Representation of dynamic spatial configurations in visual short-term memory

Frank Papenmeier · Markus Huff · Stephan Schwan

Abstract Locations of multiple stationary objects are represented on the basis of their global spatial configuration in visual short-term memory (VSTM). Once objects move individually, they form a global spatial configuration with varying spatial inter-object relations over time. The representation of such dynamic spatial configurations in VSTM was investigated in six experiments. Participants memorized a scene with six moving and/or stationary objects and performed a location change detection task for one object specified during the probing phase. The spatial configuration of the objects was manipulated between memory phase and probing phase. Full spatial configurations showing all objects caused higher change detection performance than did no or partial spatial configurations for static and dynamic scenes. The representation of dynamic scenes in VSTM is therefore also based on their global spatial configuration. The variation of the spatiotemporal features of the objects demonstrated that spatiotemporal features of dynamic spatial configurations are represented in VSTM. The presentation of conflicting spatiotemporal cues interfered with memory retrieval. However, missing or conforming spatiotemporal cues triggered memory retrieval of dynamic spatial configurations. The configurational representation of station-ary and moving objects was based on a single spatial configuration, indicating that static spatial configurations are a special case of dynamic spatial configurations.

F. Papenmeier · S. Schwan
Knowledge Media Research Center,
Tübingen, Germany

M. Huff
University of Tübingen,
Tübingen, Germany

F. Papenmeier (✉)
Knowledge Media Research Center,
Schleichstraße 6,
72076 Tübingen, Germany
e-mail: f.papenmeier@iwm-kmrc.de

The environment is highly dynamic, confronting people constantly with moving objects. Walking down an avenue one might see several children running about a playground. Each time one passes by a tree, the view of the playground is covered. With the view covered, there remains some memory of the events happening on the playground. Is this memory restricted to the locations and movement of the individual children, or is the dynamically changing spatial configuration of the children represented too? With the present research, we investigate this question. In particular, we examine whether the global spatial configuration of moving objects is represented in visual short-term memory (VSTM), and we characterize certain aspects of their representation.

In most research on VSTM, the processing of stationary stimuli has been examined. Findings have concerned the capacity of short-term memory (e.g., Cowan, 2001; Sligte, Scholte, & Lamme, 2008; Zhang & Luck, 2008), the binding of features to objects (e.g., Luck & Vogel, 1997; Wheeler & Treisman, 2002), or the relevance of global spatial configurations (e.g., Blalock & Clegg, 2010; Hollingworth, 2007; Jiang, Chun, & Olson, 2004; Jiang, Olson, & Chun, 2000). Jiang et al. (2000) emphasized the importance of the objects' global spatial configuration for the encoding of locations, as well as features such as color and shape. Like most research on VSTM, they studied stationary objects and, thus, static spatial configurations.

Once objects move individually, their spatial inter-object relations vary over time, thus creating a dynamic spatial configuration. The dynamics of the global spatial configuration of moving objects make the processing of dynamic scenes qualitatively different from that of static scenes and, therefore, do not allow for a generalization of results, especially with respect to the importance of spatial configurations in VSTM. To broaden our understanding of VSTM, it is therefore essential to study empirically whether dynamic spatial configurations of moving objects are represented in VSTM and how they relate to corresponding results of spatial configurations of stationary objects.

Spatial configurations

Jiang et al. (2000) studied the role of spatial configurations of stationary stimuli in VSTM, using a modified change detection paradigm (e.g., Phillips, 1974). Their participants encoded an array of multiple stationary objects and were instructed to ignore the spatial configuration of all objects and to encode each object location individually. Following a retention interval of 907 ms, participants performed a location change detection task for one particular object highlighted by a red frame. The critical manipulation involved the type of spatial configuration present during the probing phase. The nonprobed objects were invisible, shown at their original locations, or shown at displaced locations, or only a subset of nonprobed objects were visible and shown at their original locations. The availability of the full spatial configuration (visible nonprobed objects at their original locations) resulted in the highest performance, demonstrating the representation of spatial configurations in VSTM. Since performance was not higher with partial spatial configurations (a subset of visible nonprobed objects) than without spatial configurations (the probed object alone), object locations are encoded into a global spatial configuration of all objects. Jiang et al. (2000) concluded that spatial configurations are obligatorily processed, because even the addition of a condition with distorted spatial configurations (displaced nonprobed objects) and the explicit instruction to ignore the spatial configuration of all objects did not eliminate the higher performance with full configurations, as compared with no spatial configurations. This conclusion is further supported by a study on the influence of irrelevant grouping cues on performance in a location change detection task (Jiang et al., 2004). If the irrelevant grouping cues were changed between the memory and the probe images, the location change detection performance was impaired. Previous research also demonstrated that the representation of spatial configurations is not restricted to the simultaneous presentation of multiple static objects. With sequential presentations showing one object at a time, participants were able to extract the global spatial configuration too (Blalock & Clegg, 2010). Since participants received an explicit instruction to encode the spatial layout in that study, it remains unclear whether the participants would also have done so spontaneously.

All these results account for the encoding of a set of stationary objects not changing their spatial inter-object relations over time. Once objects start moving, they form a dynamic spatial configuration. Even though the representation of such dynamic spatial configurations in VSTM has not been examined yet, there is some evidence from the literature on visual attention that spatial configurations can be extracted from moving object locations. Evidence for this comes from a multiple-object tracking study in which participants tracked multiple objects moving on a square floor plane in a three-dimensional scene (Huff, Jahn, & Schwan, 2009). At one point in time, an abrupt rotation of the scene around the center of the floor plane occurred. Displacements of objects in retinal and scene coordinates were smaller the closer they were located to the center of the floor plane at the time of the abrupt scene rotation. The size of the displacement of the individual target objects did not affect how well they were tracked. Instead, displacements of the whole scene determined tracking performance, thus demonstrating that the objects were processed as a whole scene. This suggests that the spatial configuration of the moving objects was processed. The extraction of dynamic spatial configurations from dynamic scenes is further supported by eye-tracking studies during multiple-object tracking tasks (Fehd & Seiffert, 2008; Huff, Papenmeier, Jahn, & Hesse, 2010). They show a substantial amount of gaze time spent on the centroid of the target group, indicating target grouping during multiple-object tracking (see also Yantis, 1992) that corresponds to the use of dynamic spatial configurations. Drawing this analogy from the literature on attention to possible processes in VSTM, it is plausible to assume that dynamic spatial configurations might be represented.

Dynamics in visual memory

Early studies on the memory of visually perceived moving stimuli were interested in the processing of body movements. They identified a memory store for body movement configurations separate from the visuospatial component of working memory. These studies measured the memory span for sequences of multiple whole-body movement patterns (Smyth, Pearson, & Pendleton, 1988; Smyth & Pendleton, 1990) and local hand movement patterns (Smyth & Pendleton, 1989) with different dual-task conditions. The memory span was reduced by a concurrent movement task

only. Neither a concurrent spatial task nor a concurrent articulatory suppression task had an effect. The identified memory store has a capacity of about three movement actions (Wood, 2007). These results on the representation of body movements suggest two interesting aspects that might also hold true for the memory representation of dynamic spatial configurations. First, these results demonstrate that complex movement configurations are successfully maintained in memory. Therefore, dynamic spatial configurations of multiple objects in motion might also be encoded in VSTM. And second, complex dynamic configurations might be stored separately from static spatial configurations. Therefore, moving and stationary objects might be stored in separate spatial configurations in VSTM.

There are further studies suggesting a separation of VSTM for static and dynamic information. They investigated performance in static and dynamic versions of a visual pattern task across different age groups (Pickering, Gathercole, Hall, & Lloyd, 2001). These studies found a developmental dissociation for static and dynamic information. Developmental increases occurred faster with the static version of the task than with the dynamic version. These results suggest that moving and stationary objects might be stored in separate subsystems and, therefore, separate spatial configurations.

However, some mental representations might be dynamic in nature by incorporating a temporal dimension (Freyd, 1987). For example, participants viewing static pictures from unidirectional action scenes extract dynamic information, as was shown by the longer response times needed to correctly reject distractors depicting future, rather than past, stills from a scene (Freyd, 1983). Dynamic mental representations are also extracted from more simple visual stimuli, such as a rotating rectangle, as is known from the representational momentum effect (Freyd & Finke, 1984). The representation of spatial configurations might, therefore, be dynamic in nature by incorporating a temporal dimension. This would lead to the prediction that the representation of spatial configurations for static visual displays (Jiang et al., 2000) might be a special case of the representation of spatial configurations for dynamic visual displays. Therefore, one would predict that stationary and moving objects are stored in a single spatial configuration, instead of in separate spatial configurations.

The present study

Summarizing, spatial locations of stationary objects are stored on the basis of their global spatial configuration in VSTM. Once multiple objects in a display move individually, they form a spatial configuration that changes continually over time. Even though the encoding of such dynamic spatial configurations into VSTM has not been studied yet, there is evidence suggesting that they might be represented. We studied the representation of dynamic spatial configurations in VSTM, adapting the modified location change detection paradigm developed by Jiang et al. (2000). Participants memorized a dynamic scene and were probed for the location change of one individual object highlighted in the probing phase. We varied the type of spatial configuration present in the probing phase. Since participants did not know which specific object was going to be probed for a location change, they had to encode all the objects in the memory phase. The critical question studied with two experiments in Section 1 was whether these moving objects are represented individually in VSTM or whether the spatial configuration formed by the objects is represented too. If spatial configurations of moving objects are represented in VSTM, the presence of global spatial configurations in the probing phase should improve location change detection performance for individual objects. Since we will provide evidence for the configurational encoding of moving objection, we further characterize the representation with four additional experiments in Section 2. In particular, we demonstrate that the spatiotemporal features of dynamic spatial configurations are part of their VSTM representation. Nonetheless, spatial cues without spatiotemporal features were also able to trigger their memory retrieval if the probed object itself was stationary during retrieval. Furthermore, we show that stationary and moving objects are represented in a single spatial configuration in VSTM. We propose that the representation of static and dynamic spatial configurations is based on the same system.

Section 1: Spatial configurations with dynamic scenes

There is evidence that spatial configurations are spontaneously extracted from dynamic scenes (Fehd & Seiffert, 2008; Huff et al., 2010; Yantis, 1992). This suggests that they might also be essential for the VSTM representation of dynamic scenes. However, this has not been studied, yet. Experiments 1a and 1b fill this empirical gap. With Experiment 1a, we demonstrate that spatial configurations of dynamic scenes are represented in VSTM. We extend these results, by showing that there is a strong tendency to represent spatial configurations of dynamic scenes in VSTM by discouraging their encoding, with the addition of distorted spatial configurations in the probing phase of Experiment 1b.

Experiment 1a

After memorizing a *dynamic scene*, participants had to detect a location change of one individual object. If the

spatial configuration of all objects is represented in VSTM, we should see a performance benefit with the presence of the full configuration, as compared with no spatial configuration in the probing phase. We should not see a performance benefit for partial spatial configuration in the probing phase, since this does not correspond to the global configuration represented in VSTM. In control conditions, we presented static stimuli (Jiang et al., 2000).

Method

Participants Twenty students from the University of Tübingen participated in this experiment in exchange for partial course credit or monetary compensation. All participants reported normal or corrected-to-normal vision.

Apparatus and stimuli In all experiments, stimuli were presented on an 18-in. screen using Blender 2.49 (http://www.blender.org/) and custom software written in Python (http://www.python.org/). We presented six green squares extending 0.9° × 0.9° of visual angle each. The squares were positioned randomly, with random movement directions within an invisible bounding box of 15.6° × 15.6° of visual angle in the center of the screen. The dynamic squares moved on linear trajectories at a constant speed of 5° of visual angle per second for 400 ms. The squares were not allowed to overlap and had a minimum spacing of 0.4° of visual angle between each other at any point in time. Dynamic squares were restricted so as not to hit the invisible boundary, to prevent direction changes. The color of the background was gray, RGB = (127, 127, 127). A demonstration of the dynamic stimuli is available online at http://www.iwm-kmrc.de/cybermedia/dynamic-configurations/. Yes/no responses were recorded using the outer buttons of a DirectIN High Speed Button-Box (http://www.empirisoft.com/), with the assignment of the buttons balanced across participants. Participants kept an unrestricted viewing distance of 65 cm to the screen.

Procedure and design Participants' task was to detect a location change of an individual object in a scene lasting 400 ms. The timing within each trial was as follows (see Fig. 1). At the beginning of the trial, a white fixation cross was presented for 500 ms, followed by a blank screen for 100 ms. Thereafter, the memory phase in which six squares were shown either moving or stationary was presented for 400 ms. The screen turned blank for 900 ms, followed by a probing phase of 400 ms. The squares in the probing phase were stationary whenever they had been stationary in the memory phase and moved on linear trajectories in the probing phase whenever they had done so in the memory phase. Following the probing phase, a blank screen was shown until response. Feedback about the accuracy of the response was given visually as either "Richtig" ("correct") or "Falsch" ("incorrect"). During the probing phase, one square was highlighted by a red frame. Participants were instructed to answer whether this particular square occupied a former location and moved the same way as any square in the memory phase or not. Change trials were constructed by displacing the probed square to a previously unoccupied position. The displaced square conformed to the same restrictions (e.g., minimum spacing to all squares, including the original version of the probed square) that applied to regular squares mentioned above. The direction of motion was not changed. Analogous to Jiang et al. (2000), we instructed participants to memorize the squares individually, telling them that the configuration formed by the squares should be ignored, since it might impair memory accuracy for individual squares, and that they would need to answer for one highlighted square only. If the presence of spatial configurations provides a memory benefit despite this instruction, it would be a first indication that there is a strong tendency to represent spatial configurations of dynamic scenes. Participants were instructed to answer as accurately as possible without worrying about speed. Response times were measured from probe onset until response.

Fig. 1 Schematic of the general procedure of all experiments. Note that the color of the squares was green, the color of the frame around the probed square was red, the background color was gray, and the color of the fixation cross was white in all experiments
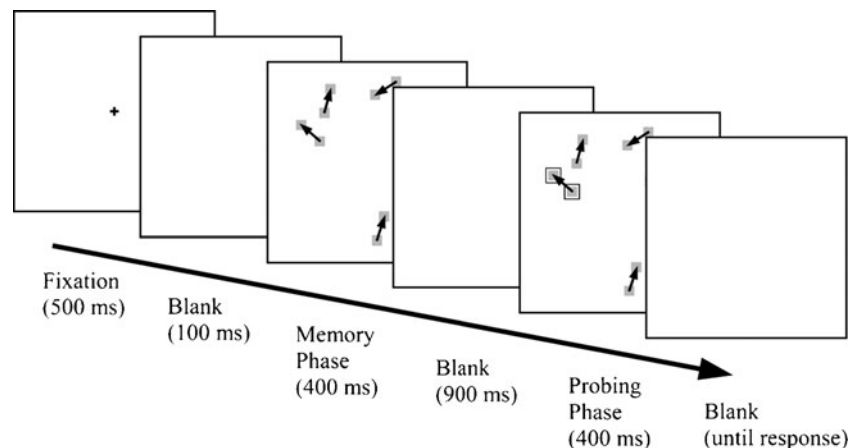
We manipulated the type of spatial configuration present in the probing phase (full, six squares; partial, three squares; no, probed square alone; see Fig. 2) and scene dynamics (moving squares, stationary squares). All squares serving as spatial configuration and not probed by the red frame were positioned at the exact positions and paths of movement occupied in the memory phase. This resulted in a 3 (spatial configuration: full, partial, no) × 2 (scene dynamics: dynamic, static) × 2 (change present: yes, no) × 20 (repetitions) within-subjects design with 240 experimental trials. At the beginning of the experiment, participants performed 24 practice trials. In the present and all subsequent experiments, the conditions were presented randomly throughout the experiment, and the conditions in the practice trials were balanced according to the experimental design.

Results

Analyses are based on the signal detection theory, and results are reported with the sensitivity measure d' as the dependent variable for location change detection perfor-

A. Memory Phase

B. Full Configuration   C. No Configuration

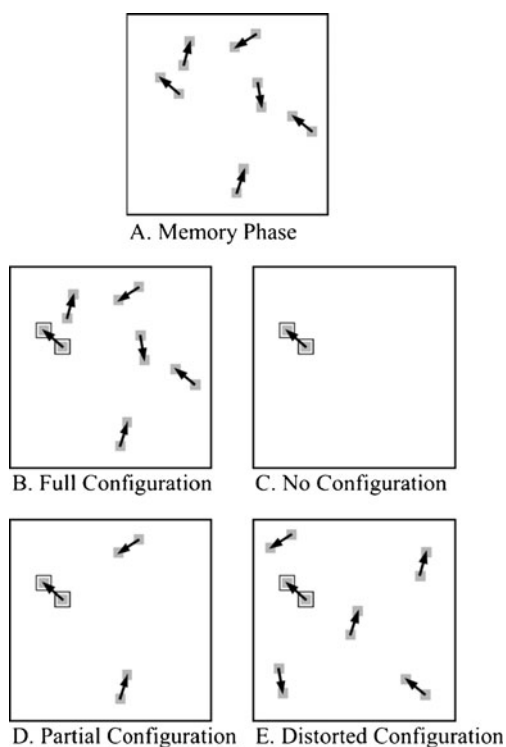D. Partial Configuration   E. Distorted Configuration

Fig. 2 Types of spatial configurations used during the probing phase in Experiments 1a, 1b, and 2b. A single object marked by a red outline in the probing phase (**b–e**) is probed for a location change, as compared with the memory phase (**a**), by showing full (**b**), no (**c**), partial (**d**), or distorted (**e**) spatial configurations. The probed object was not displaced in this example. With static scenes, the objects did not move but were stationary

mance (see Fig. 3). Since d' is not defined for hit and false alarm rates of 1.0 and 0.0, we adjusted such values to half a trial incorrect in all experiments. All trials with response times larger than 5,000 ms were considered invalid and were removed from the data set (11 trials, 0.23% of the data). We report partial eta-squared (η_p^2) as the effect size measure. Hit rates, false alarm rates, proportions correct, and response times for hits across all experiments are reported in the Appendix.

We analyzed sensitivity using a 3 (spatial configuration) × 2 (scene dynamics) repeated measures ANOVA. There was a significant main effect for the type of spatial configuration present in the probing phase, $F(2, 38) = 35.82$, $p < .001$, $\eta_p^2 = .65$, not interacting with scene dynamics, $F(2, 38) = 2.26$, $p = .118$, $\eta_p^2 = .11$. Planned contrasts using paired t-tests revealed that sensitivity was higher with full spatial configurations than with partial and no spatial configurations, both $ps \leq .001$, and that there was no significant difference between partial and no spatial configurations, $p = .518$. There was also a significant main effect for scene dynamics, $F(1, 19) = 70.77$, $p < .001$, $\eta_p^2 = .79$, with lower change detection performance for dynamic than for static scenes.

Discussion

This experiment provides the first evidence that spatial configurations of dynamic scenes are represented in VSTM. Change detection performance was higher with full spatial configurations than without spatial configurations. Full spatial configurations caused a higher change detection performance for static scenes too, thus replicating the findings of Jiang et al. (2000) with our material. The effect of spatial configurations did not differ significantly between static and dynamic scenes, as is indicated by the missing interaction effect. Thus, processes underlying the representation of static and dynamic spatial configurations might be similar. This will be further investigated in the experiments in Section 2.

The present pattern of results was obtained even though participants were explicitly instructed to ignore the spatial configuration formed by the objects. In fact, some participants reported, in a postexperiment questionnaire with the open-ended question of how they had solved the experimental task, that they could not help memorizing the spatial configuration. This provides the first evidence that there is a strong tendency to represent spatial configurations of dynamic scenes.

Partial spatial configurations were included into the present experiment in order to preclude an anchoring hypothesis (Jiang et al., 2000) for dynamic scenes. According to an anchoring hypothesis, the location change
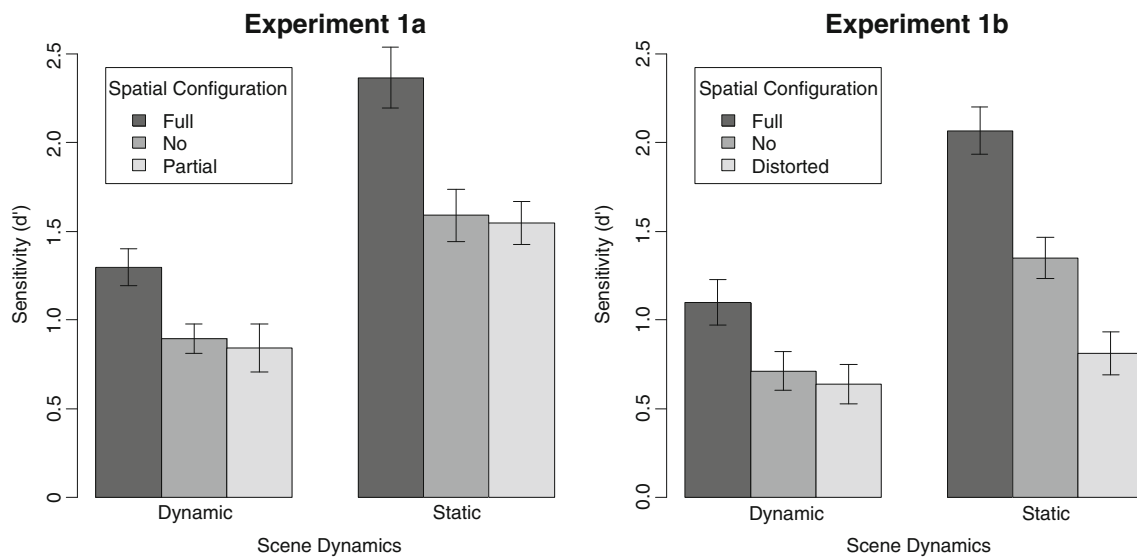
Experiment 1a

Experiment 1b

Fig. 3 Left: Results of Experiment 1a. Right: Results of Experiment 1b. Mean sensitivity is shown as a function of the type of spatial configuration present in the probing phase and scene dynamics. Full configurations increase change detection performance for static and dynamic scenes. Error bars represent the standard error of the mean (*SEM*)

detection benefit of full spatial configurations over no spatial configurations might not be related to the representation of spatial configurations at all. The presence of nonprobed objects during the probing phase might provide spatial anchors and, thus, reduce the spatial uncertainty for the exact location of the probed object on the display. If the present results were caused by spatial anchors only, we would expect a subset of visible nonprobed objects—a partial spatial configuration—to provide some spatial anchors too. These spatial anchors should result in higher change detection performance than without a spatial configuration. Because performance was not higher with a partial spatial configuration than without a spatial configuration, we can rule out the anchoring hypothesis for static and dynamic scenes. Instead, there is evidence that the global spatial configuration of the objects is represented in VSTM.

In addition to the effects of spatial configurations, we also observed a main effect of scene dynamics on location change detection performance. Location changes were detected less efficiently with dynamic than with static scenes. Since we were particularly interested in the effect of spatial configurations and their interaction with scene dynamics, this main effect of scene dynamics was not of particular interest to the initial research question for this first experiment—namely, whether spatial configurations of dynamic scenes are represented in VSTM. The effect of scene dynamics could be caused either by the type of memory representation for spatial configurations of static and dynamic scenes or by processes related to memory retrieval. Moving objects might be

represented with a higher spatial uncertainty because they change their locations continually over time, or it might be more difficult to match a dynamic probe to the corresponding memory representation due to the spatio-temporal properties of dynamic probes. Furthermore, higher performance with static conditions might also be explained by the task that required participants to bind motion information to locations in dynamic conditions, but not in static conditions. However, the present data do not allow us to make a decision on the above-mentioned possibilities, since the experiment was not explicitly designed for this.

Experiment 1b

Previous research has shown that spatial configurations are obligatorily encoded for static random dot patterns (Jiang et al., 2000). Applying the same method as Jiang et al. (2000), we tested whether these results generalize to our dynamic scenes.

Method

Participants Twenty-four students from the University of Tübingen participated in this experiment in exchange for partial course credit or monetary compensation. All participants reported normal or corrected-to-normal vision.

Procedure and design The procedure was identical to that in the first experiment, except for one modification. The partial spatial configuration condition was replaced by a

distorted spatial configuration condition. In the distorted condition, the probing phase consisted of the probed square and five nonprobed squares. Each of the five nonprobed squares occupied a new position that no object in the memory phase had occupied at any specific point in time (see Fig. 2).

Introducing this change to the experimental design resulted in a 3 (spatial configuration: full, distorted, no) × 2 (scene dynamics: dynamic, static) × 2 (change present: yes, no) × 20 (repetitions) within-subjects design with 240 experimental trials. At the beginning of the experiment, participants performed 24 practice trials.

Results

Change detection performance as measured by d' is depicted in Fig. 3. All trials with response times larger than 5,000 ms were considered invalid and were removed from the data set (35 trials, 0.61% of the data).

We analyzed sensitivity using a 3 (spatial configuration) × 2 (scene dynamics) repeated measures ANOVA. There was a significant main effect for spatial configuration, $F(2, 46) = 36.39$, $p < .001$, $\eta_p^2 = .61$, interacting with scene dynamics, $F(2, 46) = 10.50$, $p < .001$, $\eta_p^2 = .31$. Planned contrasts using paired t-tests revealed that sensitivity was higher with full spatial configurations than with distorted or no spatial configurations for both static and dynamic scenes, all ps ≤ .003. Sensitivity was lower with distorted spatial configurations than without spatial configurations for static scenes, $p < .001$, but not for dynamic scenes, $p = .617$. As in Experiment 1a, there was a significant main effect for scene dynamics, $F(1, 23) = 85.34$, $p < .001$, $\eta_p^2 = .79$, indicating lower change detection performance for dynamic than for static scenes.

Discussion

We included distorted spatial configurations in the probing phase to further discourage participants from encoding the spatial configuration formed by the objects. In the distorted condition, all nonprobed squares changed their positions between the memory and probing phases.

Despite the introduction of distorted spatial configurations into the experimental design, we found a higher location change detection performance with full than with no spatial configurations. This provides further evidence that spatial configurations of dynamic scenes might be obligatorily represented in VSTM. With static scenes, we replicated the findings of Jiang et al. (2000) that static spatial configurations might be obligatorily represented.

As in the study of Jiang et al. (2000), we presented full and distorted spatial configurations with equal probabilities. If distorted configurations were presented more often than full configurations and there still was an advantage of full configurations over no spatial configurations, an even stronger conclusion could be drawn. However, given the strength of the configuration effect in our experiments and previous research, the present results show a strong tendency to represent spatial configurations of dynamic scenes in VSTM.

We found a lower location change detection performance for distorted than for no spatial configurations with static, but not dynamic, scenes. Nonetheless, the experimental manipulation can be considered successful, as is evident from the response times (see the Appendix). The presence of distorted spatial configurations resulted in longer response times with static and dynamic scenes. This indicates that the presence of distorted configurations interfered with the memory representation that contains the global spatial configuration. We expected this interference to discourage participants from encoding the global spatial configuration. Their failure to do so supports the idea that there is a strong tendency to represent spatial configurations of dynamic scenes.

Section 2: Characteristics of the representation of dynamic spatial configurations

So far, we have shown that spatial configurations of dynamic scenes are represented in VSTM. In the experiments in Section 2, we studied the characteristics of their representation. In Experiments 2a, 2b, and 3, we investigated the role of spatiotemporal features. We demonstrated that the representation includes spatiotemporal information, thus showing the representation of *dynamic* spatial configurations in VSTM. In Experiment 4, we studied the representation of stationary and moving objects in mixed scenes. We demonstrated that stationary and moving objects are represented in a single spatial configuration, rather than in separate static and dynamic spatial configurations, in VSTM.

Experiment 2a

In the present experiment, we investigated whether the VSTM representation of dynamic scenes includes the rough (static) spatial layout of the objects or whether the dynamics of the display are also encoded. Participants encoded a dynamic scene and detected a location change of a moving object. Besides the full dynamic spatial configuration and no spatial configuration conditions from the previous experiments, we presented a critical new config-

uration condition in the probing phase of the present experiment: a static spatial configuration. In the static spatial configuration condition, all objects but the probe were static and were located at the original locations that they had occupied at the middle frame of the dynamic memory scene. That is, the static spatial configuration showed the spatial layout but not the dynamics of the scene.

If the VSTM representation of the spatial configuration of dynamic scenes includes the spatial layout of the objects only, performance with static spatial configurations should equal dynamic spatial configurations and result in a higher performance than with no configurations. If the dynamics of the scene are part of the VSTM representation, performance should be higher with dynamic spatial configurations than with static spatial configurations and no spatial configurations.

Method

Participants Eighteen students from the University of Tübingen participated in this experiment in exchange for partial course credit or monetary compensation. All participants reported normal or corrected-to-normal vision.

Procedure and design We used the same procedure as in Experiment 1a, unless otherwise noted. Participants encoded a dynamic scene showing six moving squares during the memory phase. Besides dynamic and no spatial configurations, we presented static spatial configurations during the probing phase (see Fig. 4). In the static condition, the five nonprobed squares were stationary in the probing phase and were shown at the original locations that they had occupied at the middle frame of the memory phase throughout the probing phase.

This resulted in a 3 (spatial configuration: dynamic, static, no) × 2 (change present: yes, no) × 20 (repetitions) within-subjects design with 120 experimental trials. At the beginning of the experiment, participants performed 24 practice trials.

Results

Change detection performance as measured by d' is depicted in Fig. 5. All trials with response times larger than 5,000 ms were considered invalid and were removed from the data set (nine trials, 0.42% of the data).

We analyzed sensitivity using an ANOVA including spatial configuration as a within-subjects factor. The main effect of spatial configuration reached significance, F(2,
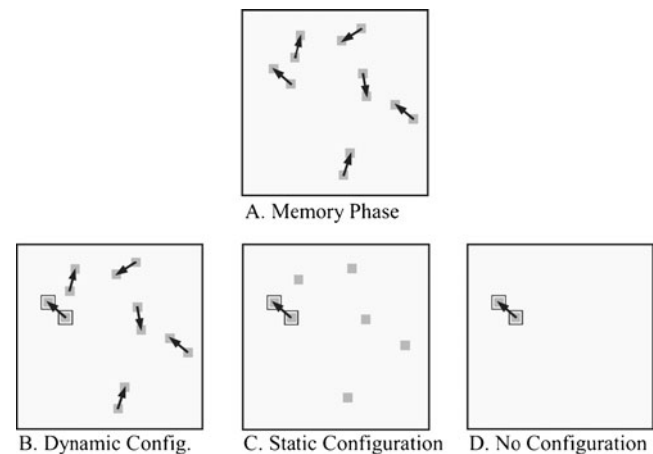

Fig. 4 Types of spatial configurations used during the probing phase in Experiment 2a. A single moving object marked by a red outline in the probing phase (**b–d**) is probed for a location change, as compared with the memory phase (**a**), by showing dynamic (**b**), static (**c**), or no (**d**) spatial configurations. The probed object was not displaced in this example

34) = 4.66, p = .016, η_p^2 = .22. The effect of spatial configuration was further examined with planned contrasts using paired t-tests. We replicated the higher performance with dynamic configurations, as compared with no spatial configurations, p = .011. The static spatial configurations resulted in a significantly lower change detection performance than did the dynamic spatial configurations, p = .022, and did not differ significantly from no spatial configurations, p = .795.

Discussion

The results of Experiment 2a show that the VSTM representation includes more than the pure spatial layout of the objects. Performance with dynamic spatial configurations was the highest, indicating that the VSTM representation of the spatial configuration of the dynamic scene includes the scene dynamics.

Surprisingly, static spatial configurations were not sufficient to cause a higher change detection performance than that for no spatial configurations. Since the rough spatial layout of the scene is a subset of the dynamic spatial configuration, one could have expected a higher performance with static spatial configurations than with no spatial configurations too. Two reasons can account for this finding. First, the static layout might not have been sufficient to trigger the retrieval of the dynamic VSTM representation. Second, interference within the display of the probing phase between dynamic probe and static reference objects might have affected memory retrieval. We examined this in Experiment 2b.
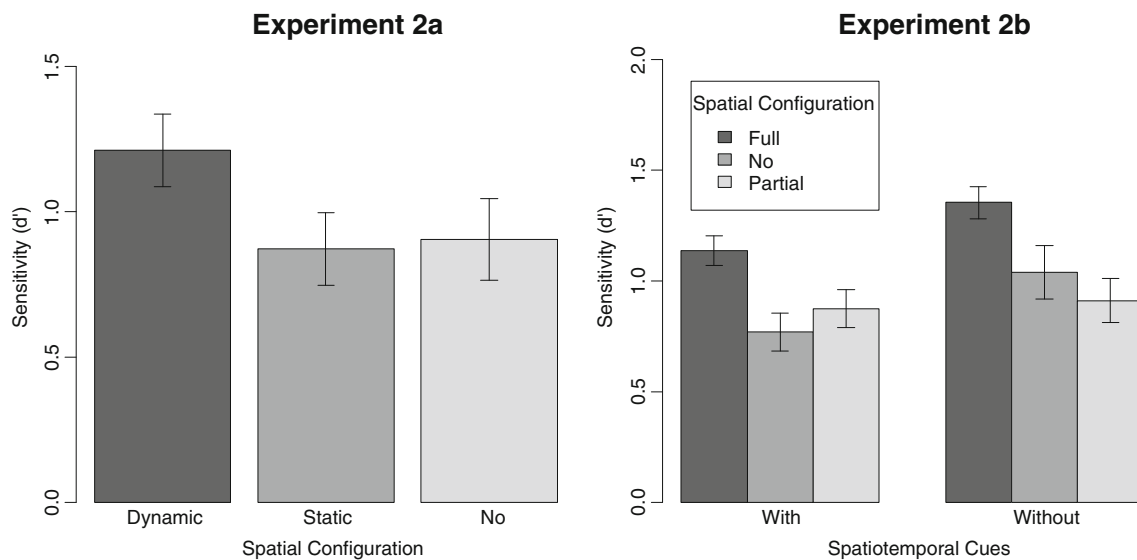
Experiment 2a

Experiment 2b

Fig. 5 Left: Results of Experiment 2a. Location change detection performance for a moving probe object is higher with the dynamic spatial configuration than with a static configuration or no configuration. Right: Results of Experiment 2b. Full spatial configurations resulted in higher location change detection performance than did no or partial spatial configurations. This was true for a dynamic presentation showing a moving probe object with dynamic context (with spatiotemporal cues). The same pattern resulted with a static snapshot of the dynamic scene showing a stationary probe object with static context (without spatiotemporal cues). Error bars represent the *SEM*

Experiment 2b

In the present experiment, we investigated whether configurational information that lacks spatiotemporal information is always insufficient to trigger memory retrieval of dynamic spatial configurations. In particular, we were interested in whether a static snapshot from a dynamic scene in the probing phase is sufficient to show the beneficial effect of full spatial configuration seen in the previous experiments. The critical difference from Experiment 2a is related with the dynamics of the probed object: It is *stationary* if a snapshot of the scene is shown, and it *moves* if the scene is shown dynamically.

If spatiotemporal cues are an integral part of memory retrieval of the representation of dynamic spatial configurations, we should see an interaction—namely, a beneficial effect of full spatial configurations with the dynamic scene, but not with a static snapshot from the scene in the probing phase. If spatial information alone can be sufficient to trigger memory retrieval, we should see a beneficial effect of full spatial configurations for the dynamic scene, as well as for a static snapshot from the scene.

Method

Participants Thirty-two students from the University of Tübingen participated in this experiment in exchange for partial course credit or monetary compensation. All participants reported normal or corrected-to-normal vision.

Procedure and design We used the same procedure as in Experiment 1a, if not otherwise noted. Participants encoded a dynamic scene showing six moving squares during the memory phase. Contrary to the previous experiments, scene dynamics of the memory and probing phases did not necessarily match any more. Instead, we varied the availability of spatiotemporal cues in the probing phase by showing either all frames of the dynamic scene (with spatiotemporal cues) or a static snapshot from the middle of the scene (without spatiotemporal cues) for 400 ms. Participants were instructed to detect a location change for the highlighted square by answering the question of whether the probed square was located on the path of movement of any moving square from the memory phase, no matter whether the probed square moved or not. In order to encourage the encoding of the scene motion, the probing phase showed the dynamic scene with all frames on two thirds of the trials, while there was a static snapshot from the middle of the scene on one third of the trials.

This resulted in a 3 (spatial configuration: full, partial, no) × 2 (spatiotemporal cues: with, without) × 2 (change present: yes, no) within-subjects design. Each condition was repeated 15 times without spatiotemporal cues and 30 times with spatiotemporal cues in the probing phase, resulting in 270 experimental trials per participant. At the

beginning of the experiment, participants performed 18 practice trials.

Results

Change detection performance as measured by d' is depicted in Fig. 5. All trials with response times larger than 5,000 ms were considered invalid and were removed from the data set (28 trials, 0.32% of the data). One participant was excluded from the analysis due to chance-level performance.

We analyzed sensitivity using a 3 (spatial configuration) × 2 (spatiotemporal cues) repeated measures ANOVA. The interaction of spatial configuration and spatiotemporal cues on sensitivity was clearly not significant, $F(2, 60) = 1.10$, $p = .339$, $\eta_p^2 = .04$, indicating that the presence of spatial configurations had the same effect on location change detection performance no matter whether spatiotemporal cues were available in the probing phase or not. The significant main effect of spatial configuration, $F(2, 60) = 14.20$, $p < .001$, $\eta_p^2 = .32$, was caused by a higher change detection performance with full spatial configurations than with partial or no spatial configurations, both $ps < .001$, and partial and no spatial configurations not differing significantly from one another, $p = .896$, as revealed by planned contrasts using paired t-tests. There was also a significant main effect of spatiotemporal cues, $F(1, 30) = 6.48$, $p = .016$, $\eta_p^2 = .18$, indicating lower change detection performance with spatiotemporal cues than without in the probing phase.

Discussion

Full spatial configurations provided a change detection benefit if spatiotemporal cues were present in the probing phase, replicating our previous results. More important, a higher performance with full spatial configurations was also observed with a static snapshot from the middle of the scene that contained only spatial but no spatiotemporal cues. Spatial cues can, therefore, be sufficient during memory retrieval for matching the memory representation of dynamic spatial configurations on the visual input during the probing phase. This provides further evidence that the mechanisms underlying the representation of static and dynamic spatial configurations in VSTM might be similar, if not the same.

The effects of spatial configuration elicited with and without spatiotemporal cues in the probing phase were comparable, as indicated by the clearly nonsignificant interaction effect between spatial configuration and spatiotemporal cues on location change detection performance. Therefore, one might argue that the VSTM representation

of dynamic spatial configurations contains only spatial but no spatiotemporal features. However, our results from Experiment 2a preclude such an argument. In addition, we further investigated the role of spatiotemporal features in Experiment 3. Anticipating the results, we will provide further evidence that detailed spatiotemporal features of the dynamic spatial configuration are represented.

There was a main effect of spatiotemporal cues on sensitivity, showing higher location change detection performance without spatiotemporal cues than with them. This is true even though participants always encoded a dynamic scene and, thus, based their responses on the same memory representation. These results support our claim made in the previous experiments that the main effect of scene dynamics on sensitivity is not at all conclusive with respect to the question of whether moving object locations are less accurately represented in VSTM than are static object locations. In contrast, scene dynamics during memory retrieval itself account for part of the scene dynamics effect.

Experiment 3

Results from Experiment 2a showed that the VSTM representation of dynamic spatial configurations includes spatiotemporal information, because only dynamic configurations, but not the static spatial layout, improved change detection performance for a moving probe object. When the probed object itself was also stationary in Experiment 2b, the spatial layout did cause a configuration effect. Since the reference objects were static in Experiment 2a, it remains unclear how detailed the spatiotemporal representation of dynamic spatial configurations is. In the present experiment, we tackled this question by designing dynamic configurations in such a way that their spatial information equaled the memory scene as closely as possible and their spatiotemporal information differed from the memory scene.

We derived two variants from the dynamic memory scene to test for two plausible ways in which the spatial information could be preserved with differing spatiotemporal features. First, the spatial locations might be averaged over time, creating a mean spatial configuration formed by the mean object locations. A partial orthogonal spatial configuration was derived from the memory scene by having part of the nonprobed objects move orthogonally to their original motion path through their original mean location. Such a partial orthogonal configuration creates the same mean spatial configuration as the memory scene but differs in its spatiotemporal features. Second, the spatial locations might be summarized over time. In this case, the spatial paths of motion would be represented without information about the motion over time. A partial reverse spatial configuration was derived from the memory scene

by having part of the nonprobed objects move along their reverse motion paths. Such a partial reverse configuration creates the same spatial paths of motion as the memory scene but differs in its spatiotemporal features.

If we observed a configurational effect with preserved spatiotemporal features, but not with these new variants with differing spatiotemporal features, this would provide further evidence that the representation of dynamic spatial configurations includes detailed information about the dynamics of the display.

Method

Participants Twenty-eight students from the University of Tübingen participated in this experiment in exchange for partial course credit or monetary compensation. All participants reported normal or corrected-to-normal vision. Some students had previously participated in Experiment 1b but were naive as to the purpose of the experiment.

Procedure and design We used the same procedure as in Experiment 1a, unless otherwise noted. Participants encoded a dynamic scene showing six moving squares during the memory phase. Besides full and no spatial configurations, we presented partial reverse and partial orthogonal spatial configurations during the probing phase (see Fig. 6). In both partial conditions, there were six squares present. Three of the nonprobed squares were manipulated. With partial reverse spatial configurations, those three squares started at their original ending position from the memory phase and moved in the reverse direction to their original starting position. With partial orthogonal spatial configurations, those three squares moved orthogonally to their original trajectory through the same mean location at the middle of the scene. One of the two possible movement directions was randomly chosen for each of the three orthogonal objects individually.

This resulted in a 4 (spatial configuration: full, partial reverse, partial orthogonal, no) × 2 (change present: yes, no) × 15 (repetitions) within-subjects design with 120 experimental trials. At the beginning of the experiment, participants performed 16 practice trials.

Results

Change detection performance as measured by d' is depicted in Fig. 7. All trials with response times larger than 5,000 ms were considered invalid and were removed from the data set (six trials, 0.22% of the data). Five participants were excluded from the analysis due to chance-level performance.
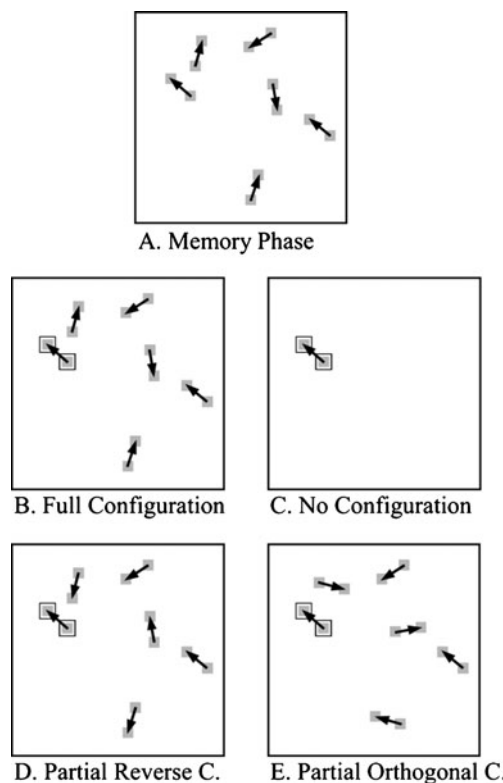

Fig. 6 Types of spatial configurations used during the probing phase in Experiment 3. A single object marked by a red outline in the probing phase (**b**–**e**) is probed for a location change, as compared with the memory phase (**a**), by showing full (**b**), no (**c**), partial reverse (**d**), or partial orthogonal (**e**) spatial configurations. The probed object was not displaced in this example

We analyzed sensitivity using an ANOVA including spatial configuration as a within-subjects factor. The main effect of spatial configuration reached significance, $F(3, 66) =$
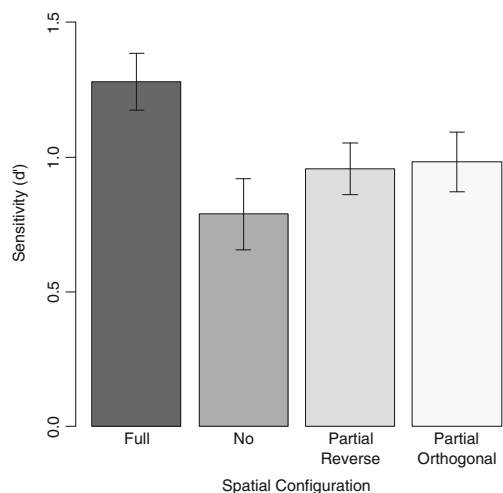

Fig. 7 Results of Experiment 3: Mean location change detection performance as a function of spatial configuration. In the partial reverse and partial orthogonal conditions, the full spatial configuration was present with three nonprobed squares moving in reverse or orthogonal directions, respectively. Error bars represent the *SEM*

3.86, $p = .013$, $\eta_p^2 = .15$. The effect of spatial configuration was further examined with planned contrasts using paired t-tests. We replicated the higher performance with full, as compared with no, spatial configurations, $p = .006$. The partial reverse and partial orthogonal spatial configurations resulted in a significantly lower change detection performance than did full spatial configurations, both $ps \leq .045$, indicating that detailed spatiotemporal features of dynamic spatial configurations are represented in VSTM. In addition, both partial conditions did not differ significantly from one another and the no spatial configuration condition, all $ps \geq .259$.

Discussion

Two new spatial configuration conditions, a partial orthogonal and partial reverse condition, were derived from the dynamic memory scene to contain the same averaged or summarized spatial features as the memory scene, respectively, but to differ in their spatiotemporal features from the memory scene. If the VSTM representation of dynamic spatial configurations does not include detailed spatiotemporal information, we would expect at least one of the two partial configuration conditions to cause the same location change detection performance as full spatial configurations. If the VSTM representation includes detailed spatiotemporal information, we would expect location change detection performance in both partial conditions to be lower than with full spatial configurations.

Our results are in line with the detailed representation of spatiotemporal features of dynamic spatial configurations in VSTM. The partial reverse and partial orthogonal conditions caused location change detection performances that were lower than those with full spatial configurations. Actually, unlike the middle frame used in Experiment 2b but similar to the static configuration in Experiment 2a, the partial reverse and partial orthogonal conditions did not provide a significantly higher performance than the no spatial configurations. Therefore, memory retrieval of dynamic spatial configurations is possible with either missing spatiotemporal cues (Experiment 2b) or spatiotemporal cues conforming to the representation (full dynamic condition in all the experiments). Presenting conflicting spatiotemporal cues, however, interferes with memory retrieval (static configuration in Experiment 2a and partial conditions in the present experiment).

Experiment 4

In the previous experiments, we used memory scenes with either moving or stationary objects only. To broaden our understanding of the mechanisms underlying the formation and representation of static and dynamic spatial configurations, we used mixed scenes containing stationary and moving objects in the present experiment. In a mixed scene, one can think of at least three reasonable spatial configurations: the stable and static spatial configurations formed by the stationary objects alone, the highly dynamic spatial configuration formed by the moving objects alone, and a single spatial configuration including both stationary and moving objects.

We were interested in whether stationary and moving objects in a mixed scene are represented in separate spatial configurations according to their dynamics or whether they are represented in a single spatial configuration including all objects. If stationary and moving objects are represented in separate spatial configurations, location change detection performance for an individual object should benefit from the presence of a partial spatial configuration containing all objects corresponding in dynamics to the probed object during the probing phase. Change detection performance with such a partial spatial configuration should, therefore, be higher than that without a spatial configuration. If they are represented in a single spatial configuration, the presence of such a partial spatial configuration containing all objects corresponding in dynamics to the probed object should not produce any other effects than did the partial spatial configurations in Experiment 1a, thus eliciting a location change detection performance not higher than that without a spatial configuration.

Even if stationary and moving objects are represented in a single spatial configuration, stationary and moving objects might form distinct perceptual groups interfering with the formation of such a configuration by effects of perceptual grouping (Jiang et al., 2004). Such effects should result in a lower mean performance with the full configuration conditions in the present experiment than with the full configuration conditions in Experiment 1a, where we presented only one perceptual group at a time.

Method

Participants Twenty students from the University of Tübingen participated in this experiment in exchange for partial course credit or monetary compensation. All participants reported normal or corrected-to-normal vision.

Procedure and design We used the same procedure as in Experiment 1a, unless otherwise noted. Instead of manipulating the dynamics of the whole scene, we manipulated the dynamics of the probed square. Participants saw mixed scenes containing three moving and three stationary squares in the memory phase. During the probing phase, either a stationary or a moving square was probed for a location change by being highlighted with a red frame, as in the previous experiments. On half of the trials, the probed square was stationary, and on the other half, it was moving.

Furthermore, we split the partial spatial configuration condition in Experiment 1a into two new conditions: a partial dynamic spatial configuration showing all remaining moving squares and a partial static spatial configuration showing all remaining stationary squares during the probing phase.

Note that we did not want participants to know during the memory phase whether the probed square would be stationary or moving. Therefore, we used equal numbers of stationary and moving squares in these scenes. This resulted in unequal total numbers of squares visible in the partial static and partial dynamic conditions during the probing phase. Three squares were present when the dynamics of the probed square and partial condition matched (e.g., the probed moving square and two additional moving squares) and four squares otherwise (e.g., the probed moving square and three additional stationary squares). This difference in total object numbers between the partial static and partial dynamic conditions did not present a problem for the experimental design. First, the main question of the present experiment addressed the comparison of performance with partial spatial configurations congruent with probe dynamics and, thus, always containing three squares with full and no spatial configurations within each probe dynamics condition. And second, we showed in Experiment 1a that pure spatial anchoring not showing the full spatial configuration does not improve change detection performance. In that sense, the partial spatial configurations incongruent with probe dynamics served as an even stronger anchoring control condition than did the partial spatial configurations in Experiment 1a, since they contained even four out of six squares in the present experiment. We expected that performance would not be higher than that without spatial configurations, thus excluding a pure spatial anchoring hypothesis for the present experiment too.

These manipulations resulted in a 4 (spatial configuration: full, partial dynamic, partial static, no) × 2 (probe dynamics: moving square, stationary square) × 2 (change present: yes, no) × 20 (repetitions) within-subjects design with 320 experimental trials. At the beginning of the experiment, participants performed 32 practice trials.

Results

Change detection performance as measured by d' is depicted in Fig. 8. All trials with response times larger than 5,000 ms were considered invalid and were removed from the data set (23 trials, 0.36% of the data).

We analyzed sensitivity using a 4 (spatial configuration) × 2 (probe dynamics) repeated measures ANOVA. There was a significant main effect of spatial configuration, $F(3, 57) =$
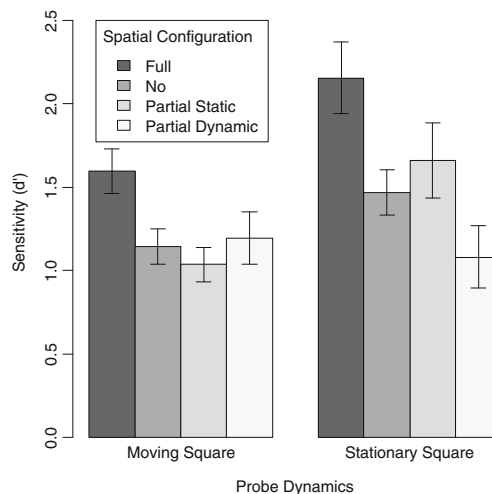

Fig. 8 Results of Experiment 4: Mean location change detection performance as a function of the type of spatial configuration present in the probing phase and the dynamics of the probed object. Error bars represent the *SEM*

13.84, $p < .001$, $\eta_p^2 = .42$, interacting with probe dynamics, $F(3, 57) = 4.42$, $p = .007$, $\eta_p^2 = .19$. Planned contrasts using paired *t*-tests revealed that sensitivity was higher with full spatial configurations than with partial dynamic, partial static, and no spatial configurations for stationary and moving probes, all ps $\leq .025$. There was no significant difference in performance between partial static and no spatial configurations for stationary and moving probes, both ps $\geq .194$. While performance with partial dynamic configurations did not differ significantly from that with partial static and no configurations for moving probes, both ps $\geq .268$, performance was lower with partial dynamic configurations than with partial static and no configurations for stationary probes, both ps $\leq .010$. Interestingly, the partial dynamic condition showing lowest performance with stationary probes contained all but two stationary squares. There was also a significant main effect of probe dynamics, $F(1, 19) = 11.38$, $p = .003$, $\eta_p^2 = .37$, indicating a higher location change detection performance for stationary than for moving probes.

To analyze the effects of perceptual grouping, we calculated a between-experiments comparison of the sensitivity in the full configuration conditions in Experiment 1a and the full configuration conditions in the present experiment. To account for the performance differences between static and dynamic stimuli, we calculated the mean performance of the static and dynamic full configuration conditions for each participant. Sensitivity for the full configuration conditions ($M = 1.83$, $SE = 0.12$) of Experiment 1a did not differ significantly from that for the full configuration conditions of the mixed scenes ($M = 1.88$, $SE = 0.15$) in the present experiment, $t(38) = -0.22$, $p = .827$, indicating that there was

no interference due to perceptual grouping in the present experiment.

Discussion

Participants encoded mixed scenes containing stationary and moving squares. If static and dynamic objects are represented in separate spatial configurations, the presence of a partial spatial configuration containing all squares congruent in dynamics with the probed square should result in a higher change detection performance than should no spatial configuration. If they are represented in a single configuration, performance should be no higher with such partial spatial configurations than without spatial configurations, because pure anchoring does not improve change detection performance, as is shown in Experiment 1a.

Change detection performance with partial spatial configurations congruent with probe dynamics was not higher than that without spatial configurations for both stationary and moving probes. The results are therefore in favor of the representation of stationary and moving objects in a single spatial configuration. This conclusion is supported by the finding that full spatial configurations containing stationary and moving objects did, indeed, result in a higher change detection performance than did no spatial configurations, thus ruling out the alternative explanation that spatial configurations might not be encoded in mixed scenes at all.

Interestingly, the formation of the single spatial configuration did not show interference effects due to perceptual grouping. Performance with full configurations showing stationary and moving objects simultaneously in the present experiment did not differ from that in the full configuration conditions in Experiment 1a, where either stationary or moving objects were shown. Therefore, it seems that spatiotemporal information does not serve as perceptual grouping cues as line segments did (Jiang et al., 2004). This might be due to the fact that line segments create a perceptual grouping instantly, while spatiotemporal information includes the temporal dimension; that is, it takes some time to divide the visual input into separate groups on the basis of spatiotemporal information.

The present results replicated the findings of Experiment 1a that the pure presence of more spatial anchors during the probing phase cannot explain the effect of higher change detection performance with full configurations, as compared with no spatial configurations. For all partial spatial configuration conditions presented in the experiment, we observed change detection performances that were not higher than those without spatial configurations.

When participants tried to detect a location change of a stationary probe with partial dynamic spatial configurations, they showed a surprisingly low change detection performance that was even below the change detection performance of stationary probes without spatial configurations. One might argue that the moving squares around the stationary probed square captured attention by their motion onset (Abrams & Christ, 2003) and reduced the resources available for the location change detection task, thus reducing performance, as compared with the partial static and no conditions. However, these moving squares were also present with full spatial configurations where change detection performance was highest, thus precluding a simple attentional capture explanation. An alternative and plausible explanation for the low performance is that the moving objects increased the spatial uncertainty of the stationary probed object's location by the process of induced motion (e.g., Duncker, 1929). This, in turn, resulted in a reduced location change detection performance for the probed object. If the full spatial configuration is present, it serves as reference for the spatial location of the probed object, thus increasing location change detection performance.

During the encoding of the mixed scenes, participants never knew whether a stationary or moving object would be probed in the probing phase. Nonetheless, location changes of moving probes were detected less accurately than location changes of stationary probes. This indicates that locations of moving objects might be less accurately represented in VSTM than locations of stationary objects.

General discussion

Previous research demonstrated that locations of multiple static objects are represented on the basis of their global spatial configuration in VSTM (e.g., Jiang et al., 2004; Jiang et al., 2000). Once objects move, they form a dynamic spatial configuration that changes continually over time. To date, there has been no research on the representation of such dynamic spatial configurations in VSTM. There is some evidence from the literature on attentional processes within the multiple-object tracking paradigm suggesting that dynamic spatial configurations in dynamic scenes are used to guide visual attention (e.g., Fehd & Seiffert, 2008; Huff et al., 2009; Yantis, 1992). With the present set of experiments, we tested the hypothesis that dynamic spatial configurations are represented in VSTM, and we characterized some aspects of this representation.

The main finding of our experiments is that dynamic spatial configurations are represented in VSTM including detailed spatiotemporal information. Across all of our experiments, we observed a location change detection benefit when the full spatial configuration was present in the probing phase. Global spatial configurations, therefore, not only are an important feature of VSTM representations

of static objects that are presented simultaneously (Jiang et al., 2004; Jiang et al., 2000; see also the static scene conditions in our Experiments 1a and 1b) or sequentially (Blalock & Clegg, 2010), but also are essential for the representation of locations of dynamic objects with varying spatial inter-object relations over time. Besides demonstrating a spatiotemporal representation in VSTM, this provides further evidence for the enormous role of global spatial configurations for VSTM representations. Using distorted spatial configurations in the probing phase, we tried to discourage the encoding of spatial configurations in the first place. The pattern of results—namely, a performance benefit for full spatial configurations—was not changed, thus replicating the strong tendency to encode static spatial configurations (Jiang et al., 2000) and extending those findings by showing a strong tendency for the encoding of dynamic spatial configurations.

In all our experiments, we ruled out an anchoring hypothesis, since performance with partial spatial configurations never reliably exceeded performance without spatial configurations. Even the presence of four out of six objects in some partial conditions of Experiment 4 did not increase performance. In addition, we demonstrated in a control experiment within our lab that the introduction of articulatory suppression does not change the pattern of results of Experiment 1a. We can therefore be confident that our effects are truly caused by the representation of dynamic spatial configurations in VSTM.

Static and dynamic spatial configurations: one system?

Having shown that dynamic spatial configurations are represented in VSTM raises the question of whether it is one system or two separate subsystems that underlie the representations of static and dynamic spatial configurations. When previous studies on the representation of static and dynamic content in general are considered, both accounts seem plausible. A developmental dissociation for static and dynamic information suggests a separation of visual memory into a static and a dynamic subsystem (Pickering et al., 2001). A separation of spatiotemporal and location information is also supported by studies on observed actions (Wood, 2007) and the architecture of visual working memory (Wood, 2011) using a dual-task paradigm. Those results show the existence of three separate buffers specialized on spatiotemporal, feature, and view-dependent snapshot information. While dynamic configurations might be stored in the spatiotemporal component, static configurations might be stored in a snapshot-like form. However, some mental representations might be dynamic in nature by incorporating a temporal dimension (Freyd, 1987). From this perspective, static representations are special cases of dynamic repre-

sentations, which, in turn, would suggest representation in the same system.

There is accumulating evidence across our experiments that suggest the existence of one system for the representation of static and dynamic spatial configurations. We demonstrated that stationary and moving objects in a mixed scene are combined into a single spatial configuration instead of two separate spatial configurations according to the objects' motion. This precludes a strong separation hypothesis predicting the separate representation of static and dynamic objects and, therefore, the formation of separate spatial configurations. Certainly, a weak separation hypothesis could still account for our results in Experiment 4 by stating that the separation is based not on the dynamics of the individual objects forming the spatial configuration but on the presence of varying spatial inter-object relations over time. From this perspective, we studied the dynamic subsystem with our mixed scenes, because the presence of some moving objects caused varying spatial inter-object relations over time, thus forming a dynamic spatial configuration. However, we also demonstrated that the memory representation of dynamic spatial configurations can be accessed with missing spatiotemporal cues during memory retrieval in the probing phase. Therefore, such a dynamic subsystem can handle missing spatiotemporal cues. To assume an additional static subsystem doing exactly the same thing would suggest that VSTM is organized inefficiently, which seems unlikely. Therefore, we propose the existence of one system responsible for the representation of static and dynamic spatial configurations. An interesting consequence of this proposal is the implication that the role of spatial configurations studied in static contexts is a special case of the role of spatial configurations in dynamic contexts.

This one-system view is also in line with other results from our experiments. Since static and dynamic spatial configurations are represented in one system, it is plausible that we observed a comparable effect of spatial configurations for static and dynamic scenes. Furthermore, there was evidence that both static and dynamic spatial configurations are encoded even if discouraged by the presence of distorted configurations, and we found for both configurations that the global spatial layout is encoded in VSTM; that is, partial spatial configurations were never sufficient to improve location change detection performance.

Even though we observed the above-mentioned similarities between static and dynamic configurations, there was also an important difference in scene dynamics observed across the experiments. Mean location change detection performance for the probed objects was lower for dynamic than for static scenes. Since the effect of spatial configurations was not affected by scene dynamics, this difference

is unlikely to have been caused by a differential representation of the global spatial configuration for static and dynamic scenes. Therefore, it does not preclude our one-system view of the representation of static and dynamic spatial configurations proposed above.

The lower performance with dynamic scenes might rather be explained in terms of the precision of the memory representation, processes of memory retrieval for the individual object locations, and the task requiring the binding of motion to locations for dynamic scenes only. Location changes of moving objects were detected less efficiently than stationary ones even when the participants encoded mixed scenes in which the objects were represented in a single spatial configuration. This suggests that moving objects are represented with a higher spatial uncertainty than stationary ones. However, we also found that memory retrieval of a dynamic representation was more difficult with a dynamic probe than with a static probe in Experiment 2b, where the memory representation was identical. The lower performance with dynamic scenes can, therefore, be explained not only in terms of a lower precision of the representation of dynamic object locations, but also by the effects of memory retrieval. Possibly, the dynamic memory representation is distorted into the future due to processes such as representational momentum (Freyd & Finke, 1984) causing an error prone process of matching the present time of the probing phase to the corresponding time in the memory representation during memory retrieval with the dynamic probes used in our experiments.

There are two limitations to the above discussed one-system view. First, our one-system view is valid for the configurational processing of static and dynamic stimuli. Our results are not necessarily valid for the processing of static and dynamic stimuli in general or other types of spatial and spatiotemporal information. Indeed, there is previous research that has demonstrated the separation of spatial and spatiotemporal information in other settings (Pickering et al., 2001; Smyth et al., 1988; Smyth & Pendleton, 1989, 1990; Wood, 2007, 2011). However, our results could also be reconciled with the findings of Wood (2011) either by assuming that our stimuli are always stored in the spatiotemporal component of VSTM or by assuming that the view-dependent snapshot component is not completely static but that it includes some kind of temporal buffer. Second, without having tested a broad range of different and realistic dynamic stimuli, we do not know whether our results on dynamic spatial configurations generalize to different kinds of dynamic events. However, previous research suggests that simple movements and more complex observed actions compete for limited resources and are, therefore, processed in the same system (see Experiment 3b in Wood, 2011). Although there are no empirical results on the configurational processing of complex dynamic stimuli such as observed actions, this gives a first indication that our results might generalize to other dynamic stimuli.

Implications for research on dynamic scenes

With our demonstration that dynamic spatial configurations are represented in VSTM, we provide additional evidence to a growing body of literature emphasizing that not only spatial features, but also spatiotemporal features are important for visual information processing. For example, spatiotemporal features cause a memory distortion through representational momentum (Freyd & Finke, 1984), are used to establish object correspondence across displacements (e.g., Kahneman, Treisman, & Gibbs, 1992; Mitroff & Alvarez, 2007; see also Hollingworth & Franconeri, 2009), are stored in a short-lived visual buffer (Smith, Mollon, Bhardwaj, & Smithson, 2011), are part of the long-term memory representation of dynamic scenes (Matthews, Benjamin, & Osborne, 2007), predict the segmentation of ongoing activity into events (Zacks, 2004), and are utilized when three or fewer objects are tracked with visual attention (Fencsik, Klieger, & Horowitz, 2007; Iordanescu, Grabowecky, & Suzuki, 2009), but not when four objects are tracked (Fencsik et al., 2007; Horowitz, Birnkrant, Fencsik, Tran, & Wolfe, 2006; Keane & Pylyshyn, 2006). The importance of spatiotemporal features is not very surprising, considering that almost any object in the environment is capable of location changes through internal or external forces.

The concept of dynamic spatial configurations in VSTM established with our experiments not only adds new information to the existing body of research, but also might prove useful as a parsimonious explanation of current findings. A recent study examined the relationship between VSTM and object files (Hollingworth & Rasmussen, 2010) by studying the binding of object features to locations and the updating of these bindings through object motion. Participants saw a layout of stationary empty squares that were briefly colored. After a short interval, they moved to new locations, such that they formed the same original layout as the start locations, with the objects swapping locations within the layout and creating an updated layout. Participants had to detect a color change with the colors shown in the original layout, updated layout, or a random layout matching neither one. Since performance was better with the updated layout than with the random layout, feature–location bindings were updated through object motion. However, performance was even better with the original layout than with the updated layout. On the basis of these results, the existence of two systems of object–position bindings was proposed: first, an object-file system

that is sensitive to object motion with a low precision of bound feature representations, and second, a system that is insensitive to object motion representing positions in a static spatial configuration and binding object features to locations with a higher precision. The latter is consistent with the findings of Jiang et al. (2000). Since our results suggest that static spatial configurations are a special case of dynamic spatial configurations, there might be an alternative explanation for this two-systems approach that is more parsimonious. This approach is based on the second system of Hollingworth and Rasmussen. However, we propose that this system might actually be organized spatiotemporally, such that object–position bindings are based on dynamic spatial configurations in VSTM. The two-systems approach and dynamic spatial configurations approach can be tested against one another in future research by explicitly varying the availability of nonprobed squares in the probing phase. The dynamic spatial configurations approach predicts a drop of color change detection performance with the original and updated layouts if some nonprobed objects are removed, thus destroying the spatial configuration in the probing phase. According to the two-systems approach, performance with the original layout, but not the updated layout, should decrease, since only the system responsible for the higher performance with the original layout is based on a static spatial configuration.

When tracking multiple objects with visual attention, observers spontaneously adopt a grouping strategy (Yantis, 1992); that is, they track the group of target objects, rather than the targets individually. The same was true for the memory of individual object locations in our present experiments; that is, observers memorized their dynamic spatial configuration. This could indicate that target grouping and the formation of dynamic spatial configurations in VSTM are based on the same mechanisms and could provide further evidence for the role of VSTM in multiple-object tracking (Makovski & Jiang, 2009). Therefore, research on attentive tracking and the memory of dynamic spatial configurations might benefit from one another. On the one hand, preloading VSTM with a dynamic spatial configuration might reduce tracking performance for grouped but not ungrouped target motion. This would provide further evidence for the role of VSTM in multiple-object tracking and the processes employed during tracking. On the other hand, previous findings on target grouping might hold true for dynamic spatial configurations. In particular, it has been shown that tracking performance decreases when target motion causes a collapse of the higher order object formed by the target objects. It would be interesting to see whether this result generalizes to the representation of dynamic spatial configurations, since it would give information on the restrictions of their formation.

With the present set of experiments, we studied short dynamic scenes of 400 ms to link the findings to previous research on VSTM. In the daily environment, one faces dynamic scenes far longer in duration than the scenes we tested in the present experiments. When watching longer dynamic scenes, observers usually segment them into meaningful events (see Zacks, Speer, Swallow, Braver, & Reynolds, 2007, for a review), and this segmentation behavior has a distinct influence on memory. Memory for an object is better if it occurred in the current rather than the previous event, with time from perception to test being controlled (Swallow, Zacks, & Abrams, 2009). This effect is caused by the updating of memory contents whenever a new event starts. Movement features of objects have been shown to have a strong influence on the perception of distinct events (Zacks, 2004). If objects do not move on distinct linear trajectories, as in our experiments, but change their directions over time, this should elicit the perception of distinct events, and an updating of memory contents across these events should occur. Therefore, when longer dynamic scenes with changes in movement directions are watched, the benefit of dynamic spatial configurations reported in the present experiments might account only for recognition tests probing contents from currently and not previously perceived events. Such results would give further information on the persistence and formation of dynamic spatial configurations in VSTM. Certainly, one should also consider the possibility that the perception of events could, in part, be the result of the processing of dynamic spatial configurations. Possibly, the change of movement directions causes the need for the formation of a new dynamic spatial configuration, which, in turn, causes an updating of memory contents and the perception of an event.

Conclusion

When objects move in dynamic scenes, they form a spatial configuration that changes continually over time. We demonstrated that the dynamic spatial configuration of 400-ms clips showing six moving squares are represented in VSTM, including their spatiotemporal features. Stationary and moving objects in mixed scenes were represented in a single spatial configuration. We propose that static and dynamic spatial configurations can be represented within the same system and that static spatial configurations were a special case of dynamic spatial configurations in our experiments.

Appendix

Table 1 Hit rates, false alarm rates, proportions correct, and response times for hits across all experiments (standard errors in parentheses)

	Spatial Configuration	Hit Rate	False Alarm Rate	Proportion Correct	Response Times (ms)
Experiment 1a					
Dynamic scene	Full	.79 (.03)	.36 (.03)	.71 (.02)	1,073 (56)
	No	.56 (.03)	.25 (.03)	.65 (.01)	1,135 (84)
	Partial	.65 (.03)	.35 (.03)	.65 (.02)	1,132 (69)
Static scene	Full	.92 (.02)	.23 (.03)	.85 (.02)	927 (55)
	No	.80 (.03)	.27 (.03)	.77 (.02)	977 (64)
	Partial	.75 (.02)	.23 (.03)	.76 (.02)	1,031 (60)
Experiment 1b					
Dynamic scene	Full	.81 (.02)	.46 (.03)	.68 (.02)	1,110 (49)
	No	.52 (.04)	.28 (.03)	.62 (.02)	1,122 (51)
	Distorted	.52 (.03)	.30 (.04)	.61 (.02)	1,250 (67)
Static scene	Full	.91 (.02)	.31 (.03)	.81 (.02)	934 (38)
	No	.73 (.03)	.28 (.03)	.73 (.02)	1,029 (43)
	Distorted	.56 (.03)	.28 (.03)	.64 (.02)	1,217 (74)
Experiment 2a					
	Dynamic	.85 (.03)	.50 (.04)	.68 (.02)	1,151 (65)
	Static	.59 (.04)	.28 (.02)	.66 (.02)	1,215 (76)
	No	.63 (.03)	.31 (.04)	.66 (.02)	1,242 (73)
Experiment 2b					
With spatiotemporal cues	Full	.79 (.02)	.40 (.02)	.69 (.01)	1,122 (39)
	No	.51 (.03)	.25 (.02)	.63 (.02)	1,138 (53)
	Partial	.64 (.03)	.32 (.02)	.66 (.01)	1,203 (53)
Without spatiotemporal cues	Full	.80 (.02)	.36 (.03)	.72 (.01)	1,111 (53)
	No	.70 (.04)	.36 (.03)	.67 (.02)	1,081 (47)
	Partial	.69 (.03)	.37 (.03)	.66 (.02)	1,188 (54)
Experiment 3					
	Full	.82 (.02)	.39 (.03)	.72 (.02)	992 (53)
	No	.57 (.04)	.30 (.03)	.64 (.02)	1,040 (52)
	Partial reverse	.67 (.03)	.33 (.03)	.67 (.02)	1,068 (59)
	Partial orthogonal	.68 (.03)	.33 (.02)	.67 (.02)	1,008 (49)
Experiment 4					
Moving square	Full	.88 (.02)	.39 (.04)	.74 (.02)	1,146 (51)
	No	.58 (.03)	.20 (.02)	.69 (.02)	1,173 (60)
	Partial static	.63 (.04)	.26 (.03)	.68 (.02)	1,183 (67)
	Partial dynamic	.70 (.04)	.29 (.03)	.70 (.03)	1,171 (48)
Stationary square	Full	.88 (.03)	.23 (.04)	.82 (.03)	1,153 (44)
	No	.73 (.04)	.25 (.02)	.74 (.02)	1,140 (51)
	Partial static	.72 (.04)	.21 (.04)	.75 (.03)	1,110 (52)
	Partial dynamic	.67 (.04)	.32 (.05)	.68 (.03)	1,237 (56)

References

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science, 14,* 427–432. doi:10.1111/1467-9280.01458

Blalock, L. D., & Clegg, B. A. (2010). Encoding and representation of simultaneous and sequential arrays in visuospatial working memory. *Quarterly Journal of Experimental Psychology, 63,* 856–862. doi:10.1080/17470211003690680

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–114. doi:10.1017/S0140525X01003922

Duncker, K. (1929). Über induzierte Bewegung [On induced motion]. *Psychologische Forschung, 12,* 180–259. doi:10.1007/BF02409210

Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition, 108,* 201–209. doi:10.1016/j.cognition.2007.11.008

Fencsik, D. E., Klieger, S. B., & Horowitz, T. S. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception & Psychophysics, 69,* 567–577. doi:10.3758/BF03193914

Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics, 33,* 575–581. doi:10.3758/BF03202940

Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review, 94,* 427–438. doi:10.1037/0033-295X.94.4.427

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 126–132. doi:10.1037/0278-7393.10.1.126

Hollingworth, A. (2007). Object–position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 31–47. doi:10.1037/0096-1523.33.1.31

Hollingworth, A., & Franconeri, S. L. (2009). Object correspondence across brief occlusions is established on the basis of both spatiotemporal and surface feature cues. *Cognition, 113,* 150–166. doi:10.1016/j.cognition.2009.08.004

Hollingworth, A., & Rasmussen, I. P. (2010). Binding objects to locations: The relationship between object files and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 543–564. doi:10.1037/a0017836

Horowitz, T. S., Birnkrant, R. S., Fencsik, D. E., Tran, L., & Wolfe, J. M. (2006). How do we track invisible objects? *Psychonomic Bulletin & Review, 13,* 516–523. doi:10.3758/BF03193879

Huff, M., Jahn, G., & Schwan, S. (2009). Tracking multiple objects across abrupt viewpoint changes. *Visual Cognition, 17,* 297–306. doi:10.1080/13506280802061838

Huff, M., Papenmeier, F., Jahn, G., & Hesse, F. W. (2010). Eye movements across viewpoint changes in multiple object tracking. *Visual Cognition, 18,* 1368–1391. doi:10.1080/13506285.2010.495878

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision, 9*(4, Art. 1), 1–12. doi:10.1167/9.4.1

Jiang, Y., Chun, M. M., & Olson, I. R. (2004). Perceptual grouping in change detection. *Perception & Psychophysics, 66,* 446–453. doi:10.3758/BF03194892

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. doi:10.1037/0278-7393.26.3.683

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219. doi:10.1016/0010-0285(92)90007-O

Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cognitive Psychology, 52,* 346–368. doi:10.1016/j.cogpsych.2005.12.001

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. doi:10.1038/36846

Makovski, T., & Jiang, Y. V. (2009). The role of visual working memory in attentive tracking of unique objects. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1687–1697. doi:10.1037/a0016453

Matthews, W. J., Benjamin, C., & Osborne, C. (2007). Memory for moving and static images. *Psychonomic Bulletin & Review, 14,* 989–993. doi:10.3758/BF03194133

Mitroff, S. R., & Alvarez, G. A. (2007). Space and time, not surface features, guide object persistence. *Psychonomic Bulletin & Review, 14,* 1199–1204. doi:10.3758/BF03193113

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16,* 283–290. doi:10.3758/BF03203943

Pickering, S. J., Gathercole, S. E., Hall, M., & Lloyd, S. A. (2001). Development of memory for pattern and path: Further evidence for the fractionation of visual-spatial memory. *Quarterly Journal of Experimental Psychology, 54A,* 397–420. doi:10.1080/02724980042000174

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS ONE, 3,* e1699. doi:10.1371/journal.pone.0001699

Smith, W. S., Mollon, J. D., Bhardwaj, R., & Smithson, H. E. (2011). Is there brief temporal buffering of successive visual inputs? *Quarterly Journal of Experimental Psychology, 64,* 767–791. doi:10.1080/17470218.2010.511237

Smyth, M. M., Pearson, N. A., & Pendleton, L. R. (1988). Movement and working memory: Patterns and positions in space. *Quarterly Journal of Experimental Psychology, 40A,* 497–514.

Smyth, M. M., & Pendleton, L. R. (1989). Working memory for movements. *Quarterly Journal of Experimental Psychology, 41A,* 235–250.

Smyth, M. M., & Pendleton, L. R. (1990). Space and movement in working memory. *Quarterly Journal of Experimental Psychology, 42A,* 291–304.

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General, 138,* 236–257. doi:10.1037/a0015631

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131,* 48–64. doi:10.1037/0096-3445.131.1.48

Wood, J. N. (2007). Visual working memory for observed actions. *Journal of Experimental Psychology: General, 136,* 639–652. doi:10.1037/0096-3445.136.4.639

Wood, J. N. (2011). A core knowledge architecture of visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 357–381. doi:10.1037/a0021935

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology, 24,* 295–340. doi:10.1016/0010-0285(92)90010-Y

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science, 28,* 979–1008. doi:10.1016/j.cogsci.2004.06.003

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind–brain perspective. *Psychological Bulletin, 133,* 273–293. doi:10.1037/0033-2909.133.2.273

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453,* 233–235. doi:10.1038/nature06860