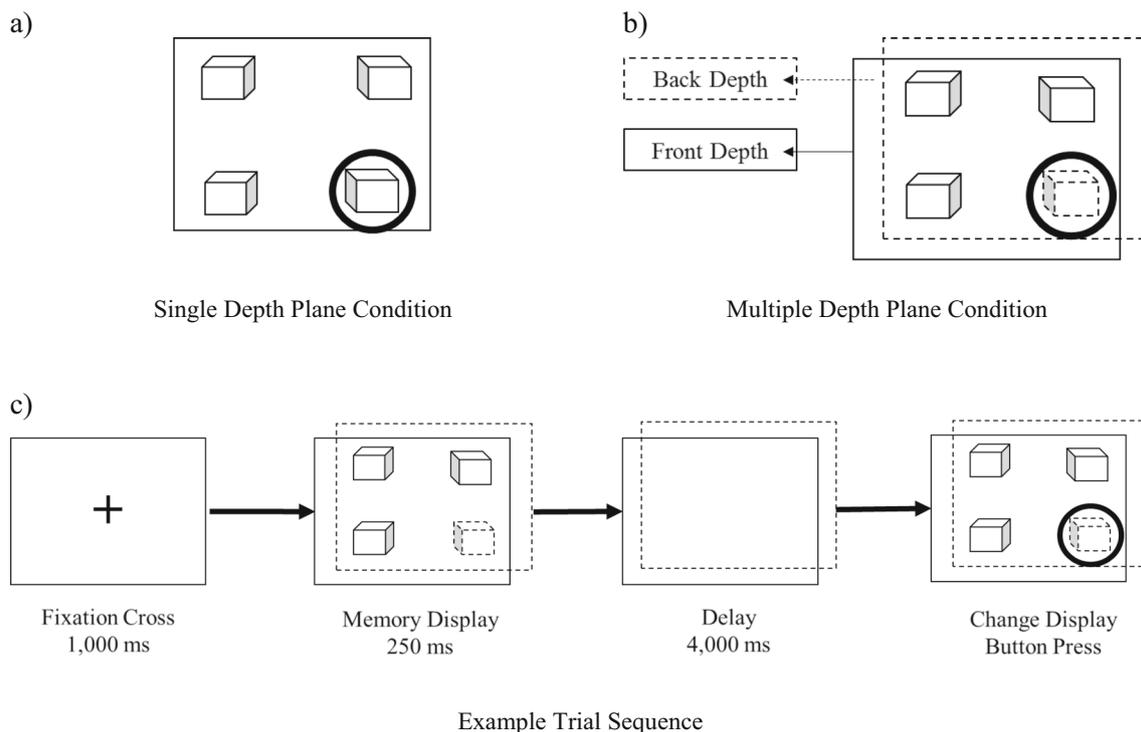
A pilot study was conducted to determine an appropriate sample size. Using our pilot data, we conducted a repeated-measures ANOVA power analysis in G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007), using a Cohen's f of 0.70, power of 0.99, an alpha probability of 0.01, a 0.12 sphericity correction, and a correlation of 0.42 between the repeated measures. A high sphericity correction was used due to the large number of set sizes, which increase the likelihood of a sphericity violation. Based on this analysis, 23 participants should be satisfactory to find a moderate effect size for differences in change-detection accuracy in our task.

Stimuli and procedure The experiment was programmed and run in SR Research Ltd's Experiment Builder. Stimuli were first generated using Blender and then transformed into

anaglyph images in Adobe Photoshop. The experiment was presented on a Dell Professional P190S, 19-in. monitor, 1280 × 1050 pixels, with participants seated in a chin rest approximately 44 cm from the screen. All participants wore red–blue anaglyph glasses to provide the percept of depth. Consistent with work in a similar vein conducted by Donnelly et al. (2017), the anaglyph method was used because it allowed us to easily create the percept of depth within a wide variety of stimuli. Stimuli consisted of arrays of 3-D cubes presented randomly in one of seven colors—black, blue, green, violet, red, white, and yellow—on a checkered background. Each color was carefully generated to avoid possible distortions of the red–blue filters of the anaglyph glasses. The visual angle of each cube was 7.34° × 7.16°. The large size of the cubes was necessary in order to give a convincing perception of depth within the anaglyph method. The cubes appeared to be in a depth plane either behind or in front of the screen. On half of the change trials all cubes appeared in one depth plane, and on the other half the cube to be changed (i.e., the target) appeared in a separate depth plane from the distractors (see Fig. 1a–b). On nonchange trials, all cubes also appeared in one depth plane, or one cube was randomly placed in a separate depth plane. Stimuli were presented in a 3 × 3 grid (with the center location excluded), with up to 2.65° of horizontal jitter and 0.86° of vertical jitter. The target's location was randomly selected from the eight possible positions. Trials varied based on the number of depth planes (1 vs. 2), number of items (2–

8), and change presence (no change vs. change). All conditions were randomized within blocks. There were a total of 294 trials distributed across six experimental blocks.

All participants provided informed consent upon entering the lab and were prescreened for vision (visual acuity [20/32 or better corrected vision on a Snellen eye chart], color vision [Ishihara's test for color blindness; 13 plates], and stereopsis [Stereo Fly test]). Participants were then seated at a computer station for the remainder of the study. Participants were instructed to complete a change-detection task. Participants were told to focus on a fixation cross that appeared for approximately 1,000 ms, then a memory display appeared for 250 ms, followed by a delay period of 4,000 ms, and finally the change display. The long delay period was chosen in order to increase the difficulty of the task, thereby increasing the likelihood that participants would use the depth information, to the extent that they were at all able (see Unsworth & Robison, 2015, for an example of a similar task using a similar delay period). On the final change display, the memory array was presented again, with one of the cubes circled (see Fig. 1). On half of the trials the circled cube had changed colors from the original presentation. Participants were asked to indicate, via button press, if this change had occurred. To familiarize themselves with the task, participants were given 14 practice trials where they were provided with accuracy feedback. Capacity (K) was determined using Cowan's methodology (Cowan, 2001). Capacity (K) was averaged across all seven set sizes

a) Single Depth Plane Condition

b) Multiple Depth Plane Condition

Back Depth

Front Depth

c)

| Fixation Cross 1,000 ms | Memory Display 250 ms | Delay 4,000 ms | Change Display Button Press |

Example Trial Sequence

Fig. 1 Example of the two depth plane conditions and trial sequence. Both **a** and **b** show the change portion of the trial with the target circled. The target was either in the same depth as the distractors (**a**) or in a separate depth by itself (**b**). Each trial began with a fixation cross, then the memory display, delay screen, and change display (**c**)

Performance by Depth
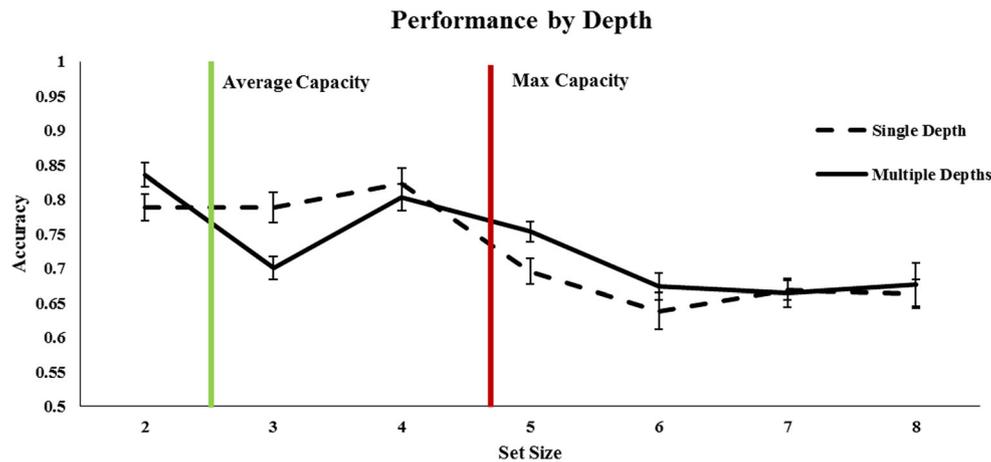
Fig. 2 Experiment 1 accuracy across the seven set sizes for the two conditions. Average capacity (K) and max capacity (K) were collapsed across conditions. Error bars indicate the standard error of the mean corrected for a within-subjects design (Morey, 2008)

in order to develop an average capacity (average K), and the highest capacity across the seven set sizes determined a max capacity (max K). There were optional breaks between experimental blocks (56 trials each), with a mandatory break between Blocks 2 and 3. Following the study participants filled out a brief demographics questionnaire. The entire experiment took approximately an hour with breaks.

Results

Accuracy Four participants were excluded because their accuracy fell below chance performance (50% accuracy). The remaining 23 participants were included in the analyses. Unless otherwise stated, Greenhouse–Geisser corrections were used in any analysis where sphericity was violated. To determine the effect of multiple depth planes on VWM performance, accuracy was submitted to a 2 (depth: one depth plane vs. two depth planes) × 7 (set size: 2–8) repeated-measures ANOVA. The analysis revealed a significant main effect of set size, $F(3.83, 84.27) = 15.13$, $p < .001$, $\eta_p^2 = .41$. Consistent with most VWM tasks, accuracy decreased as the number of items in the display increased. There was no main effect of depth, $F(1, 22) = .44$, $p = .51$, $\eta_p^2 = .02$. However, there was a significant interaction between depth and set size, $F(4.47, 98.23) = 3.14$, $p = .012$, $\eta_p^2 = .13$, indicating the effect of multiple depth planes varied by the number of items in the display (see Fig. 2).

In order to break down this interaction further, separate repeated-measures ANOVAs were conducted to compare depth effects for each set size. These analyzes revealed that multiple depth planes benefited performance at Set Size 5, $F(1, 22) = 5.70$, $p = .026$, $\eta_p^2 = .21$, but hindered performance at Set Size 3, $F(1, 22) = 8.20$, $p = .009$, $\eta_p^2 = .27$. There were no other significant effects of depth ($ps > .068$). Overall, these results suggest that the extent to

which depth information influences VWM processes might be related to the VWM demands of the task.

Capacity Participants' average K and max K were analyzed to further interpret the ANOVA results. Neither average K nor max K varied by the depth condition ($ps > .343$), and were collapsed across groups. When averaged across participants, the average K was 2.58 ($SE = 0.17$), and the max K was 4.72 ($SE = 0.27$). Interestingly, these two-capacity indicators map on remarkably well to the set sizes (three and five) at which we observed accuracy differences associated with depth (see Fig. 2). Specifically, when the number of items in the display is near the average capacity at Set Size 3, participants experience a decrement in accuracy performance associated with depth. However, when participants are just above their max capacity at Set Size 5, they seem to be able to use the additional depth plane to improve VWM performance.

Overall, these results suggest that multiple depth planes can, in fact, improve VWM performance, but that this benefit may only manifest when the task gets more difficult, such as when the set size just exceeds an individual's max capacity in the case of the current task.[1] In

---

[1] This account that depth benefits only arise when VWM load just exceeds max capacity makes the prediction that individual differences in max capacity should map onto the set size at which VWM depth benefits occur. Broadly speaking, in Experiment 1 we only found significant improvements associated with depth at Set Size 5, which was consistent with the average max capacity across participants. However, inspection of individual participant data (which we do not provide here) also suggested some anecdotal evidence that depth effects were somewhat dependent upon individual working memory capacity (some participants with lower capacity show depth benefits at lower set sizes below 5, whereas others with higher max capacity showed benefits at set sizes greater than 5). Out data in the current study are not robust enough to statistically evaluate these patterns in a reasonable manner; however, it is worth noting that individual differences in working memory capacity may ultimately be an important component in understanding the relationship between VWM and depth information.

contrast, depth information also appears to be a potential hindrance to VWM performance when the task is easier. Still, Experiment 1 only examined how multiple depth planes affect performance when the target was isolated in its own depth, making it unclear whether the benefit of depth at Set Size 5 was due to the target being isolated relative to distractor items, or due to some sort of depth tag strategy. In Experiment 2, we examined if depth benefits were solely due to cueing by exploring how equally distributing both the target and distractor items across depth planes affected VWM performance on the same task. If depth benefits are only due to the target being isolated, and hence cued in depth, we would expect performance to be poorer when the items are more evenly distributed across depths. On the contrary, if benefits are more closely associated with a different strategy, such as depth tagging, then we might expect performance to be similar, or better when the items are evenly distributed across depth planes.

Experiment 2

Method

Participants A total of 22 (M_{age} = 19.24 years, 17 female) participants were recruited from the University of Central Florida in exchange for course credit. All participants had normal or corrected-to-normal vision.
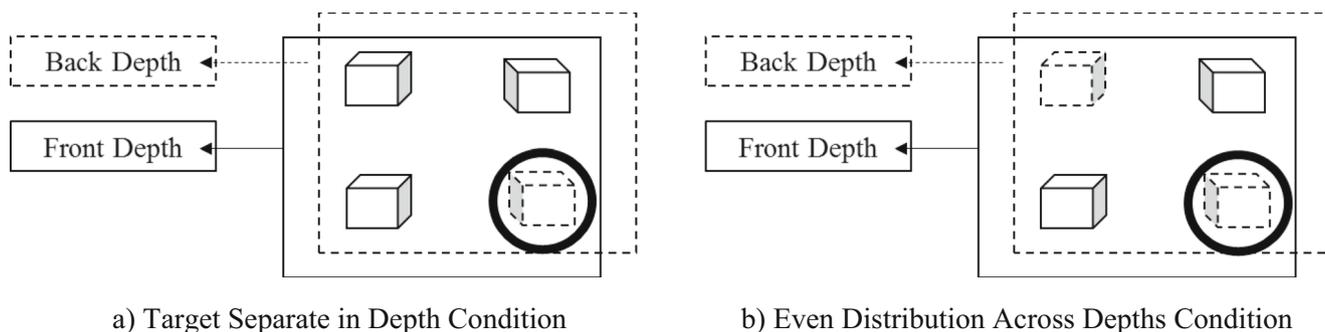
Given the lack of a main effect of depth in Experiment 1, a new power analysis was completed for Experiment 2, assuming a small effect size. Using G*Power 3 (Faul et al., 2007), we calculated a repeated-measures ANOVA power analysis using a Cohen's f of 0.4, power of 0.99, an alpha probability of 0.01, a 1 sphericity correction, and a correlation of 0.5 between the repeated measures. No sphericity correction was used because of the smaller number of set sizes. Based on this analysis, 20 participants should allow

us to find a small effect size for differences in change-detection accuracy between depths.

Stimuli and procedure The stimuli and procedure were identical to that of Experiment 1, with the following exceptions. To determine if multiple planes increase VWM memory performance due to the target being singled out in depth or due to some depth tagging (e.g., grouping or chunking), we varied the distribution of items across depth planes (see Fig. 3). Specifically, in one condition the number of items in the display were evenly distributed across the two depth planes, and in the other condition the target was separated in its own depth plane with the distractors in the other depth (as in Experiment 1). The two depth conditions were presented on an equal number of trials. In order to evenly distribute items across depth planes, only Set Sizes 4, 6, and 8 were used. Set Size 2 was not used, as it would have rendered the target isolated in depth, as in Experiment 1. There were a total of 252 trials, with one block of 12 practice trials and five blocks of 48 experimental trials.

Results

Accuracy Two participants were excluded because their accuracy fell below chance performance (50%). Twenty participants remained to be included in the following analyses. A 2 (depth: even distribution of items vs. target separate) × 3 (set size: 4, 6, 8) repeated-measures ANOVA was conducted on accuracy. A main effect of set size was found, $F(1.57, 29.88) = 24.24$, $p < .001$, $\eta_p^2 = .56$, indicating accuracy generally decreased as set size increased. Importantly, there was a main effect of depth, $F(1, 19) = 8.22$, $p = .010$, $\eta_p^2 = .30$, such that accuracy was higher on trials in which the items in the display were more evenly distributed between the two depths (see Fig. 4). This result indicates that the more distributed items are in depth, the larger the benefit to performance. There was no significant interaction between depth and set size, $F(1.99, 37.88) = 0.66$, $p = .523$, $\eta_p^2 = .03$.

a) Target Separate in Depth Condition

b) Even Distribution Across Depths Condition

Fig. 3 Example of distribution of items in each depth condition. Both **a** and **b** show the change portion of the trial with the target circled. The distractors were either in a depth separate from the target (**a**) or all the items were evenly distributed between depths (**b**)
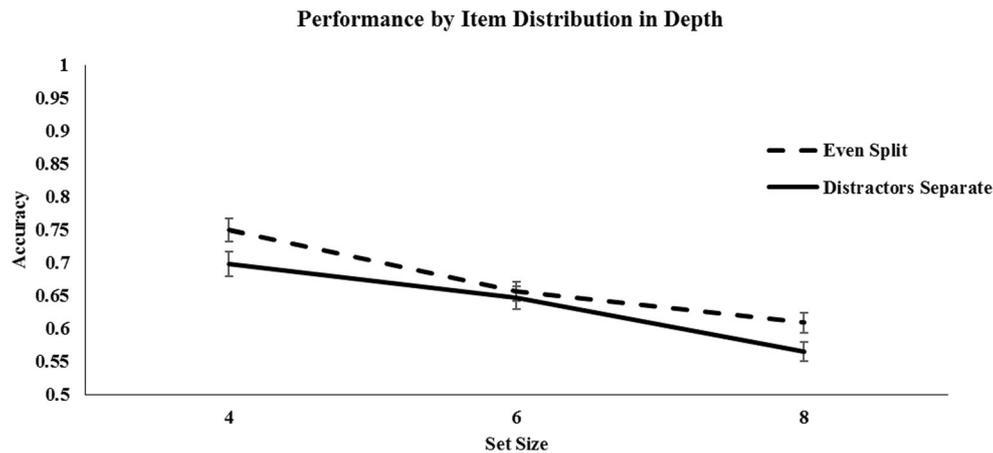
Fig. 4 Experiment 2 accuracy for each condition across set sizes. Error bars indicate the standard error of the mean corrected for a within-subjects design (Morey, 2008)

Capacity Unlike in Experiment 1, both average K and max K varied as a function of the depth conditions. Specifically, average K in the evenly distributed-depth condition ($M = 2.43$) was significantly higher than in the target-separate condition ($M = 1.84$), $F(1, 19) = 5.17$, $p = .035$, $\eta_p^2 = .21$. Additionally, max K was higher in evenly distributed-depth condition ($M = 3.69$) than when the target was in its own depth plane ($M = 3.07$), $F(1, 19) = 8.62$, $p = .008$, $\eta_p^2 = .31$. Both of these results suggest that capacity can be increased by depth tagging, with the greatest benefits seen with a more even distribution of items across depths.

Together, these results support the interpretation that VWM capacity is somewhat malleable and can be substantially improved by the availability of depth information. However, it still remains unclear if all individuals benefit from multidimensional displays equally. Since previous work has indicated that high-capacity individuals may be more effective at using relevant information in visual memory tasks (Vogel et al., 2005), it seems possible that individual differences in capacity may determine the benefit of multiple depth planes on VWM. In Experiment 3, we included an independent measure of working memory capacity to explore how depth benefits may vary as a function of working memory ability.

Experiment 3

Method

Participants A total of 64 ($M_{age} = 19.30$ years, 36 female) participants were recruited from the University of Central Florida in exchange for course credit. All participants had normal or corrected-to-normal vision.

A repeated-measures ANOVA power analysis in G*Power 3 was conducted (Faul et al., 2007), using a Cohen's f of 0.25, power of 0.99, an alpha probability of 0.01, a 1 sphericity correction, and a correlation of 0.5 between the repeated measures. Based on this analysis, 62 participants should be sufficient to find a small effect size for differences in change-detection accuracy.

Stimuli and procedure The stimuli and procedure that were used matched Experiments 1 and 2, with the following exceptions. The single depth plane condition from Experiment 1 was compared with the even distribution condition from Experiment 2 (see Fig. 5). Additionally, prior to beginning the experiment participants completed the same change-detection task without depth information. This task was used to achieve an independent measure of working memory capacity for all participants. This initial measure of capacity was 98 trials (12 practice trials, two blocks of 43 experimental trials), with optional breaks between practice and the experimental blocks. With the addition of the capacity measurement, the entire experiment took approximately an hour and a half to complete.

Results

Accuracy Two participants were excluded because their accuracy fell below chance performance (50%). Sixty-two participants remained to be included in the following analyses. A 2 (depth: even distribution of items between depths vs. single depth) × 3 (set size: 4, 6, 8) repeated-measures ANOVA was conducted on accuracy in order to examine how evenly distributing items across depths compares to performance when all items are in a single depth. As expected, a main effect of set size was found, $F(2, 122) = 84.33$, $p < .001$, $\eta_p^2 = .58$, indicating accuracy decreased as set size increased. As in Experiment 2, there was a main effect of depth, $F(1, 61) = 13.04$, $p < .001$, $\eta_p^2 = .18$, with higher accuracy on trials where the items were evenly distributed between the two depths (see Fig. 6). These results reaffirm the first two

a) Single Depth Plane Condition   b) Even Distribution Across Depths Condition
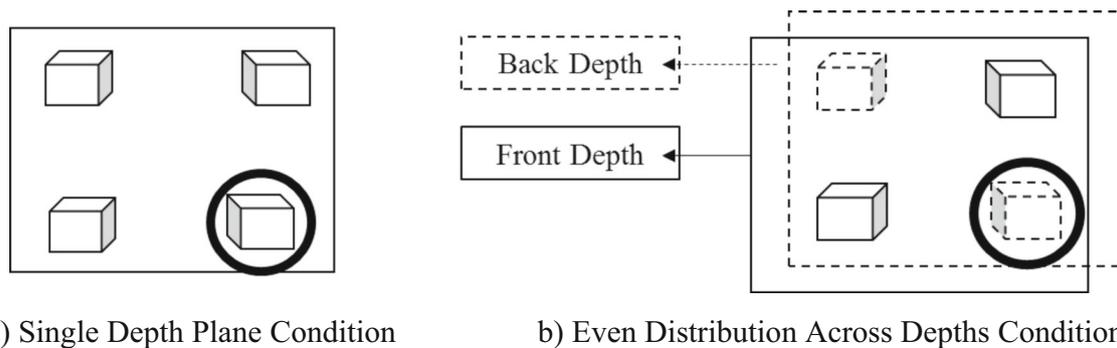
Fig. 5 Example of distribution of items in each depth condition. Both **a** and **b** show the change portion of the trial with the target circled. All items were either in the same depth plane (**a**) or all the items were evenly distributed between depths (**b**)

experiments findings that working memory performance can be aided with multiple depth planes. There was not a significant interaction between depth and set size, $F(2, 122) = 0.86$, $p = .428$, $\eta_p^2 = .01$.

High-capacity versus low-capacity individuals Capacity was calculated by averaging participant's Cowan's K across all three set sizes. A median split was then used to separate all participants into high-capacity and low-capacity groups (see Table 1). A 2 (capacity: high capacity vs. low capacity) × 2 (depth: even distribution of items between depths vs. single depth) × 3 (set size: 4, 6, 8) repeated-measures ANOVA was conducted in order to determine if both high-capacity and low-capacity individuals would see working memory benefits in multidimensional displays.

A main effect of set size was found, $F(2, 120) = 84.02$, $p < .001$, $\eta_p^2 = .58$, again demonstrating that performance decreased as the number of items to be remember increased. There was also a significant main effect of depth, $F(1, 60) = 13.47$, $p < .001$, $\eta_p^2 = .18$, with better working memory performance on trials where the items were evenly distributed across two depth planes. It should be noted that, consistent with the benefits in Experiments 1 and 2, the low-capacity

individuals' max capacity was below all presented set sizes (~3.5 items). Although the high-capacity individuals' max capacity was just above the lowest set size (~5.3 items), it appears their benefit of depth increases with the number of items in the display. Importantly there was a significant main effect of capacity, $F(1, 60) = 4.86$, $p = .031$, $\eta_p^2 = .08$, indicating that individuals with higher working memory capacity outperformed their lower capacity counterparts (see Fig. 7). There were no other significant effects ($ps > .089$). It should be noted that the Depth × Capacity interaction approached significance, $F(1, 60) = 3.00$, $p = .089$, $\eta_p^2 = .05$, suggesting that high-capacity and low-capacity individuals may differ in their ability to benefit from multidimensional displays. The observed power for the interaction term was .40, suggesting it is possible a different and/or larger sample may have led to a significant interaction effect. In such a case, the current data would suggest that depth related differences in VWM are sensitive to individual capacity (see Fig. 7) such that high-capacity individuals continue to benefit at higher loads, whereas low-capacity individuals do not. Again, we would emphasize here that such an interpretation must be made cautiously, given the current data. Despite possible differences between individuals, the results indicate that all individuals
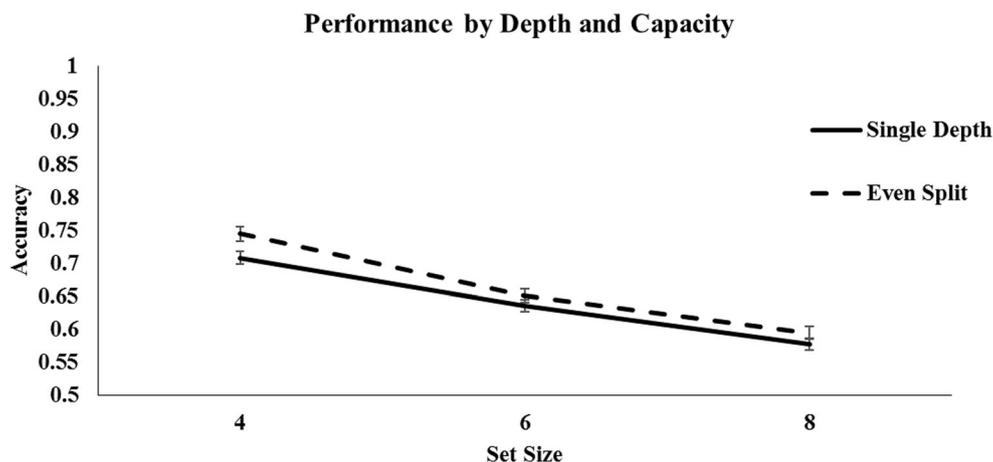
Fig. 6 Experiment 3 accuracy for each condition across set sizes. Error bars indicate the standard error of the mean corrected for a within-subjects design (Morey, 2008)

Table 1 Means and standard deviations for change accuracy and capacity by group

Group	Set Size 4	Set Size 6	Set Size 8	Average K	Max K
High capacity Single depth	.73 (.16)	.65 (.07)	.60 (.10)	4.20 (.51)	5.35 (.87)
High capacity Even distribution	.76 (.14)	.69 (.11)	.63 (.10)		
Low capacity Single depth	.69 (.11)	.62 (.09)	.56 (.11)	2.13 (1.17)	3.54 (.93)
Low capacity Even distribution	.74 (.13)	.61 (.13)	.56 (.10)		

Note. Values displayed represent mean accuracy trials each condition for each group and each group's average K. Parenthetical values represent standard deviations

can benefit from multidimensional displays at some point, regardless of individual differences in VWM capacity.

General discussion

Previous research exploring the effects of depth information on VWM performance has had inconsistent results. Qian et al. (2017) reported that multiple depth planes do not elicit performance benefits over single depth plane displays. Rather, there are only benefits for items closer in depth compared with those farther in depth. However, this study and others exploring VWM in depth used relatively short retention intervals and restricted set sizes that may be below the memory load necessary to require the usage of depth information (e.g., Reeves & Lei, 2014). By exploring performance across several set sizes, the present studies demonstrate that depth information can improve VWM performance with sufficient VWM load. Experiment 1 showed that depth benefits are observed at some set sizes, and that depth can hinder performance at others. Experiment 2 determined that depth benefits are not due to the target being cued in depth, but rather are likely associated with some sort of depth tagging (e.g.,

chunking or grouping). Lastly, Experiment 3 demonstrated that individuals can benefit from depth information, regardless of differences in capacity, but the magnitude may vary depending on memory load.

Previous work exploring depth benefits in visual attention indicates that a certain level of cognitive load may be required to observe performance benefits associated with depth. Our results are consistent with this literature; for depth to be useful, a certain level of general task difficulty (in our case, memory load) appears to be necessary (Atchley et al., 1997; Finlayson, Remington, Retell, & Grove, 2013). Our Experiment 1 findings were consistent with this notion, as depth benefits were only observed at Set Size 5. Interestingly, this set size was just above the participants' max capacity for the task as measured by Cowan's K, indicating a relatively high memory load. However, it must also be noted that when the memory load of the task was just above participants' average capacity, depth information appears to have been a hindrance. Tentatively, such a pattern might suggest that when a given task is well within an individual's ability, they do not need depth information. Under such circumstances (e.g., lower set sizes), the depth information may become extraneous information (e.g., perceptual noise)
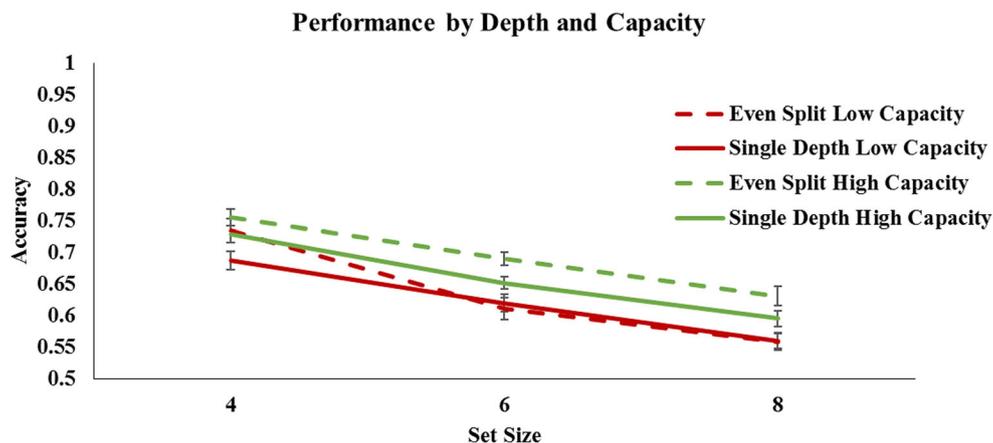
Fig. 7 Experiment 3 accuracy for each condition across set sizes broken down by individual capacity. Error bars indicate the standard error of the mean corrected for a within-subjects design (Morey, 2008)

that ultimately harms VWM performance. This hypothesis is consistent with previous VWM research that demonstrates that attention to irrelevant information decreases VWM abilities (Vogel et al., 2005). At larger set sizes, depth information may be considered as relevant information that is used to enhance VWM abilities. A bolder but not necessarily incompatible explanation may be that sufficient memory load is necessary to engender performance benefits from multiple depth planes in the VWM domain. This account does, however, raise the question of why we did not find depth benefits in displays with more than five items. One possibility is that the memory demand of the task at higher set sizes was simply too high for depth cues to be helpful. That is, depth information can provide a crutch for VWM, but its utility has limits. Still, what limits do exist appear to be somewhat malleable and dependent on how items are distributed across depth dimensions. In Experiment 1, all items but the target were relegated to a single depth plane, but when items were distributed equally across depth planes, as they were in Experiments 2 and 3, overall depth benefits were observed out to set sizes ranging from four to eight items. One might interpret these findings as evidence that VWM processes can tag or group items in depth to aid overall performance. Regardless of the precise mechanism by which depth information is employed, our data clearly indicate that the additional information it provides can serve as a useful aid for VWM.

Closely intertwined with our finding that depth benefits depend on the VWM load of the task are our findings regarding individual differences in working memory. Previous research has established that high-capacity individuals are better at using the relevant information in a display compared with low-capacity individuals (Vogel et al., 2005). These findings can be extended to suggest that high-capacity individuals may be more likely to use depth information as a strategy to improve their performance. Experiment 3 explored this possibility, but did not produce evidence supporting it; all participants benefitted from multiple depth planes (underlying the fact that these findings are fairly robust). However, there did appear to be an insignificant trend toward an interaction of depth and capacity, such that low-capacity individuals benefitted more at the lower set sizes while their high-capacity counterparts continued to benefit out to displays with eight items. We are reticent to speculate overly much on nonsignificant findings, but we would cautiously suggest that the possibility that depth effects in VWM might interact in some way with individual VWM capacity is a possibility that warrants further consideration.

Although our studies suggest that multidimensional displays do extend working memory abilities for most individuals, the precise mechanism of the benefit is still unclear. In Experiment 1, the target was singled out in depth from the rest of the distracting items, indicating that performance benefits might be due, in part, to some sort of depth cueing. Previous visual search studies have indicated that cueing participants to the depth plane of a target can improve performance (Atchley et al., 1997; Dent, Braithwaite, He, & Humphreys, 2012; Finlayson et al., 2013). However, Experiments 2 and 3 both found similar, if not more robust, patterns in item arrays where all items (i.e., the target and distractors) were equally distributed across two depth planes. In light of those findings, the Experiment 1 data seem to provide less convincing support for the strict argument that depth benefits are solely due to depth cueing. If the benefits at set size 5 were due to cueing alone, one would expect the benefit to persist across the larger set sizes, since participants would still only need to hold one item in memory. Broadly, we view these combined data as evidence against an account where depth benefits in VWM are solely derived from some sort of depth cueing, though such cueing might provide some explanatory power for circumstances where a target is isolated in depth. Instead, it seems more likely that these benefits are more largely associated with some other mechanism, possibly some sort of tagging by depth or perceptual grouping by depth. Xu and Nakayama (2007) found comparable benefits by presenting two different 3-D surfaces in a sequential paradigm. Interestingly, they did not find any differences with depth to VWM when both 3-D surfaces were presented simultaneously. This suggests that although individuals may be able to group or tag items by depth, it could vary by the type of depth presentation. Interestingly, one could also speculate that the depth-tagging strategy may be representative of serial processing of items in the display, with participants reviewing each item and coding the depth information at some level. Another possibility is that there is a preattentive filtering of the less densely populated depth layer that would be congruent with a solely cueing explanation. However, a form of serial processing seems more harmonious with our findings. Regardless, additional studies will be necessary to further elucidate these underlying mechanisms.

Generally speaking, the results from our study are consistent with Xu and Nakayama (2007) and suggest that depth information can improve performance in VWM tasks when the memory load of the task is sufficiently high. This benefit of depth information does not rely on the target being singled out in depth; it is observed when items are distributed across depth planes as well. Importantly, these benefits can be seen across both high-capacity and low-capacity individuals, suggesting the depth can act as a general crutch to VWM process. Overall, our findings suggest that it is possible for multidimensional displays to extend VWM abilities beyond typical capacity limitations given the appropriate memory load.

References

Atchley, P., Kramer, A. F., Andersen, G. J., & Theeuwes, J. (1997). Spatial cuing in a stereoscopic display: Evidence for a 'depth-aware' attentional focus. *Psychonomic Bulletin & Review*, *4*(4), 524–529.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628. doi:https://doi.org/10.1111/j.1467-9280.2007.01949.x

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.

Dent, K., Braithwaite, J., He, X., & Humphreys, G. (2012). Integrating space and time in visual search: How the preview benefit it modulated by stereoscopic depth. *Vision Research*, *65*, 45–61.

Donnelly, N., Godwin, H. J., Menneer, T., Liversedge, S. P., Cave, K. R., & Holliman, N. S. (2017). Adding depth to overlapping displays can improve visual search performance. *Journal of Experimental Psychology: Human Perception And Performance*, *43*(8), 1532–1549.

Downing, C. J., & Pinker, S. (1985). The spatial structure of visual attention. In M. P. O. Martin (Ed.), Attention and performance XI (pp. 171–187). Hillsdale: Erlbaum.

Enns, J., & Rensink, R. (1990). Influence of scene based properties on visual search. *Science*, *247*(4943), 721–723.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Finlayson, N., Remington, R., Retell, J., & Grove, P. (2013). Segmentation by depth does not always facilitate visual search. *Journal of Vision*, *13*(8), 1–14.

Ghirardelli, T. G., & Folk, C. L. (1996). Spatial cuing in a stereoscopic display: Evidence for a "depth-blind" attentional spotlight. *Psychonomic Bulletin & Review*, *3*(1), 81–86. doi:https://doi.org/10.3758/BF03210744

Haladjian, H. H., Montemayor, C., & Pylyshyn, Z. W. (2008). Segregating targets and nontargets in depth eliminates inhibition of nontargets in multiple object tracking. *Visual Cognition*, *16*(1), 107–110.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. doi:https://doi.org/10.1037/h0043158

Morey, R. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64.

Peterson, D. J., & Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychonomic Bulletin & Review*, *20*(6), 1282–1289.

Qian, J., Li, J., Wang, K., Liu, S., & Lei, Q. (2017). Evidence for the effect of depth on visual working memory. *Scientific Reports*, *7*(1), 6408. doi:https://doi.org/10.1038/s41598-017-06719-6

Reeves, A., & Lei, Q. (2014). Is visual short-term memory depthful?. *Vision Research*, *96*, 106–112. doi:https://doi.org/10.1016/j.visres.2014.01.010

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*(3), 757–765.

van Lamsweerde, A. E., & Beck, M. R. (2015). Incidental learning of probability information is differentially affected by the type of visual working memory representation. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, *69*(4), 283–296. doi:https://doi.org/10.1037/cep0000058

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. doi:https://doi.org/10.1038/nature04171

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, *10*(1), 80–87. doi:https://doi.org/10.3758/BF03196470

Xu, Y. (2006). Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness. *Perception & Psychophysics*, *68*(5), 815–828.

Xu, Y., & Nakayama, K. (2007). Visual short-term memory benefit for objects on different 3-D surfaces. *Journal of Experimental Psychology: General*, *136*(4), 653–662.