Cognition 214 (2021) 104748

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Using space to remember: Short-term spatial structure spontaneously improves working memory

Sami R. Yousif [a,*], Monica D. Rosenberg [b], Frank C. Keil [a]

[a] Yale University, Department of Psychology, United States of America
[b] University of Chicago, Department of Psychology, United States of America

ARTICLE INFO

Keywords:
Working memory
Spatial cognition
Visual working memory, verbal working memory
Visuospatial bootstrapping

ABSTRACT

Spatial information plays an important role in how we remember. In general, there are two (non mutually exclusive) views regarding the role that space plays in memory. One view is that objects overlapping in space interfere with each other in memory. For example, objects presented in the same location (at different points in time) are more frequently confused with one another than objects that are not. Another view is that spatial information can 'bootstrap' other kinds of information. For example, remembering a phone number is easier one can see the arrangement of a keypad. Here, building on both perspectives, we test the hypothesis that task-irrelevant *spatial structure* (i.e., objects appearing in stable locations over repeated iterations) improves working memory. Across 7 experiments, we demonstrate that (1) irrelevant spatial structure improves memory for sequences of objects; (2) this effect does not depend on long-term spatial associations; (3) this effect is unique to space (as opposed to features like color); and (4) *spatial structure* can be teased apart from *spatial interference*, and the former drives memory improvement. We discuss how these findings relate to and challenge 'spatial interference' accounts as well as 'visuospatial bootstrapping'.

Spatial representations are implicated in a diverse array of cognitive processes — from aesthetics (Palmer et al., 2013), to numerical cognition (Dehaene et al., 1993; Zorzi et al., 2002), to social cognition (Parkinson & Wheatley, 2013). For millennia, there has also been a notion that space plays some role in how we remember, or else can be used to improve memory (as in the 'method of loci'). In all of these cases, spatial representations apparently matter, but *how* space is involved in these processes is largely unclear. For example, are spatial-numerical associations a product of long-term mappings between numbers and space, or online mappings formed in working memory (for discussion, see van Dijck & Fias, 2011)? Here, we investigate the role of space in working memory to better understand the flexibility of spatial representations and to address the ways in which spatial representations influence memory in the first place.

Here, we investigate working memory, which refers to the short-term maintenance and manipulation of information in the mind (Baddeley, 1992). We consider two distinct possibilities regarding the relationship between space and working memory. The first we call 'spatial interference'. This view suggests that items appearing in the same location in space interfere with one another (e.g., Treisman & Zhang, 2006). The second is 'visuospatial bootstrapping'. Work on visuospatial bootstrapping suggests that items presented with stable spatial mappings (e.g., as in the digits on a keypad) are better remembered than items without such mappings (for review, see Darling, Allen, & Havelka, 2017). While these views are not mutually exclusive, they stem from different approaches. We will briefly highlight typical 'spatial interference' and 'visuospatial bootstrapping' effects.

1. Spatial interference

One theory is that space supports the binding of features to objects, and therefore that objects overlapping in space interfere with one another (e.g., Treisman & Zhang, 2006). That is, if you see a blue circle and a green triangle in the same location, you will be more likely to experience a feature binding error, misremembering having seen either a green circle or a blue triangle. Similar effects are observed across a wide range of paradigms (cf. Jiang, Olson, & Chun, 2000; Rajsic & Wilson, 2014; Woodman, Vogel, & Luck, 2012). For example, participants tasked with remembering the orientation of a line perform worse when multiple lines appeared in the same location (Pertzov & Husain, 2014). Such effects are also specific to space: lines with overlapping colors do *not* result in the same kind of interference (see also Rajsic &

* Corresponding author at: Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205, United States of America.
E-mail address: sami.yousif@yale.edu (S.R. Yousif).

https://doi.org/10.1016/j.cognition.2021.104748
Received 30 September 2020; Received in revised form 22 April 2021; Accepted 23 April 2021
Available online 26 May 2021

Wilson, 2014).

2. Visuospatial bootstrapping

Suppose you are memorizing a phone number; perhaps you would try to visualize where each number is located on a keypad. Indeed, people are better at remembering verbal information when it is mapped onto consistent spatial locations: digits are better remembered when they are presented in a keypad formation, as opposed to being presented in a single location, or even in a line (Darling & Havelka, 2010; Race, Palombo, Cadden, Burke, & Verfaellie, 2015). However, other work suggests this is only true when those mappings exist in long-term memory; people do *not* better remember digits that are displayed in a rearranged keypad formation (Darling, Allen, Havelka, Campbell, & Rattray, 2012). This phenomenon of *'visuospatial bootstrapping'* is said to speak to communication between verbal and visual information systems in working memory (for review, Darling et al., 2017).

A related body of work suggests that items in working memory are automatically spatialized (Abrahamse, Van Dijck, Majerus, & Fias, 2014; Aulet, Yousif, & Lourenco, 2021; Guida et al., 2018; Guida & Campitelli, 2019; Guida, Leroux, Lavielle-Guida, & Noël, 2016; van Dijck, Abrahamse, Acar, Ketels, & Fias, 2014; van Dijck & Fias, 2011). In other words, imagine a task in which individuals must remember a sequence of objects: 'orange', 'apple', 'pear', 'banana', 'cherry'; after memorizing this sequence, individuals respond relatively faster to earlier items with their left hand compared to their right hand, and relatively faster to later items with their right hand compared to their left hand. Such lateralization suggests that the sequence had been mapped onto space in some way, and perhaps that this mapping was functionally involved in the maintenance of that information in working memory. This is known as the Spatial–Positional Association of Response Codes ('SPoARC').

So far, we have conceptualized working memory broadly as a system for maintaining and manipulating information online. However, there are many distinct models of the working memory system. For example, earlier working memory models differentiated discrete subsystems of working memory: a visuospatial sketchpad, which manipulates visual information, a phonological loop, which manipulates verbal information, and a central executive (Baddeley & Hitch, 1974). Later, a fourth subsystem was proposed, the episodic buffer, to explain the interactions of information across modality and across memory systems (i.e., short-term and long-term memory; Baddeley, 2000). Other views characterize working memory more continuously and do not segregate visual and verbal working memory (and thus have no need for a fourth system to moderate between them; e.g., Cowan, 1998). The exact nature of working memory remains a topic of ongoing debate.

Note, however, that visuospatial bootstrapping and related phenomena such as the 'SPoARC' effect address interactions between short-term and long-term memory (e.g., Darling et al., 2012; Race et al., 2015) as well as interactions between verbal and visuospatial information in working memory (see also Alloway, Gathercole, & Pickering, 2006). As such, this body of work speaks not only to how spatial information influences memory, but also to the organization of our memory systems in the first place (in that information in long-term memory influences the retention of information in short-term memory, and in that visuospatial information influences the retention of verbal information). In this way, the present results may bear on debates regarding the extent to which working memory is modality specific (see, e.g., Allen, Havelka, Falcon, Evans, & Darling, 2015; Morey, 2018).

3. Current study

Here, we test the hypothesis that *short-term spatial structure* supports working memory maintenance, even in tasks that pose no explicit spatial requirements. Spatial structure could take many forms, and here we operationalize it as a consistent mapping between objects and space (as in 'visuospatial bootstrapping'; Darling et al., 2017). For example, shopping in a grocery store whose aisles are rearranged every time you visited would be an experience with low spatial structure, whereas attending a meeting in which participants always sat in the same seats would be an experience with high spatial structure. In the current set of studies, we ask whether objects repeatedly appearing in the same location (although on a shorter time scale) are better remembered than objects repeatedly appearing in different locations.

Based on prior work suggesting a role of spatial representations in a broad range of cognitive processes (both in the domain of working memory, e.g., Darling et al., 2017; Pertzov & Husain, 2014; and beyond the domain of working memory; e.g., Dehaene et al., 1993; Parkinson & Wheatley, 2013; Zorzi et al., 2002), we hypothesize that task-irrelevant spatial information will benefit visual working memory more than matched non-spatial visual information (i.e., color) and non-visual information (i.e., audio information). We will also further probe *how* space is special, specifically contrasting the idea of 'spatial structure' with both 'spatial interference' and 'visuospatial bootstrapping' effects.

These broad goals can be broken down into a few specific aims. *First*, we asked about the interaction of space, long-term memory, and working memory. Prior work has suggested that spatial mappings facilitate memory only when those spatial mappings are held in long-term memory (Darling et al., 2012). However, another possibility that visuospatial bootstrapping does not *depend* on mappings in long-term memory, but that mappings in long-term memory *interfere* with short-term mappings. Here, we test short-term spatial mappings (i.e., established over a span of 6–10 s) in a case where there are no long-term mappings. *Second*, we asked *how* spatial information influenced working memory (assuming an effect of space in the first place). For example, what if objects are mapped onto stable locations, but other objects are mapped onto those same locations? Work on neither 'spatial interference' nor 'visuospatial bootstrapping' makes clear predictions about such cases. *Finally*, we want to understand how spatial structure and verbal rehearsal compete when participants can rely on both. The task used here allowed for both visual rehearsal (see Awh, Jonides, & Reuter-Lorenz, 1998) and verbal rehearsal of the relevant information through memorizing sequences of common shapes (i.e., circles, pentagons, and diamonds). Those shapes appeared in different locations and in different colors, but participants were specifically instructed to remember only (a) what shapes they saw and (b) what order they saw them in. Participants were free to verbally rehearse the sequences (which they frequently did) — but this was not a requirement. Therefore, akin to visuospatial bootstrapping tasks (Darling et al., 2017), the task here may speak to interactions between verbal and visual working memory.

To address these questions, we present seven experiments, all with the same essential components. (1) Participants always remembered sequences of shapes that they were free to verbally rehearse, and (2) the sequences were often structured such that either location information or color information (and in one case audio information) covaried with the different shapes. E.g., in a 'space-structured' condition, any shape that appeared multiple times in a sequence always appeared in the same location, but the colors of those shapes are random (and vice-versa for a 'colored-structured' condition). Combined, these features allow us to probe when and how we use space to remember and how these effects inform current research on working memory.

4. Experiments 1a and 1b: space vs. color

In a first set of experiments, participants completed the simplest version of our paradigm: they saw a series of 5–7 shapes (comprised of three unique shapes appearing at least once each, and in a random order) and had to recall (a) what those shapes were and (b) the order that they saw them in. Crucially, participants completed two blocks of trials, each of which was structured in a unique way. In the 'space-structured' block, any shape appearing multiple times appeared in the same location; no other shapes appeared in that location, and the colors of all shapes were randomized. In the 'color-structured' block, any shape appearing multiple times appeared in the same color; no other shapes appeared in that color, and the locations of all shapes were randomized. We had two key questions: (1) Does either location-based or color-based structure uniquely influence working memory, and (2) If so, does structure influence working memory *even when* participants report verbal rehearsal strategies?

4.1. Method

This experiment, and all subsequent experiments, were pre-registered. Those pre-registrations, as well as experiment code, raw data, and analyses, can be accessed here: https://osf.io/85sjd/. Experiments 1a and 1b were identical, except for their sample sizes and one change to the instructions (as noted below).

4.1.1. Participants

24 undergraduate students participated in Experiment 1a and 16 undergraduate students participated in Experiment 1b in exchange for course credit. The sample sizes were chosen in advance based on pilot data and were pre-registered. The sample size of Experiment 1b was chosen based on repeatedly sub-sampling data from Experiment 1a and finding that 16 participants were sufficient to demonstrate the primary effect. This study was approved by the relevant Institutional Review Board.

4.1.2. Apparatus

The experiment was conducted with custom software written in Python with the PsychoPy libraries (Peirce et al., 2019). Participants sat without restraint approximately 60 cm from a $49^\circ \times 29^\circ$ display, with all spatial extents reported below computed based on this distance.

4.1.3. Stimuli

The display consisted of four black squares ($5.10^\circ \times 5.10^\circ$) on a grey (50% black; 50% white) background (Fig. 1). The squares were located in each of the four quadrants of the screen, each 7.66° horizontally and 7.66° vertically displaced from the center of the screen. The shapes themselves (a circle, a pentagon, and a diamond) were all just shorter than 5.10° in height and appeared centered within one of the four black squares. They appeared in one of four colors (the default 'red', 'green', 'blue', or 'yellow' in PsychoPy).

4.1.4. Procedure & design

On each trial, participants watched as shapes appeared one at a time in different locations and in different colors. The shapes appeared for 1000 ms, with 500 ms between presentations.[1] Any given trial had either 5, 6, or 7 shapes (see more on how the sequences were constructed below). After all shapes were presented, the three shapes appeared in white side-by-side in the center of the screen, in a random order. The four black squares in each quadrant remained on the screen during this time. Participants then had to click on the shapes in the order that they saw them. Even though the shapes varied in color and location, participants knew that they would only have to report what shapes they saw and what order they saw them in. They were specifically told that they could only click one time for each shape that they saw (e.g., if they saw seven shapes, they were instructed to click seven times; they were allowed to click the same shape multiple times), and that the next trial would automatically start when they had pressed the correct number of shapes. The purpose of this was to ensure that for each trial we had a number of responses equal to the number of items in the sequence, thus simplifying our measure of performance. There was no counter or any other indicator reminding them how many shapes they had seen. Each time a shape was selected, it briefly flashed yellow (as a form of visual feedback, so that participants would know their response was recorded). Once a certain number of shapes were selected (equal to the number of items that had been in the previous sequence), the experiment automatically moved onto the next trial (after a 1.5 s delay).

The sequences of shapes were constructed in a few important ways. First, there were two distinct trial types, divided into two unique blocks. In the 'space-structured' block, any shape appearing multiple times appeared in the same location; no other shapes appeared in that location, and the colors of all shapes were randomized. In the 'color-structured' block, any shape appearing multiple times appeared in the same color; no other shapes appeared in that color, and the locations of all shapes were randomized. The number of colors and locations were matched (4). The first three shapes of each sequence were always unique; in other words, participants always saw all three shapes within the first three. The remaining two to four shapes were random in every respect (except that they adhered to the relevant structure, depending on the block).

Each participant completed 48 trials, divided into two equal, counterbalanced blocks: a space-structured block, and a color-structured block. In each of these blocks, participants completed 24 trials (3 difficulties [5, 6, or 7 shapes] × 8 unique trials). Between the two blocks, a message appeared encouraging participants to rest briefly before continuing. Participants completed one representative practice trial (the data from which were not recorded) before beginning the task. Including instructions and practice trials, the total task duration was about 25 min.

In Experiment 1a, but not in Experiment 1b or subsequent experiments, participants were explicitly cued to the relevant structure. They were specifically told that, although both color and location information were irrelevant to their task, they were free to use this information if it benefited them. We explained the way that color and location would be structured, in general, and told them that the two blocks of trials would be distinct in this way. However, participants were also reminded that they could disregard or ignore this information as they wished.

4.2. Results & discussion

Results from Experiment 1a can be seen in Fig. 2 (panels A & B). Accuracy was generally higher for space-structured trials ($M = 0.84$, $SD = 0.08$) compared to color-structured trials ($M = 0.79$, $SD = 0.11$), and this effect was most pronounced at higher set sizes. Indeed, a repeated measures ANOVA revealed a main effect of set size ($F[2,46] = 20.30$, $p < .001$, $\eta_p 2 = 0.47$), a main effect of trial type ($F[1,23] = 13.10$, $p = .001$, $\eta_p 2 = 0.36$), and an interaction between the two ($F[2,46] = 7.87$, $p = .001$, $\eta_p 2 = 0.26$). Post-hoc tests confirmed that accuracy was higher for space-structured trials than color-structured trials ($t[23] = 3.62$, $p = .001$, $d = 0.74$), and that accuracy was higher for set size 5 than 6 ($p = .003$), and higher for set size 6 than 7 ($p = .002$).

For all experiments, we calculated Bayes factors for the key experimental contrasts (i.e., between the space-structured and color-structured trials) to assess null effects. We report Bayes factors for significant results (such as ones here) for consistency. Bayes factors are reported as measure of relative evidence for an alternative hypothesis (here, a difference in accuracy between experimental conditions) relative to a null hypothesis (no difference between conditions). Whereas Bayes factors greater than 3 are considered substantial evidence in favor of the alternative hypothesis, Bayes factors less than 1/3 are considered substantial evidence in favor of the null hypothesis (see Wetzels et al., 2011). Bayes factors for the space-structured vs. unstructured comparison provided substantial evidence for the alternative hypothesis (BF = 25.47).

We also coded participants' responses during debriefing to identify whether they spontaneously identified either a verbal rehearsal (e.g., "I said the shapes in the order in my head") or spatial (e.g., "If I forgot the

[1] Our pre-registration states that the shapes should appear for 500 ms with 1000 ms between presentations. That was our intent. However, due to a bug in our experiment code, the timing was flipped. This applies to in-lab experiments (1, 2a, 3, and 4) but not online experiments (2b, 5), meaning that results replicate with both timings.

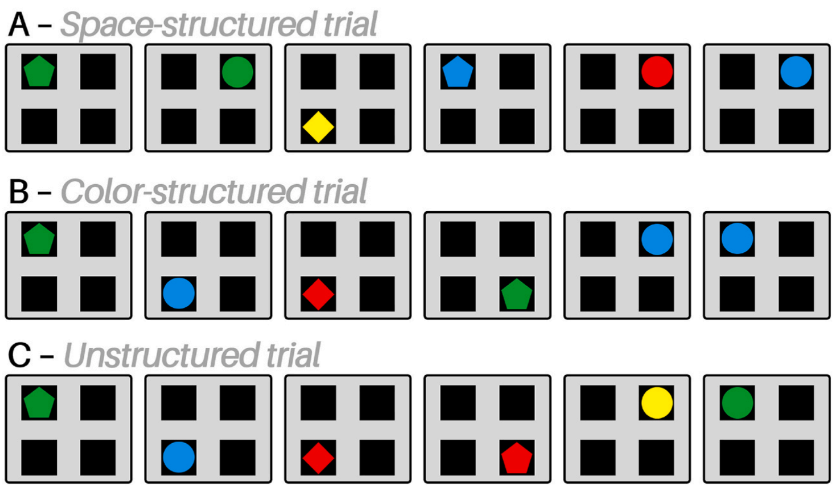

Fig. 1. Schematic of the trial structure and three two unique trial types. Shapes appeared one at a time with a brief ISI between shapes. (A) An example of a 'space-structured' trial. In this example, circles always appear in the top right corner, but their color is random. This condition was used for Experiments 1–3 and Experiment 5. (B) An example of a 'color-structured' trial. In this example, circles always appear in blue, but their location is random. This condition was used for Experiments 1–3 and Experiment 5. (C) An example of an unstructured trial. In this example, both colors and locations are random. This condition was used in Experiments 2–3 and Experiment 5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

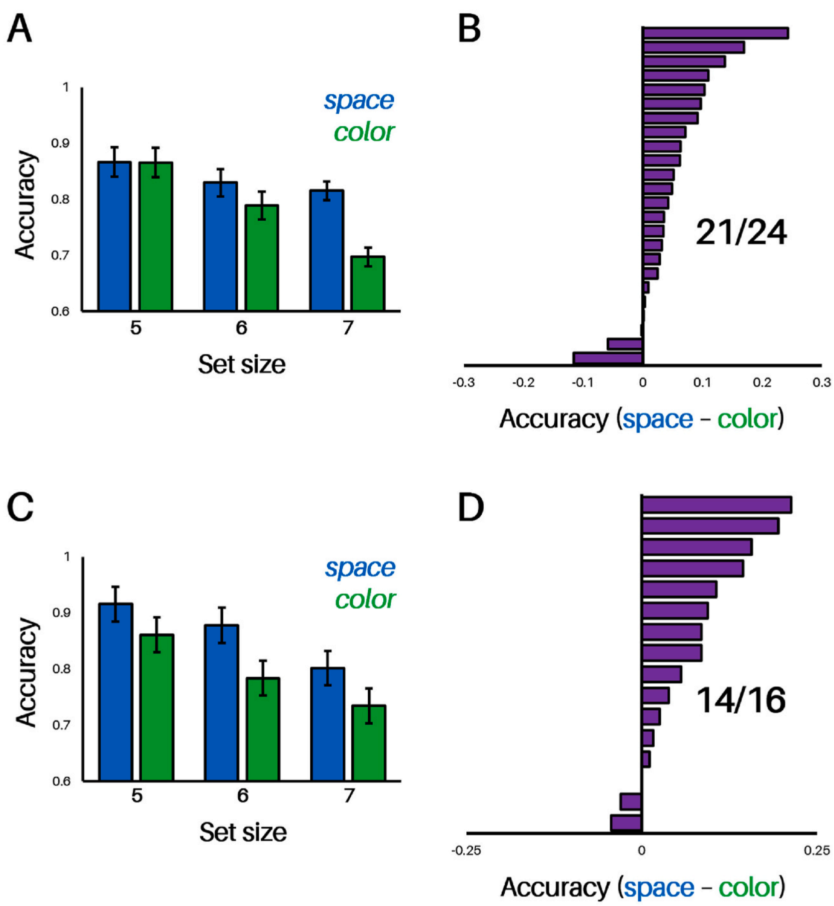

Fig. 2. Results from Experiment 1a (A & B) and 1b (C & D). On the left (A & C) average accuracy is broken down set size and by condition. On the right (B & D), difference scores are shown between the space-structured and color-structured condition for each participant. The number of participants showing the predicted effect are shown within each figure. Error bars represent +/− 1 standard error.

pattern, I would try to remember the locations") strategy. Of the 24 participants, 15 explicitly indicated the use of a verbal rehearsal strategy, whereas only 2 participants indicated the use of a spatial strategy. (We intentionally coded participants' responses in a conservative manner; the full debriefing questions and participant responses are available on our OSF page.) Therefore, these results are unlikely to results from an explicit spatial strategy.

Results from Experiment 1b can be seen in Fig. 2 (panels C & D). As is evident from the figure, accuracy was generally higher for space-structured trials ($M = 0.86$, $SD = 0.07$) compared to color-structured trials ($M = 0.78$, $SD = 0.10$), and this effect was equally pronounced at all set sizes. Indeed, a repeated measures ANOVA revealed a main effect of set size ($F[2,30] = 18.71$, $p < .001$, $\eta_p 2 = 0.56$), a main effect of trial type ($F[1,15] = 13.82$, $p = .002$, $\eta_p 2 = 0.48$), and no interaction between the two ($F[2,30] = 0.45$, $p = .64$, $\eta_p 2 = 0.03$). Post-hoc tests confirmed that accuracy was higher for space-structured trials than color-structured trials (t[15] = 3.75, p = .002, d = 0.94, BF = 21.83), and that accuracy was higher for set size 5 than 6 (p = .01), and higher for set size 6 than 7 (p = .005).

Once again, we coded participants' responses during debriefing to identify whether they spontaneously identified either a verbal rehearsal or spatial strategy. Of the 16 participants, 16 explicitly indicated the use of a rehearsal strategy (though one of these 16 reported rehearsing musical notes rather than verbal information), whereas only 1 participant indicated the use of a spatial strategy.

These experiments provide converging evidence that spatial structure benefits working memory even compared to another, matched type of structure (in that the color-structured condition, like the space-structured condition, had four options). Experiment 1b demonstrates that this is true even when participants are not cued to think about the structure at all. In fact, only 7 of the 16 participants reported noticing anything about the structure of the sequences during debriefing, and only 3 of those 7 believed that structure had anything to do with what we were testing (while 14/16 participants in the experiment showed an effect of spatial structure).

Notably, our task allows participants to verbally rehearse. Although from some perspectives this could defeat the point of studying *visual* working memory (but see 'visuospatial bootstrapping'; Darling et al., 2017), we see this as a strength of the present task. Given that participants *could* rehearse verbally (and they clearly did), an effect of spatial structure is especially notable. This spatial information is affecting visual working memory in spite of its irrelevance and in spite of participants' explicit engagement with verbal working memory, theorized to be a different sub-system (see, e.g., Allen et al., 2015; Morey, 2018). This pattern of results suggests one of two things: (1) our minds are capable of recruiting visual and verbal working memory simultaneously, as needed, or (2) visual working memory is *automatically* engaged (at least when there is salient, even if task-irrelevant, spatial information), and spatial structure boosts working memory even when participants are not explicitly relying on this information.

Like 'visuospatial bootstrapping', the results here speak to communication between verbal and visual information systems; however, *unlike* visuospatial bootstrapping, here we observe effects of short-term spatial mappings that are not stored in long-term memory (in contrast with 'bootstrapping' effects; see Darling et al., 2012). The relation between this these effects and visuospatial bootstrapping will be explored further in the following experiments.

5. Experiments 2a and 2b: space vs. color vs. unstructured

The previous results may be understood in one of several ways. For example, the results could be explained as a *benefit* of spatial structure or as a *decrement* of color structure for shape working memory. Alternatively, it could be that *both* spatial and color structure benefit working memory for shapes, but that spatial structure benefits working memory *more*. Here, we tested space and color structure vs. an unstructured baseline where both location and color were randomized.

5.1. Method

These experiments were identical to Experiment 1 except as noted. 12 participants completed Experiment 2a in exchange for course credit; this sample size was chosen based on sub-sampling of data from Experiments 1a and 1b and was pre-registered. However, anonymous reviewers raised concerns about the small sample size of Experiment 2a. As such, in Experiment 2b, we collected usable data from 158 participants via Amazon Mechanical Turk to reach the desired pre-registered sample size of 120 participants who met all of our inclusion criteria; an additional 19 participants were excluded for failing an attention check or failing to complete the correct number of trials (see below). Unlike Experiments 1a and 1b, these experiments included a third condition, in which both space and color were unstructured. As a result, Experiment 2a had 54 trials, 18 trials (3 difficulties [5, 6, or 7 shapes] × 6 unique trials) in each block. Experiment 2b had exactly half that many trials, to accommodate constraints imposed by the online format of the task. In Experiment 2b, the shapes appeared for 500 ms, with 1000 ms between presentations.

For our online experiment (2b), filters and checks were included to ensure high-quality data. At the outset, participants were eligible to complete the task if they (a) had an approval rate on Mechanical Turk greater than 98%, (b) lived in the United States, and (c) had completed at least 500 tasks. Participants were excluded prior to data analysis based on an attention check at the end of the task, in which participants were asked which shapes they saw (of 6 options). They had to select all that applied. Participants were excluded if they missed two or more items. Participants were also excluded if they failed to complete the task correctly (i.e., if they did not finish, or if they restarted partway through). Our pre-registered analysis plan also stated that we would analyze the data before and after excluding participants with at least 50% accuracy overall; this was to ensure that we had a high-powered sample with performance comparable to what we observed in a laboratory setting.

5.2. Results & discussion

Results from Experiment 2a can be seen in Fig. 3 (panels A & B). Accuracy was generally higher for space-structured trials (M = 0.88, SD = 0.09) compared to color-structured trials (M = 0.80, SD = 0.14) and unstructured trials (M = 0.77, SD = 0.15). Indeed, a repeated measures ANOVA revealed a main effect of set size (F[2,22] = 4.60, p = .02, η_p2 = 0.30), a main effect of trial type (F[2,22] = 8.15, p = .002, η_p2 = 0.43), and no interaction between the two (F[4,44] = 1.80, p = .15, η_p2 = 0.14). Post-hoc tests confirmed that accuracy was higher for space-structured trials than both color-structured trials (t[11] = 2.40, p = .04, d = 0.69, BF = 2.16) and unstructured trials (t[11] = 3.93, p = .002, d = 1.13, BF = 19.43), whereas color-structured trials and unstructured trials did not differ (t[11] = 1.43, p = .18, d = 0.41, BF = 0.65). Of the 12 participants, 11 explicitly indicated the use of a verbal rehearsal strategy, whereas only 1 observer indicated the use of a spatial strategy.

Results from Experiment 2b can be seen in Fig. 3 (panels C & D; results shown are from the final sample of 120 participants, after exclusion based on accuracy). Per our pre-registered analysis plan, we separately analyzed the data including and excluding participants with overall task accuracy greater than 50%; this was to account for the fact that Amazon Mechanical Turk pilot data revealed worse overall performance than our in-lab sample. First, we report analyses on the set of 158 participants who passed the attention checks, prior to our accuracy exclusion. Accuracy was generally higher for space-structured trials (M = 0.70, SD = 0.21) compared to color-structured trials (M = 0.66, SD = 0.21) and unstructured trials (M = 0.67, SD = 0.20). A repeated measures ANOVA revealed a main effect of set size (F[2,314] = 45.00, p < .001, η_p^2 = 0.22), a main effect of trial type (F[2,314] = 6.64, p = .002, η_p2 = 0.04), and no interaction between the two (F[4,628] = 2.34, p = .05, η_p2 = 0.02). Replicating Experiment 2a, post-hoc tests confirmed that accuracy was higher for space-structured trials than both color-structured trials (t[157] = 3.80, p < .001, d = 0.30, BF = 77.66) and unstructured trials (t[157] = 2.43, p = .016, d = 0.19, BF = 1.55), whereas color-structured trials and unstructured trials did not differ (t[157] = 0.89, p = .38, d = 0.07, BF = 0.13).

Next, we report analyses for the final set of 120 participants who met our accuracy inclusion criteria (>50%). Accuracy was generally higher for space-structured trials (M = 0.80, SD = 0.14) compared to color-structured trials (M = 0.76, SD = 0.15) and unstructured trials (M = 0.76, SD = 0.14). A repeated measures ANOVA revealed a main effect of set size (F[2,238] = 68.91, p < .001, η_p2 = 0.37), a main effect of trial type (F[2,238] = 6.58, p = .002, η_p2 = 0.05), and no interaction between the two (F[4,476] = 1.37, p = .24, η_p2 = 0.011). Again, post-

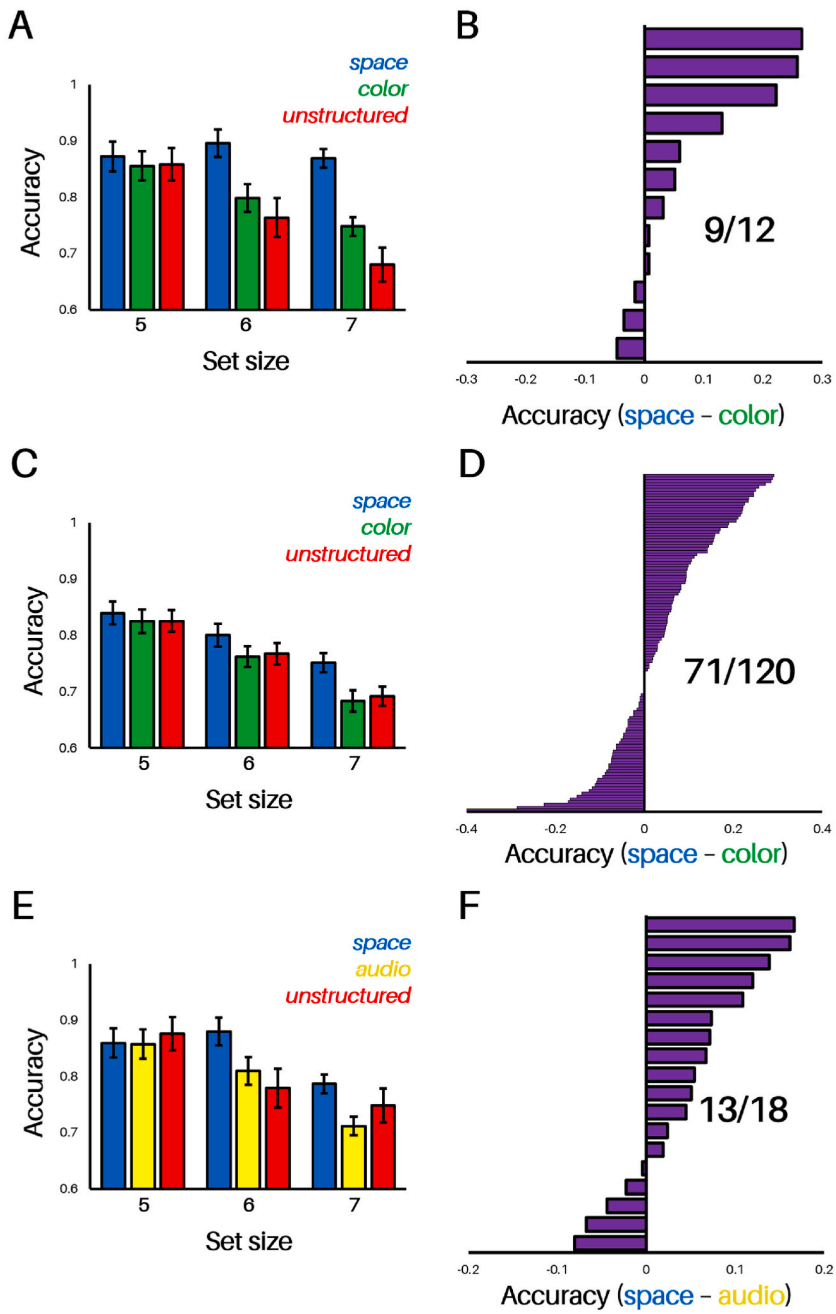

Fig. 3. Results from Experiment 2a (A & B), Experiment 2b (C & D), and Experiment 3 (E & F). On the left (A & C & E) average accuracy is broken down by set size and by condition. On the right (B & D & F), difference scores are shown between the two most relevant conditions. The number of participants showing the predicted effect are shown within each figure. Error bars represent +/− 1 standard error.

hoc tests confirmed that accuracy was higher for space-structured trials than both color-structured trials ($t[119] = 3.53$, $p < .001$, $d = 0.32$, BF = 33.13) and unstructured trials ($t[119] = 2.79$, $p = .006$, $d = 0.25$, BF = 4.02), whereas color-structured trials and unstructured trials did not differ ($t[119] = 0.39$, $p = .70$, $d = 0.04$, BF = 0.11). We did not ask our online participants about their strategies in the task.

These experiments provide converging evidence with Experiment 1 that spatial structure benefits working memory. Here, we clarify what kinds of structure matter. For example, it could have been the case that both space-structure and color-structure improve shape working memory but that space-structure does so to a larger extent. Alternatively, it could have been that space-structure does not benefit shape working memory, but that color-structure somehow interferes with shape working memory. However, it seems that neither of these accounts are true. Instead, spatial structure benefits working memory whereas there is no evidence of a color-structure benefit: although performance in the color-structure condition was numerically higher than the unstructured condition in Experiment 2a, it was actually *lower* in Experiment 2b (which had a sample size 10 times greater). This coheres with other working suggesting a privileged status of spatial information in working memory (e.g., Pertzov & Husain, 2014).

6. Experiment 3: space vs. sound vs. unstructured

The previous results establish that spatial structure benefits working memory — but is space special? One possibility is that *many* kinds of structure (i.e., repetition) benefit working memory, and that color simply isn't a salient or valuable kind of structure. Here, we compared spatial structure to *audio* structure. In other words, the relevant block featured 'audio-structured' trials in which any repeating shape was paired with the same tone each time. Is there still a greater benefit to spatial structure?

6.1. Method

This experiment was identical to Experiment 2 except as noted. 18 participants completed this experiment in exchange for course credit. This sample size was pre-registered and was chosen to be approximately identical to Experiment 1b (but rounded to a different number to account for a difference in the number of conditions). Instead of a color-structured condition, there was an audio-structure condition in which each shape was paired with a tone of a specific note. To match the number of locations, there were four possible notes: 'A', 'C', 'E', or 'G'. Matching Experiment 2a, there were 54 trials, 18 trials (3 difficulties [5, 6, or 7 shapes] × 6 unique trials) in each block. Due to the difficulty of administering audio experiments online (e.g., our inability to ensure that participants have their audio turned on, etc.), we opted not to run a replication study of this effect on Mechanical Turk.

6.2. Results & discussion

Results from Experiment 3 can be seen in Fig. 3 (panels E & F). As is evident from the figure, accuracy was generally higher for space-structured trials ($M = 0.84$, $SD = 0.08$) compared to audio-structured trials ($M = 0.79$, $SD = 0.12$) and unstructured trials ($M = 0.80$, $SD = 0.12$). Indeed, a repeated measures ANOVA revealed a main effect of set size ($F[2,34] = 18.89$, $p < .001$, $\eta_p 2 = 0.53$), a main effect of trial type ($F[2,34] = 3.47$, $p = .04$, $\eta_p 2 = .0.17$), and no interaction between the two ($F[4,68] = 2.46$, $p = .05$, $\eta_p 2 = 0.13$). Post-hoc tests confirmed that accuracy was higher for space-structured trials than both audio-structured trials ($t[17] = 2.81$, $p = .01$, $d = 0.66$, BF = 4.49) and unstructured trials ($t[17] = 2.28$, $p = .04$, $d = 0.54$, BF = 1.89), whereas audio-structured trials and unstructured trials did not differ ($t[17] = 0.35$, $p = .73$, $d = 0.08$, BF = 0.26). We also coded participants' responses during debriefing to identify whether they spontaneously identified either a verbal rehearsal or spatial strategy. Of the 18 participants, 16 explicitly indicated the use of a verbal rehearsal strategy, whereas none indicated the use of a spatial strategy.

This experiment provides converging evidence with Experiments 1 and 2 that spatial structure selectively benefits spatial working memory and further demonstrates that this benefit of structure is *unique*: neither equivalent color structure nor audio structure yielded similar benefits. Note that some other paradigms, such as those used in 'visuospatial bootstrapping' experiments (see Darling et al., 2017), are not readily adaptable to comparing spatial structure with other kinds of structure. In this way, we see the present set of studies as an extension of that research program. For example, we clarify that the 'visuospatial' aspect of visuospatial bootstrapping is uniquely important because there seems to be no evidence for an effect of 'audio' bootstrapping. Prior work supporting the 'spatial interference' view also employed primarily visual controls; in this way, the audio control of this experiment extends that research program, as well, by showing that spatial information is uniquely beneficial in working memory, not just compared to other visual cues, but compared to information in other modalities. Finally, we view this control as an especially strong one. Prior work has shown that tones can themselves be spatialized (Lidji, Kolinsky, Lochy, & Morais, 2007); therefore, one may have expected audio structure to be more useful than color structure. Nevertheless, we still observer that spatial structure is unique: audio structure did not improve retention in working memory.

7. Experiment 4: what structure matters?

Experiments 1–3 demonstrate a benefit of short-term spatial, but not color or audio, structure on working memory. But why? In previous work investigating the role of space in working memory, spatial overlap often results in memory interference (Pertzov & Husain, 2014; Treisman & Zhang, 2006). In other words, items appearing in the same location were remembered *worse* than items that appeared in different locations (but had some other overlapping feature, like color). However, many previous paradigms were unable to decouple spatial interference from spatial structure. The present paradigm has several features that enable decoupling. (1) Participants remember sequences comprised of a small set of recurring items, and (2) These items belong to distinct categories (as opposed to something like oriented lines). So, here we asked: is the effect of spatial structure caused by the *presence* of structure (i.e., the fact that any given object appears in a consistent location) or the *absence* of overlap (i.e., the fact that no two objects appear in the same location)?

To test this difference, we created two opposing conditions — an 'overlapping' condition in which different items (e.g., circle and pentagon) always appear in consistent locations but may overlap with each other, and a 'separate' condition in which different items may appear in multiple locations but will *never* overlap with each other (see Fig. 4, panels A & B). According to interference accounts, memory performance should be higher in the separate condition; although shapes appear in many unique locations, no two shapes ever overlap with one another (and thus never interfere with each other). Alternatively, the opposite might be true: the presence of structure might drive memory performance, and what matters is not whether items overlap with each other, but whether they are consistent with themselves (i.e., whether the circles always appear in one location, regardless of where the other shapes appear). This pattern of results may would be more consistent with visuospatial bootstrapping, although such studies have never tested different items overlapping in one location.

7.1. Method

This experiment was identical to Experiment 2 except as noted. 18 participants completed this experiment in exchange for course credit.[2] The color-structured and space-structured conditions were replaced with two new conditions. In a 'separate' condition, shapes could appear in any location, but no two unique shapes ever appeared in the same location (on a given trial). Although the locations were partially constrained by the shapes, there was no 'spatial structure' because the shapes did not appear in stable locations across presentations. Conversely, in an 'overlapping' condition, each shape *always* appeared in the same location, and two of the three shapes *always* overlapped with each other. To maximize the difference between the 'separate' and 'overlapping' conditions, the display was altered so that there were 6 locations (black squares) instead of 4. They were arranged in a hexagonal structure, all roughly 10.23° from the center of the screen. To account for the two new locations, there were two new colors: the default 'purple' and 'orange' in PsychoPy.

7.2. Results & discussion

Results from Experiment 4 can be seen in Fig. 4. As is evident from the figure, accuracy was generally slightly higher for 'overlapping' trials ($M = 0.85$, $SD = 0.06$) compared to both 'separate' trials ($M = 0.80$, $SD = 0.10$) and unstructured trials ($M = 0.80$, $SD = 0.10$). A repeated measures ANOVA revealed a main effect of set size ($F[2,34] = 22.56$, $p < .001$, $\eta_p 2 = 0.57$), a main effect of trial type ($F[2,34] = 4.46$, $p = .02$, $\eta_p 2 = 0.21$), and no interaction between the two ($F[4,68] = 0.04$, $p = .99$, $\eta_p 2 = 0.003$). Post-hoc tests confirmed that accuracy was higher

[2] The pre-registration incorrectly states we sought a sample size of 16. However, given that there are three conditions that appear in a counterbalanced order, the sample size must be divisible by 3. The intended sample size was 18, to match the sample size of Experiment 3.

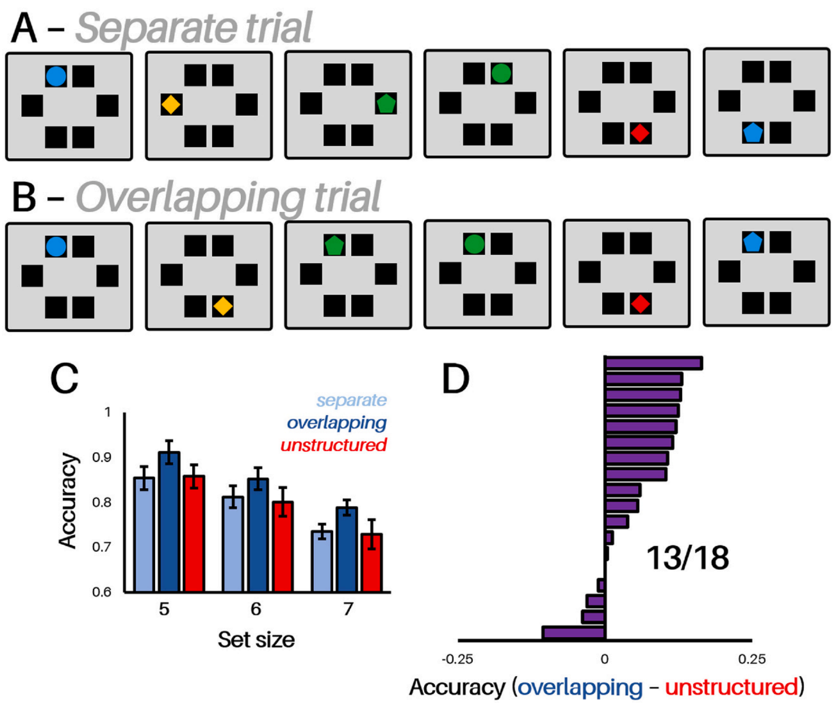

Fig. 4. Schematic and results from Experiment 4. (A) A depiction of a 'separate' trial. In this example, each shape appears in different locations, but no two shapes ever overlap with one another. (B) A depiction of an overlapping trial. In this example, all shapes appear in the same locations each time they appear, but the circle and pentagon appear in the same location. (C) Average accuracy is broken down by set size and by condition. (D) Difference scores are shown between the two most relevant conditions. The number of participants showing the predicted effect are shown within the figure. Error bars represent +/− 1 standard error.

for 'overlapping' trials than both 'separate' trials ($t[17] = 2.25$, $p = .038$, $d = 0.53$, BF = 1.80) and unstructured trials ($t[17] = 3.11$, $p = .006$, $d = 0.73$, BF = 7.73), but no difference between 'separate' and unstructured trials ($t[17] = 0.22$, $p = .83$, $d = 0.05$, BF = 0.25). Similar to the previous experiments, most of the participants (15/18) explicitly reported a verbal rehearsal strategy, and no participants explicitly reported a spatial strategy.

In the previous experiments, there was a robust effect of spatial structure (as compared to color structure, audio structure, and no structure). Here, we asked what *kind* of structure matters. Specifically, we asked whether the effects of spatial structure were caused by the *presence* of structure (as in the 'overlapping' condition) or the *absence* of overlap (as in the 'separate' condition). We found that performance was better in the 'overlapping' condition, suggesting that the benefit seen in prior experiments may have been due to the presence of structure rather than the absence of overlap.

The findings in this experiment are different from both 'spatial interference' and 'visuospatial bootstrapping' effects. For example, the spatial interference account predicts that objects overlapping in space should lead to memory impairments; however, this is not what we found. Exactly the opposite, we found that the condition in which shapes overlapped had the best memory performance. Similarly, visuospatial bootstrapping makes no specific predictions about what we should observe in the overlapping vs. separate conditions. The predictions of this view depend on what is being bootstrapped, and what it is being bootstrapped to. One possible prediction could have been that each shape needs to be bootstrapped to a single, unique location. In this case, we would have predicted equal performance across all three conditions (because in the overlapping condition, multiple shapes are bound to the same location, and in the separate condition, individual shapes are bound to multiple locations). However, this is not what we observe. Instead, the present results clarify the process of visuospatial bootstrapping: memory benefits from binding information to *specific* locations, but not necessarily to *unique* locations.

However, the key difference in this experiment (between the overlapping vs. separate conditions) could be explained by a difference in the number of locations used across conditions. Notably, the 'overlapping' condition only ever utilized 2 of the 6 locations, whereas the 'separate' condition could have utilized up to 6 locations. Thus, the effect of spatial structure could be explained by *attention* to that structure (i.e., participants can focus on a subset of locations in the 'overlapping' condition, thus reducing attentional demands), rather than the underlying structure per se. Regardless of the underlying mechanism, these results run counter to predictions of interference accounts (e.g., Pertzov & Husain, 2014) and are unexplained by accounts that emphasize long-term spatial associations (as in visuospatial bootstrapping, e.g., Darling et al., 2017; see also Darling, Havelka, Allen, Bunyan, & Flornes, 2020). Future work could adopt a similar approach to the one taken here so as to further probe the inference account and better understand the scope of visuospatial bootstrapping.

8. Exp 5: how robust is the effect of spatial structure?

Experiment 4 address two different kinds of spatial structure. However, one key difference between conditions was the number of locations participants had to attend to: in the 'overlapping' condition, in which accuracy was highest, participants had to attend to only 2 of the 6 locations, whereas in the 'separate' condition, they had to attend to all of the possible locations. The same is true, though to a lesser degree, of our previous experiments. Given the nature of the spatial structure manipulation, participants noticing the structure could realize they need only to attend to 3 of the 4 locations. It is possible that this enhanced attention to 3 of the 4 locations explains the effect of spatial structure observed so far. Here, we address this possibility (as well as other methodological details) to provide a stronger test of the 'spatial structure' account. We make three key changes to the task: (1) the number of locations and colors was reduced to 3, to match the number of shapes; (2) the first three items on each trial always had a unique color, unique location, *and* a unique shape (previously, only the shapes had to be unique, except in the space-structured and color-structured conditions in which location and color would also be unique); and (3) we arranged the 3 locations in a line, rather than in a grid format (to account for the fact that spatial information contained two dimensions, perhaps providing an advantage over the other information types).

If the previously observed effects in the space-structured conditions were caused by one or more of these three factors, then we should not expect to observe an effect of spatial structure here. Similarly, if the lack of effects in the color structure conditions were caused by the lack of these advantages (i.e., the predictability of the first few items, or the fewer locations one needed to attend to), then we should expect to observe an effect of color structure.

8.1. Method

This experiment was identical to Experiment 2b except as noted. We collected usable data from 158 participants participated via Amazon Mechanical Turk (to reach the desired pre-registered sample size of 120 participants who met all of our inclusion criteria); another 63 participants were excluded for failing an attention check or failing to complete correct number of task trials. Note: the majority of these exclusions came from participants who completed *extra* trials, ostensibly because they refreshed the task halfway through. Because we do not have a way to know *why* they refreshed the task, we exclude all such participants. There were three key changes to the task, detailed below. All three of these changes were made to better equate the information presented across conditions.

First, the number of locations and colors was reduced to 3, to match the number of shapes. This means that, across all three conditions, participants would now only have to attend to 3 locations. Previously this was not the case. In the space-structured conditions, participants would have to attend to only 3 locations, whereas in the color-structured and unstructured conditions, participants would have to attend to up to 4 locations.

Second, the first three items on each trial always had a unique color, unique location, *and* a unique shape. Previously this was not the case. In Experiments 2a and 2b, the first three shapes were always unique, but the first three colors/locations differed across conditions. In the space-structured condition, for example, the first three locations would be unique, but the first three colors would be random (and could include repeats). The opposite was true for the color-structured condition. However, this meant that certain items appeared in slightly more predictable locations. Consider a trial in the space-structured condition. If a shape appeared in Location #1, the participant would then know that they need only to attend to Locations #2, #3, and #4 to see where the next shape will appear. If the second shape appeared in Location #2, then the participant would know that they need only to attend to Locations #3 and #4 to see where the third shape will appear. By contrast, in the color-structured condition, any object could appear at any location at any time. By ensuring that shape, location, and color were unique for the first three items, the location of each shape was equally predictable across conditions. However, this also means that space and color structure were partially confounded (because the first three items are always 'structured'), meaning that we may expect reduced effects overall.

Third, we arranged the 3 locations in a line, rather than in a grid format. This is to account for the fact that the spatial information contained two-dimensions, whereas the color information did not.

8.2. Results & discussion

Results from Experiment 5 can be seen in Fig. 5 (results shown are from the final sample of 120 participants, after exclusion based on accuracy). Per our pre-registered analysis plan, we separately analyzed the data including and excluding participants with accuracy greater than 50%; this was to address the fact that Amazon Mechanical Turk pilot data revealed worse overall performance than our in-lab sample. First, we report analyses on the set of 158 participants who passed the attention checks and completed the correct number of trials prior to our accuracy exclusion. Accuracy was generally higher for space-structured trials ($M = 0.69$, $SD = 0.21$) compared to color-structured trials ($M = 0.67$, $SD = 0.21$) and unstructured trials ($M = 0.65$, $SD = 0.22$). A repeated measures ANOVA revealed a main effect of set size ($F[2,314] = 41.99$, $p < .001$, $\eta_p 2 = 0.02$), a main effect of trial type ($F[2314] = 5.00$, $p = .007$, $\eta_p 2 = 0.003$), and no interaction between the two ($F[4,628] = 0.54$, $p = .71$, $\eta_p 2 < 0.001$). Post-hoc tests confirmed that accuracy was higher for space-structured trials than unstructured trials ($t[157] = 3.14$, $p = .002$, $d = 0.25$, BF = 9.70). However, space-structured trials and color-structured trials ($t[157] = 1.51$, $p = .13$, $d = 0.12$, BF = 0.27) as well as color-structured trials and unstructured trials ($t[157] = 1.67$, $p = .10$, $d = 0.13$, BF = 0.34) did not differ.

Next, we report analyses for the final set of 120 participants who met our accuracy inclusion criteria (>50%). Accuracy was generally higher for space-structured trials ($M = 0.78$, $SD = 0.14$) compared to color-structured trials ($M = 0.76$, $SD = 0.14$) and unstructured trials ($M = 0.74$, $SD = 0.17$). A repeated measures ANOVA revealed a main effect of set size ($F[2,238] = 42.98$, $p < .001$, $\eta_p 2 = 0.27$), a main effect

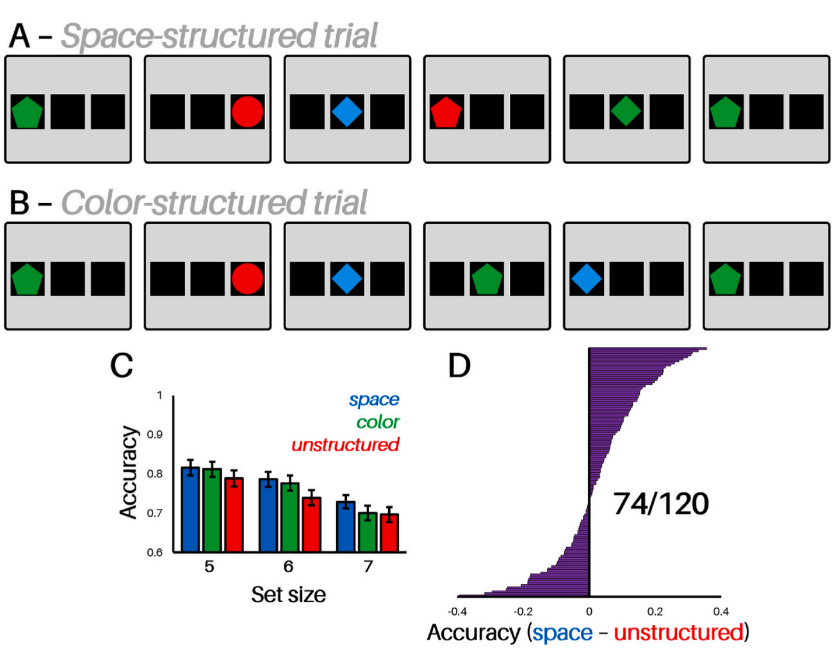

Fig. 5. Schematic and results from Experiment 5. (A) A depiction of a space-structured trial. (B) A depiction of a color-structured trial. Here, unlike previous experiments, there were only three locations, and those three locations appeared in a line. Note that locations are randomized in the color-structured condition and colors in the space-structured condition are randomized, such that it is possible for items to repeat in the same location/color. (C) Average accuracy is broken down by set size and by condition. (D) Difference scores are shown between the space and unstructured conditions. The number of participants showing the predicted effect are shown within the figure. Error bars represent +/− 1 standard error.

of trial type ($F[2,238] = 3.76$, $p = .025$, $\eta_p 2 = 0.03$), and no interaction between the two ($F[4,476] = 0.59$, $p = .67$, $\eta_p 2 = 0.005$). Post-hoc tests confirmed that accuracy was higher for space-structured trials than unstructured trials ($t[119] = 2.76$, $p = .007$, $d = 0.25$, BF = 3.78), however space-structured trials and color-structured trials ($t[119] = 1.07$, $p = .29$, $d = 0.10$, BF = 0.18) as well as color-structured trials and unstructured trials ($t[119] = 1.63$, $p = .11$, $d = 0.15$, BF = 0.37) did not differ. Thus, Bayes factors provided substantial evidence in favor of the alternative hypothesis for the space-structured vs. unstructured contrast, substantial evidence in favor of the null hypothesis for the space-structured vs. color-structured contrast, and moderate evidence in favor of the null hypothesis for the color-structured vs. unstructured contrast. We did not ask online participants about their explicit strategies in the task.

This experiment was designed to provide a strong test of the spatial structure account, by equating the information presented across conditions as much as possible. Despite this, we nevertheless observe a robust effect of spatial structure, and we do *not* observe an effect of color structure.

Interestingly, the spatial-structure and color-structure conditions did not differ from one another. There are two ways to interpret this null effect. First, color structure does benefit working memory performance, despite the null effect in Experiments 2a, 2b, and here, and that we were underpowered to detect such an effect. Yet this experiment and Experiment 2b have a combined 316 participants, and we failed to detect a reliable effect. This would be extremely unlikely if color-structure had a true effect; i.e., we consistently found moderate to strong evidence in favor of the null hypothesis for the relevant comparisons. Second, color structure does *not* benefit working memory performance, but performance in the Experiment 5 color-structured condition benefits from the spatial structure of the first three items. One of the key changes made in this experiment, in contrast with Experiments 2a and 2b, was the fact that the first three items always have a unique color and location. In this way, there will be a non-trivial number of trials in the color-structured condition, which have spatial structure (because, by random chance, shapes later in the sequence may appear in their initial locations; this will be more likely for the lower set sizes). This could cause smaller differences between conditions overall (compared to Experiment 2b). Note, however, that Experiments 2a and 2b *were* explicitly designed to better de-confound these conditions. In those cases (perhaps for this reason, or for some of the other reasons we have mentioned here), effects between conditions were much larger. Looking ahead, future work can independently test the effect of each factor (the number of spatial dimensions, the number of possible locations, the predictability of each item's location) on working memory maintenance.

9. General discussion

These experiments investigated the interactions between spatial and verbal information in working memory, as well as the specific role of spatial information in working memory. Across all seven experiments, 'spatial structure' improved memory. This enhancement existed both when participants were cued to this structure (Experiment 1a) and when they were not (Experiment 1b, 2a, 2b, and 5). The benefit of space is unique (compared to other visual features, like color, and also auditory features, like tones; Experiment 3). Further, these effects may be best understood as caused by the *presence* of structure, rather than an *absence* of overlap (Experiment 4). Finally, we showed that these effects are generally robust to various changes in the experimental design (Experiment 5).

All told, these data inform several aspects of working memory. *First*, these experiments illustrate that although people seem to have the default tendency to explicitly engage in verbal rehearsal (i.e., 83% of participants who were queried reported a rehearsal strategy), they nevertheless benefit from spatial structure; this may be seen as broadly consistent with the fact that visuospatial bootstrapping also does not depend on 'executive resources' (Calia, Darling, Havelka, & Allen, 2019). *Second*, these experiments support the view that visuospatial and verbal information interact to facilitate memory. Despite the irrelevance of spatial information, it influenced memory (as in visuospatial bootstrapping; Darling et al., 2017). In this way, these findings also constrain working memory models. For example, they are potentially at odds with the view that working memory is supported by two independent subsystems for verbal and visual working memory (e.g., Baddeley, 1992). *Third*, these experiments show that space is special compared to seemingly equivalent *visual* features, like color (see also Pertzov & Husain, 2014), and also compared to equivalent non-visual features, like auditory tones. *Fourth*, these experiments shed light on *how* space influences memory. Whereas previous work emphasized either 'spatial interference' and the role that space plays in binding features to objects (Pertzov & Husain, 2014; Treisman & Zhang, 2006), or 'visuospatial bootstrapping' and the way that spatial information facilitates memory (Darling et al., 2017), the present work asks specifically about spatial *structure*. Here, the presence of spatial structure, not the absence of overlap, influences working memory (seemingly in contrast with prior accounts). The results presented here are not mutually exclusive with either spatial interference accounts or visuospatial bootstrapping, but they do present some conflicting results. For example, an interference account should predict that memory performance in the two key conditions in Experiment 4 (separate vs. overlapping) would be equal, yet that is not what we observed; the absence of overlap alone is not all that influences memory. In contrast, the data here imply that stability (or, structure) benefits working memory, regardless of overlap. Furthermore, visuospatial bootstrapping has been said to depend on associations in long-term memory (Darling et al., 2012; but see Darling et al., 2020); however, the present effects could be seen as effects of visuospatial bootstrapping that do not depend on long-term spatial associations. We suggest that the findings in prior work may not be because bootstrapping depends on long-term associations, but rather than long-term associations interfered with the ability to make new associations over that same configuration (as when rearranging a keypad, for instance; Darling et al., 2012).

We have highlighted one key difference between our account and a spatial interference account. In Experiment 4, items overlapping with one another (in our 'overlapping' condition) are better remembered than items that never overlap (in our 'separate' condition). At first blush, these results seem directly at odds with a spatial interference account. Consider, for example, the view that features are bound to objects via space (e.g., Treisman & Zhang, 2006). What should be made of better memory for overlapping objects? Perhaps both accounts are correct and the 'interference' predicted in our task does not involve the shapes themselves but instead involves the binding of color and shape. For example, if we had a separate measure of color memory performance, participants would have worse color memory (but better shape memory) in the overlapping condition. Yet even if that were true, a spatial interference account would not necessarily predict better memory for shapes in the overlapping condition. In this way, we see some of the results presented here as highlighting gaps in our understanding of classic location binding results (see also Pertzov & Husain, 2014; Rajsic & Wilson, 2014) — raising questions about the underlying mechanisms of feature binding and motivating future work.

Another possibility is that these results do not contradict the basic idea of spatial interference, but instead speak to many different *kinds* of possible interference. Various types of interference have been conceptualized, each with unique consequences. For example, interference by feature overwriting predicts that similar items are more likely to interfere with one another, whereas interference by superposition predicts that different items are more likely to interfere with one another (Oberauer, Farrell, Jarrold, & Lewandowsky, 2016). In other words, it is possible that the seemingly different patterns of results observed in our studies compared to previous interference studies may come down to the details of the stimuli themselves. For example, maybe "circle" and "pentagon" are more similar to the mind than "blue" and "green", resulting in less interference for shape than color. This highly speculative possibility can be addressed by future work quantifying the relative similarity of different stimuli.

We see these results as generally more consistent with the idea of visuospatial bootstrapping (Darling & Havelka, 2010), although with a few key differences. First, unlike classic bootstrapping designs, the patterns here do not depend on long-term spatial mappings; the 'spatial structure' defined here is confined to, at most, a ~ 10s trial (but see Darling et al., 2020). Second, we test cases where items not only possess stable spatial mappings, but also *share* stable spatial mappings with other objects (Experiment 4). While these results are not necessarily at odds with visuospatial bootstrapping, they provide unique insight not obvious in the canonical bootstrapping designs. Third, whereas visuospatial bootstrapping studies often focus on a contrast between spatial structure vs. no structure, we *compare* spatial structure to two other forms of structure (color structure and audio structure). In so doing, we show that the effect of spatial structure is highly specific. Concretely, this means that the term 'visuospatial bootstrapping' may be an apt name; our results demonstrate, e.g., that visual information alone (in the form of color information) does not result in the same form of bootstrapping. Finally, our results show that the notion of 'spatial structure' (whether in the form of visuospatial bootstrapping, or otherwise) is highly robust: we replicate the same key effect 7 times, after numerous critical manipulations to the basic paradigm.

10. Possible explanations

Three possible explanations for these data come to mind. The first is *simplicity*: perhaps performance in the space-structured condition was better than performance in the unstructured and color-structured condition because the task was easier. One may note, for example, that in Experiments 1–3, all four of the locations were never used in a single space-structured trial (because there were fewer shapes than locations), whereas all four locations could potentially be engaged in a single color-structured or unstructured trial (because no constraints were imposed on spatial location). However, Experiment 5 rules out this possibility: when the number of shapes and locations was equated, such that it all three conditions participants would always have to attend to all of the visible locations, there was nevertheless a benefit of spatial structure relative to no structure. Although questions remain about how exactly spatial structure benefits working memory, Experiment 5 demonstrates that the primary result reported here (a difference between spatial structure and no structure) cannot be explained solely by a difference in the number of locations.

The second possible explanation is *predictability*: could our results be explained by some difference in the predictability of the sequences? One may note, for example, that the location of the second and third shapes in the space-structured conditions were partially predictable; if spatial information is structured, and participants are aware of that structure, then they know that the second and third shapes cannot appear where the prior shapes had. Although we cannot rule out this possibility for the earlier experiments, Experiment 5 was specifically designed to equate predictability. There, the second and third items were equally predictable across conditions; nevertheless, we observe a benefit of spatial structure relative to no structure.

The third possible explanation is *eye movements*. In our task, participants are free to move their eyes around the screen as they wish. Spatial location per se might not influence working memory but instead eye movements may lead directly to differential encoding. Indeed, eye movements do influence working memory, at least at retrieval (e.g., Awh et al., 1998; Theeuwes, Belopolsky, & Olivers, 2009). Yet, even controlling eye gaze (and thus overt shifts of attention) by having subjects fixate would not rule out the possibility that covert shifts of spatial attention nevertheless could explain the present results. Thus, future work might characterize the effect of eye movements on the benefits of spatial structure for working memory maintenance.

11. Conclusion

Space may be *foundational* to working memory: not only does spatial structure benefit working memory, it does so even when that information is task-irrelevant, and even when participants rely on distinctly non-spatial strategies (e.g., verbal rehearsal). Further, spatial structure seems unique in its influence; neither non-spatial visual structure (color) nor non-visual auditory structure benefitted memory, even compared to an unstructured baseline. These results raise questions about the nature of working memory, its subsystems, and their interactions, whilst emphasizing the importance of spatial *structure* in working memory.

Author note

For helpful conversation and/or comments on previous drafts, we thank Brynn Sherman and the members of the Yale Cognition and Development Laboratory. For thorough, thoughtful, and critical feedback during the review process, we thank Steve Darling. This project was funded by an NSF Graduate Research Fellowship awarded to SRY.

Author contributions

S. R. Yousif designed the experiments, analyzed the data, and wrote the manuscript with extensive, critical input from M. D. Rosenberg and F. C. Keil.

Supplementary materials

Pre-registrations for all experiments as well as experiment code, raw data, and analyses, can be accessed here: https://osf.io/85sjd/

References

Abrahamse, E., Van Dijck, J. P., Majerus, S., & Fias, W. (2014). Finding the answer in space: The mental whiteboard hypothesis on serial order in working memory. *Frontiers in Human Neuroscience, 8*, 932.

Allen, R. J., Havelka, J., Falcon, T., Evans, S., & Darling, S. (2015). Modality specificity and integration in working memory: Insights from visuospatial bootstrapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 820–830.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*, 1698–1716.

Aulet, L. S., Yousif, S. R., & Lourenco, S. F. (2021). Spatial–numerical associations from a novel paradigm support the mental number line account. In *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218211008733.

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 780.

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556–559.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *, Vol. 8. Psychology of learning and motivation* (pp. 47–89). Academic press.

Calia, C., Darling, S., Havelka, J., & Allen, R. J. (2019). Visuospatial bootstrapping: Binding useful visuospatial information during verbal working memory encoding does not require set-shifting executive resources. *Quarterly Journal of Experimental Psychology, 72*, 913–921.

Cowan, N. (1998). Visual and auditory working memory. *Trends in Cognitive Sciences, 2*, 77–78.

Darling, S., Allen, R. J., & Havelka, J. (2017). Visuospatial bootstrapping: When visuospatial and verbal memory work together. *Current Directions in Psychological Science, 26*, 3–9.

Darling, S., Allen, R. J., Havelka, J., Campbell, A., & Rattray, E. (2012). Visuospatial bootstrapping: Long-term memory representations are necessary for implicit binding of verbal and visuospatial working memory. *Psychonomic Bulletin & Review, 19*, 258–263.

Darling, S., & Havelka, J. (2010). Visuospatial bootstrapping: Evidence for binding of verbal and spatial information in working memory. *Quarterly Journal of Experimental Psychology, 63*, 239–245.

Darling, S., Havelka, J., Allen, R. J., Bunyan, E., & Flornes, L. (2020). Visuospatial bootstrapping: Spatialized displays enhance digit and nonword sequence learning. *Annals of the New York Academy of Sciences, 1477*, 100–112.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*, 371.
van Dijck, J.-P., Abrahamse, E. L., Acar, F., Ketels, B., & Fias, W. (2014). A working memory account of the interaction between numbers and spatial attention. *Quarterly Journal of Experimental Psychology, 67*, 1500–1513.
van Dijck, J. P., & Fias, W. (2011). A working memory account for spatial–numerical associations. *Cognition, 119*, 114–119.
Guida, A., & Campitelli, G. (2019). Explaining the SPoARC and SNARC effects with knowledge structures: An expertise account. *Psychonomic Bulletin & Review, 26*(2), 434–451.
Guida, A., Leroux, A., Lavielle-Guida, M., & Noël, Y. (2016). A SPoARC in the dark: Spatialization in verbal immediate memory. *Cognitive Science, 40*, 2108–2121.
Guida, A., Megreya, A. M., Lavielle-Guida, M., Noël, Y., Mathy, F., van Dijck, J. P., & Abrahamse, E. (2018). Spatialization in working memory is related to literacy and reading direction: Culture "literarily" directs our thoughts. *Cognition, 175*, 96–100.
Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 683–702.
Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1189–1207.
Morey, C. C. (2018). The case against specialized visual-spatial short-term memory. *Psychological Bulletin, 144*, 849–883.
Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin, 142*, 758–799.
Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology, 64*, 77–107.
Parkinson, C., & Wheatley, T. (2013). Old cortex, new contexts: Re-purposing spatial perception for social cognition. *Frontiers in Human Neuroscience, 7:645*, 1–7.
Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203.
Pertzov, Y., & Husain, M. (2014). The privileged role of location in visual working memory. *Attention, Perception, & Psychophysics, 76*, 1914–1924.
Race, E., Palombo, D. J., Cadden, M., Burke, K., & Verfaellie, M. (2015). Memory integration in amnesia: Prior knowledge supports verbal short-term memory. *Neuropsychologia, 70*, 272–280.
Rajsic, J., & Wilson, D. E. (2014). Asymmetrical access to color and location in visual working memory. *Attention, Perception, & Psychophysics, 76*, 1902–1913.
Theeuwes, J., Belopolsky, A., & Olivers, C. N. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica, 132*, 106–114.
Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition, 34*, 1704–1719.
Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t-tests. *Perspectives on Psychological Science, 6*, 291–298.
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*, 1–28.
Zorzi, M., Priftis, K., & Umiltà, C. (2002). Brain damage: Neglect disrupts the mental number line. *Nature, 417*, 138–139.