

Do objects in working memory compete with objects in perception?

Hang Zhang, Yuming Xuan, and Xiaolan Fu

*State Key Laboratory of Brain and Cognitive Science, Institute of Psychology,
Chinese Academy of Sciences, Beijing, China*

Zenon W. Pylyshyn

*Rutgers Centre for Cognitive Science, Rutgers University, New Brunswick,
NJ, USA*

It is generally assumed that “perceptual object” is the basic unit for processing visual information and that only a small number of objects can be either perceptually selected or encoded in working memory (WM) at one time. This raises the question whether the same resource is used when objects are selected and tracked as when they are held in WM. In two experiments, we measured dual-task interference between a memory task and a Multiple Object Tracking task. The WM tasks involve explicit, implicit, or no spatial processing. Our results suggest there is no resource competition between working memory and perceptual selection except when the WM task requires encoding spatial properties.

Keywords: Visual working memory; Object; Multiple object tracking; Dual-task interference.

It has been assumed that object is the basic unit for people to organize the everyday scenes (Kahneman, Treisman, & Gibbs, 1992; Pylyshyn, 1989; Wolfe & Bennett, 1997; Yantis & Johnson, 1990; see Scholl, 2001, for a review). The representation of an object persists when the object moves or changes, which is essential to our understanding of dynamic scenes. For example, to judge whether a car has violated the traffic rules, a traffic

Please address all correspondence to Yuming Xuan, Institute of Psychology, Chinese Academy of Sciences, 4A Datun Road, Chaoyang District, Beijing 100101, China. E-mail: xuanym@psych.ac.cn

This research was supported in part by grants from 973 Program (2006CB303101) and the National Natural Science Foundation of China (90820305, 30500157, and 30600182). We thank Chris Olivers, Geoffrey Woodman, and an anonymous reviewer for their helpful comments on a prior draft of the paper.

policeman has to identify the car as the same object after it moves. On the other hand, only a certain number of objects can be represented at a time. The ability to preserve objects is thus an important constraint on visual cognition. The present study was motivated by an interest in the nature of this ability.

One of the first theories to provide an account of perceptual objecthood is that of Kahneman et al.'s (1992). That theory, called Object File Theory, focuses on the objects whose properties have encoded and entered into working memory (WM). Object files are defined as temporary episodic tokens that may be empty or may contain the current description (e.g., location, orientation, and distance) of real objects around us. They parse the world, mediate the representation of novel objects or multiple identical objects, and serve as vehicles to bind features. The persistence of object files over time and space was demonstrated by a priming methodology. Naming an object was shown to be faster when the same object was viewed earlier at a different position (Kahneman et al., 1992), thus showing that priming tends to remain with a particular object rather than a particular location. More recent studies have shown that priming stays with moving objects even when the interval between preview and naming is as long as 8 s (Noles, Scholl, & Mitroff, 2005) and when several objects are primed and tracked in a Multiple Object Tracking (MOT) paradigm (Haladjian & Pylyshyn, 2008). Our capacity to hold information in object files is limited in terms of the number of objects, as reflected by the capacity of visual WM. The limit appears to be about four objects, no matter whether the objects have one relevant feature (colour), or two (colour and orientation), or even four (colour, orientation, length, and a gap in the line) (Luck & Vogel, 1997). But more recent evidences suggest that Visual WM capacity is determined both by a fixed number of objects and by object complexity (Alvarez & Cavanagh 2004; Xu & Chun, 2006).

Another object-based theory, Pylyshyn's visual index theory is concerned with individuating and keeping track of perceptual objects (Pylyshyn, 1989, 2001, 2007). It assumes that at the preattentive stage of visual perception a small number of objects are individuated and selected. A unique index is assigned to each object up to the maximum number of indexes available (about four) and sticks with the object as the object moves or changes properties so that only the object's individuality is relevant to the function of the index. The theory claims that the index itself does not include description of the real object but can be used to facilitate subsequent access to the object. The multiple object tracking (MOT) paradigm in which subjects must keep track of several target objects that move unpredictably among a set of (identical) nontargets, was introduced as a direct test of this indexing ability (Pylyshyn & Storm, 1988). In the basic procedure of MOT, the subset of objects that have to be tracked (called the targets) are indicated at the start

of each trial by flashing them a few times. Then all objects move randomly for some fixed period of time—typically 5–10 s. When they stop, participants must either indicate which ones were the targets (using a computer mouse) or, in some experiments, must indicate whether a particular object (e.g., one that is flashed at the end of the trial) was a target. Because all the objects have the same appearance during the tracking period, the only way participants can identify targets is by their continuous history or trajectory. Generally only about four objects can be correctly tracked (Pylyshyn, 2001, 2003; Pylyshyn & Storm, 1988; Yantis, 1992).

Cowan (2001) claimed that the number of objects that the focus of attention can cover is about four, whether the objects are in WM or in perception. In commenting on Cowan's target article, Davis (2001) and Rensink (2001) raised the question of whether objects in WM and objects being tracked compete for the same cognitive resources. Although what exactly constitutes a resource is not generally specified, it is assumed that when a limited resource is accessed by one task there is less available to another task that uses the same resource. In the case of a memory resource one might think of a resource as something like the number of slots in a memory bank that get used by certain tasks. In our case this presumably means the number of object-specific items of information that can be stored or the number of available object files.

The question of whether a particular resource is used by two different tasks is typically studied by measuring dual task interference. If performance on one of the tasks is not impaired by the simultaneous execution of a second task, this would suggest that the two tasks do not use the same resource. For example, no decrement in performance is observed when a search task (Alvarez, Horowitz, Arsenio, DiMase, & Wolfe, 2005) or a monitoring task (Leonard & Pylyshyn, 2003) are carried out at the same time as an MOT task, nor is there a decrement when two MOT tasks are carried out simultaneously in different hemifields (Alvarez & Cavanagh, 2005). Such findings are consistent with these pairs of tasks using distinct resources. In the present study we examine the question whether MOT and a memory task can be carried out without performance decrement.

Visual Index theory assumes that tracking is carried out directly by an index mechanism in early vision and not by a process that involves matching perceptual information with information stored in memory. Although information about objects can eventually be stored in object files, the assumption is that such information is not present initially and even if it is, is not used in tracking. According to the theory, indexes stick to objects independently of what properties, if any, are encoded in the object files. Thus, Visual Index theory predicts that memory resources, such as those provided by short-term memory, are not used in cases of "pure" tracking—i.e., in tasks like MOT that do not require recall of object properties.

Using a dual-task MOT-Memory method, Fougne and Marois (2006) asked participants to perform MOT while holding colours and locations of coloured patches in WM and examined the dual-task interference. Participants memorized the colours and locations of briefly presented circles and then performed an MOT task. At the end of each trial, participants responded to the WM task first and then the MOT task. They found that when one tracks more targets, recognition performance is worse, or in other words, fewer objects can be stored. Nevertheless, this effect is disproportional in MOT and WM. When the number of objects being tracked increases by 1, the number of objects that can be stored drops by only 0.5. This drop in WM is also smaller than that caused by a second WM task.

Although there is evidence for some overlap in cognitive resources between the capacity of WM and the number of objects that can be perceptually individuated and tracked this overlap is relatively small. Moreover, at least four lines of analyses suggest that this overlap may only arise under special conditions.

First, object files (objects in WM) and indices (objects selected in perception) carry different type of information. Object files may store various visual and spatial features of objects, such as colour, orientation, and location (Treisman, 2006). In contrast, indices only select and refer to individual objects (Pylyshyn, 2007)—although some writers have proposed that in some cases there may be a compulsory encoding of location in the associated object files (Fencsik, Klieger, & Horowitz, 2007; Keane & Pylyshyn, 2006; Suganuma & Yokosawa, 2006). But perceptual information does not appear to be routinely encoded. When one of the objects being tracked changes its colour or shape occasionally, this change is rarely noticed (Bahrami, 2003), nor are observers able to report these features at better than chance level (Scholl, Pylyshyn, & Franconeri, 1999).

Second, there is evidence that the number of objects is not the only constraint on visual WM, so the correlation or competition between objects in WM and in perception may actually come from other constraints rather than limits to the memory capacity for objects. Wheeler and Treisman (2002) found that the number of features on a dimension is also a constraint on visual WM, allowing for resource competition based on features themselves. In contrast with Luck and Vogel's (1997) findings, Delvenne and Bruyer (2004) found that when half of the objects are defined by shape and the other half are defined by texture, the recognition performance of visual WM is better than when all the objects are defined by shape or texture. The maximum number of features can range from 1.6 to 4.4 objects, depending on stimuli types (Alvarez & Cavanagh, 2004). Yet it seems that the limit on MOT is confined to the number of objects tracked since the visual properties of individual objects do not appear to be encoded at all.

Third, if there were common cognitive resources for objects in WM and in perception we would expect a general dual-task interference, but many studies have shown that selective interference of perception in WM occurs only when both WM and perception tasks are visual or spatial (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; Hecker & Mapperson, 1997; Quinn & McConnell, 1996). For example, in Klauer and Zhao's (2004) study, memorizing a Chinese ideograph was impaired by the concurrent task of finding a stationary asterisk among 12 moving asterisks but not by the task of deciding whether a colour belonged to the red family or the blue family. For memorizing a dot location, the interference pattern was the reverse. Also Olivers, Meijer, and Theeuwes (2006) have recently shown the reverse case: Visual WM content interferes with attentional capture in a content-specific manner. There are also many other studies that fail to show a dual-task interference when the nontracking task was not spatial (Alvarez et al., 2005; Leonard & Pylyshyn, 2003).

Fourth, the influence of WM load on perceptual tasks also appears to be confined to tasks that involve spatial properties. Keeping two or four positions in WM during a visual search slows down the search (Oh & Kim, 2004; Woodman & Luck, 2004), whereas retaining four shapes has no influence on the search, although such a visual WM task is difficult (Woodman, Vogel, & Luck, 2001).

Such considerations suggest that objects in WM and objects being tracked in MOT might have no significant resource overlap. It may be that the weak relation or interference, found by Fougne and Marois (2006) arise because both the WM task and the perception task in these studies involve spatial processing. In the present study, we test this possibility by examining the dual-task interference between WM and MOT when the WM task is nonspatial.

EXPERIMENT 1: THE DUAL-TASK INTERFERENCE OF NONSPATIAL WM AND MOT

A visual WM task usually requires maintaining features of several visual items. When location is among the features to be remembered, the task would involve spatial processing. In the present paper, such a task is called a spatial WM task, and a visual WM task for nonspatial features is called a nonspatial WM task. The purpose of Experiment 1 was to compare the dual-task interference between nonspatial WM and MOT with that between spatial WM and MOT. As illustrated in Figure 1, participants were asked to remember colour-shape bindings of coloured patches or colour-location bindings of coloured squares. During the retention interval, they performed

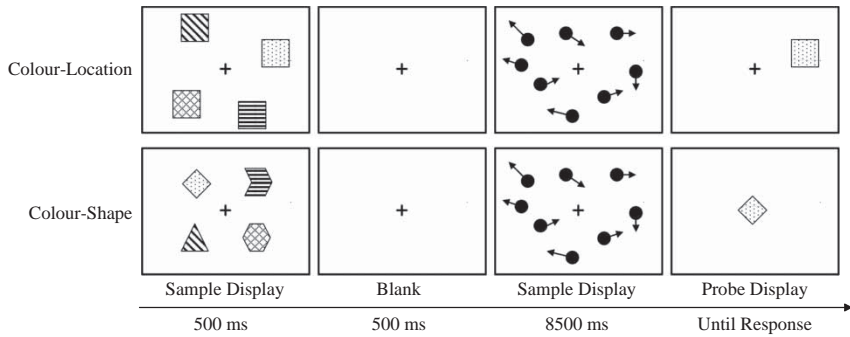


Figure 1. Time course of the WM-MOT task used in Experiment 1. Each different fill pattern represents a different solid colour.

an MOT task. Dual-task interferences were estimated as the difference of performances in such tasks and the corresponding single-task baselines.

Colour-shape conjunctions were used as WM material because it would make our results comparable to those of Fougny and Marois (2006), where similar stimuli were used. In addition it has been shown in many studies that recall and recognition of conjunctions is more difficult (more attentive or resource-demanding) than recall or single features (e.g., Delvenne & Bruyer, 2004; Postma & de Haan, 1996; Wheeler & Treisman, 2002; for counter-arguments, see Brockmole, Parra, Della Sala, & Logie, 2008; Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001).

If the weak interference, found by Fougny and Marois (2006), arises because both the WM task and the tracking task involve spatial processing, such interference effects should only be observed in the colour-location condition but not in the colour-shape condition.

Method

Participants

Fourteen paid participants, 17–24 years of age, were tested. All participants reported normal or corrected-to-normal vision and normal colour vision.

Stimuli

The display was presented on a PC with a 17-inch CRT monitor and was controlled by E-Prime experimental software (Schneider, Eschman, & Zuccolotto, 2002). The viewing distance was about 60 cm. The background

was white and the fixation point was a black cross subtending 0.6° in the centre of the screen.

Each item in the WM tasks subtended 1.0° . In the colour-location condition, the sample display consisted of two or four coloured squares. For each trial, the colours and locations of the items were separately selected at random. The colours were selected from seven preset colours. The locations were selected from a set of seven locations and the selection of each was made without replacement. The colours (and RGB values) were red (255, 0, 0), green (0, 255, 0), yellow (255, 255, 0), blue (0, 0, 255), cyan (0, 255, 255), saddle brown (128, 64, 0), and magenta (255, 0, 255). The locations were 2.9° from the fixation cross and equally spaced on an imaginary circle. The probe display was a single square whose colour and location was drawn from one item or separately from two items in the sample display.

In the colour-shape condition, the initial display consisted of two or four coloured patches, which occupied either the left two or all four vertexes of an imaginary square and were 1.6° from the fixation cross. The colours and shapes of the items were randomly selected without replacement from seven colours identical to those in the colour-location condition, and seven shapes as shown in Figure 2. The probe display was a centred coloured patch whose colour and shape corresponded to that of one item or to two different items in the sample display.

Objects of the MOT task were black disks. Each disk subtended 0.5° . There were eight disks and three of them were targets. Disks were scattered in a $15.3^\circ \times 11.5^\circ$ area around the fixation cross. The initial location of the disks was randomly selected. Each disk moved with a certain velocity and direction for a random duration averaged 500 ms, and then changed its velocity and direction unpredictably. The velocities varied from $8^\circ/s$ to $12^\circ/s$. The disks bounced away from area borders, from the fixation cross, and from each other. The distance between the centres of the disks was never less than 1.0° .

Procedure

There were three experimental conditions: Single-task of MOT (MOT-only), single-task of WM (WM-only), and dual-task of WM and MOT (WM-MOT).



Figure 2. The seven shapes used in Experiment 1.

MOT-only. At the start of a trial, eight identical disks appeared on the screen and three of them flashed for 1500 ms to indicate that they were the targets. Then all disks began to move and kept moving for 4000 ms. When they stopped, one disk expanded to a larger size. On 50% of the trials, this disk was among the three that had flashed (the targets). Participants were required to indicate whether this disk was one of the targets and they then made an unspeeded “yes” or “no” response, respectively by pressing “f” or “j”. The probe display would not disappear until a response had been made or 3000 ms had passed.

WM-only. The sample display was presented for 500 ms, followed by a retention period of 9000 ms. Then the probe display was presented and kept visible until response. In the colour-location condition, participants were instructed to memorize the colours and locations of the items in the sample display; in the colour-shape condition, participants were instructed to memorize the colours and shapes of the items. The memory set size was either two or four. On 50% of the trials, the probe item was one of the sample items; on the other 50% of the trials, the probe item was a combination of two sample items. Participants were asked to decide whether the probe item had occurred in the sample display and make an unspeeded “yes” or “no” response respectively by pressing “f” or “j”.

During the 9000 ms retention period, a blank screen was displayed in the first 500 ms and then the display was just like that in MOT-only condition except that the probe display lasted for 3000 ms. Participants were told to view the screen but ignore the moving disks.

WM-MOT. The procedure of WM-MOT resembled that of WM-only. The only difference was that participants were asked to perform MOT during the retention period of WM. After participants had made their response for the MOT task, a blank screen was used to fill up the remaining time of the 9000 ms retention period.

Each trial began with a fixation cross at the centre of the screen for 1000 ms. The intertrial interval was a random value between 2500 and 3500 ms. To prevent the possible phonological rehearsal of visual or spatial information, participants were required to repeat saying “1, 2, 3, 4” in Chinese throughout a trial in all trials.

There were five types of task: (1) MOT-only, (2) colour-location WM-only, (3) colour-location WM-MOT, (4) colour-shape WM-only, and (5) colour-shape WM-MOT. The five tasks were each presented in a block and the latter four were presented with two different WM set size (two or four). Overall there were nine conditions, each of which was repeated 16 times. Trials within a block were randomized for each participant and the sequence of the blocks was counterbalanced across participants. Before each block,

there were five practice trials. Each participant completed all the 169 trials in a 50-minute session.

Results

To correct for guessing, effective number of objects tracked (m) was computed as the measure of MOT performance and number of conjunctions remembered (K) as the measure of WM performance. The concept of effective number of objects tracked, which was used in Scholl, Pylyshyn, and Feldman (2000), and Pylyshyn and Annan (2006), refers to the number of targets actually tracked in an MOT task, with the assumption that participants track m of n targets and guess on the others. Our method of computing m is based on the hit rate (H , proportion correct when a target is probed) and the correct rejection rate (CR , proportion correct when a distractor is probed) of MOT:

$$m = n \frac{H + CR - 1}{CR}$$

(see Appendix A for complete proof).

For WM of feature conjunctions, we assume that a participant remember K conjunctions out of an array of N conjunctions. If at least one of the features of the probe is contained in the K conjunctions, she or he can make a correct response, otherwise make a guess (that's because each colour and each location occurs only once in a trial). For instance, if a participant remembers red-triangles and cyan-hexagon, when the probe is a red-diamond, she or he knows the probe is a new conjunction. Assuming H is the hit rate (proportion correct when the probe is an old conjunction) and CR is the correct rejection rate (proportion correct when the probe is a new conjunction)

$$K = \frac{N*H + N - 1 - \sqrt{(N*H + N - 1)^2 - 4*N*(N - 1)*(H + CR - 1)}}{2}$$

(see Appendix B for complete proof).

As our interest was in the dual-task interference between MOT and WM, our statistics focused on dual-task cost, the difference between the performance of MOT or WM in the single task and that in the dual task. What concerned us was whether the dual-task costs varied with conditions and whether the dual-task costs were different from zero. An alpha level of .05 was used as the criterion for a significant difference for all statistical tests in this paper.

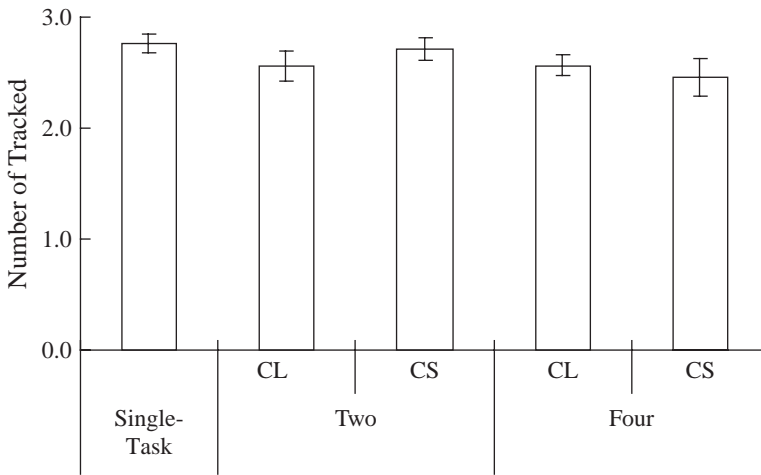


Figure 3. Mean effective number of objects tracked for the single-task and dual-task conditions of Experiment 1, with the latter as a function of WM material (colour-location or colour-shape) and set size of WM (two or four). Colour-location is abbreviated as CL and colour-shape as CS. Error bars show standard errors.

MOT

The mean effective number of objects tracked in the single-task condition and in each memory load condition is plotted in Figure 3. The dual-task costs of MOT were obtained by subtracting the effective number of objects tracked in the dual-task conditions from the effective number of objects tracked in the single-task condition. According to a 2 (WM material: Colour-location, colour-shape) \times 2 (set size of WM: Two, four) repeated measures ANOVA, WM material and set size of WM had no effects on the dual-task costs of MOT. We then collapsed the dual-task costs of MOT in these conditions and conducted a Bonferroni corrected *t*-test on the mean. The mean was not statistically different from zero, which meant that the performance of MOT was not influenced by WM load.

WM

The mean number of conjunctions remembered in the single-task and dual-task for each memory condition is plotted in Figure 4. Similar to the dual-task costs of MOT, dual-task costs of WM were obtained by subtracting the number of conjunctions remembered in the dual-task conditions from the number of conjunctions remembered in the single-task conditions. A 2 (WM material: Colour-location, colour-shape) \times 2 (set size of WM: Two, four) repeated-measures ANOVA analysis revealed that WM

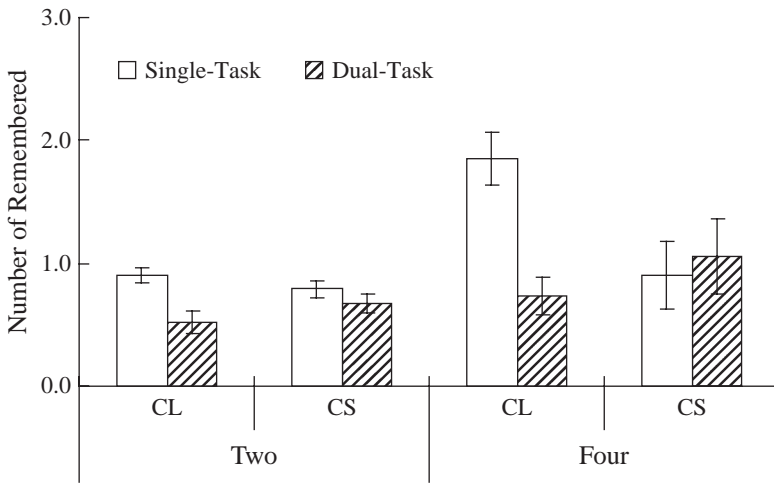


Figure 4. Mean number of conjunctions remembered, for the single-task and dual-task conditions of Experiment 1, as a function of WM material (colour-location or colour-shape) and set size of WM (two or four). Colour-location is abbreviated as CL and colour-shape as CS. Error bars show standard errors.

material had a significant effect on the dual-task cost of WM, $F(1, 13) = 17.16, p < .01$, and that WM material and set size of WM had an interaction, $F(1, 13) = 15.77, p < .01$. So we did *t*-tests separately on each condition to examine whether the dual-task cost in that condition was different from zero. The dual-task costs of WM in the colour-location condition at set size two and four were significantly above zero, $t(13) = 2.83, p < .01$, and $t(13) = 4.16, p < .01$, respectively, whereas the dual-task costs of WM in the colour-shape condition at set size two and four were not different from zero.

To be sure that the effect of WM material on the dual-task cost of WM was not due to the difference between the baselines of these two kinds of materials, we carried out a 2 (WM material: Colour-location, colour-shape) \times 2 (set size of WM: Two, four) repeated measures ANOVA analysis on the number of conjunctions remembered in single-task conditions. The main effect of WM, the main effect of set size of WM, and the interaction were significant, $F(1, 13) = 8.25, p = .01$, $F(1, 13) = 8.52, p = .01$, and $F(1, 13) = 8.96, p = .01$, respectively. A further simple effect analysis showed that memories for colour-location and colour-shape were significantly different at set size four (memories for colour-location were better than those for colour-shape), but were similar at set size two.

To rule out the effect of different baseline performances, a relative dual-task cost of WM (K_r) was computed as:

$$K_r = \frac{K_{\text{single task}} - K_{\text{dual task}}}{K_{\text{single task}}}$$

Then a 2 (WM material: Colour-location, colour-shape) \times 2 (set size of WM: Two, four) repeated measures ANOVA analysis on K_r was conducted. Results showed that the main effect of WM material was significant, $F(1, 13) = 8.04$, $p = .014$, thus indicating the dual-task cost of WM in the colour-location condition (mean $K_r = .49$) was greater than that in the colour-shape condition (mean $K_r = .08$); other effects were not significant. Again, we did t -tests separately on each condition to examine whether the relative dual-task cost (K_r) in that condition was different from zero. The dual-task costs of WM in the colour-location condition at set size two and four were significantly above zero, $t(13) = 3.91$, $p < .01$, and $t(13) = 6.83$, $p < .01$, respectively, whereas the dual-task costs of WM in the colour-shape condition at set sizes two and four were not different from zero. Therefore, after the correction for the baseline performances it is safe to conclude that WM performance was impaired in the dual task condition only when the content of the WM was spatially specific.

Discussion

As we expected, the dual-task interference between visual WM and MOT is specific to spatial WM tasks. Parallel to that of Fougny and Marois (2006), we found that the performance of spatial WM is impaired by the concurrent MOT task. But we also found that the maintenance of nonspatial WM is not influenced by performing MOT (at least with a set size of 2). It has been shown that spatial WM and nonspatial WM are impaired to the same extent by random-number generation, which requires generation of a random sequence of numbers between 1 and 10 and is supposed to demand on general executive functions such as the inhibition of dominant responses and information updating (Klauer & Zhao, 2004). Therefore, the disproportional effect of MOT on spatial WM and nonspatial WM found in this experiment suggests that a primarily spatial resource might be involved in performing the MOT task, which is not consistent with the Visual Index theory. We discuss this issue more in the General Discussion.

Besides, the performance of MOT does not become worse when WM is occupied, nor does it decline as the memory load increases from two objects, a modest load, to four objects, a severe load. If objects in WM and objects in perception have one and the same limit, one would predict a competition between the objects being tracked and the objects being remembered. Yet the present results seem to indicate the dissociation between visual objects and objects held in WM.

The K index seemed quite low (as low as about one object) in Experiment 1. In fact, according to the computation equation, for perfect performance, when $H = CR = 1$, $K = N - 1$. Thus, when the set size is 2, the maximum K is 1; and when the set size is 4, the maximum K is 3. Therefore the results of Experiment 1 showed that participants' performance were near perfect in the condition of set size 2; and lower but still acceptable in the condition of set size 4. According to Xu and Chun (2006), WM capacity is determined both by a fixed number of objects and by object complexity. Previous studies found K can be as low as 1.6 or even 1.4.

EXPERIMENT 2: THE ROLE OF IMPLICIT SPATIAL PROCESSING IN THE DUAL-TASK INTERFERENCE OF WM AND MOT

In Experiment 1 we observed that WM for colour-shape feature conjunctions is not affected by MOT, whereas Fougnie and Marois (2006) found in their Experiment 6 that WM for colours is impaired by MOT, which allowed them to eliminate a domain-specific account of the dual-task interference. Although memorizing colour-shape bindings and memorizing colours are not the same thing, this inconsistency cannot be predicted from the existing theories.

A comparison of the stimuli and procedures suggested that the difference in the spatial layouts of the WM sample display might be the cause. In the colour-shape condition of our Experiment 1, the locations of sample items were fixed and more centred. A sample item appeared 1.6° from the fixation point and 2.3° from its neighbour. In contrast, in their Experiment 6, the locations of sample items varied from trial to trial and were more peripheral. The locations were randomly selected from six possible locations that were 3.3° from the fixation cross and equally spaced on an imaginary circle. As automatic encoding or representation of spatial information is frequently reported in behavioural and neuroimaging studies (Coull & Frith, 1998; Epstein & Kanwisher, 1999; Lu & Proctor, 1994; Meegan & Honsberger, 2005), it is possible that different types of spatial layouts lead to different amounts of automatic encoding or representation of spatial information and finally different interference patterns.

To test this possibility directly, in Experiment 2 we introduced spatial layout of WM sample display as an independent variable, which had two levels. One level was identical to that of the colour-shape condition of Experiment 1 and the other was the same as that of Experiment 6 of Fougnie and Marois (2006), namely, the centred condition and the dispersed condition. Another independent variable of Experiment 2 was number of targets in MOT. We hypothesized that the spatial information

in the centred condition would be far less processed than that in the dispersed condition because locations of the stimuli in the centred condition were fixed, whereas those in the dispersed condition were randomly selected.

Method

Participants

Twenty-eight paid participants, 19–25 years of age, were tested. All reported normal or corrected-to-normal vision and normal colour vision. None of them had participated in Experiment 1. Fourteen of them were randomly assigned to the centred condition and the other fourteen to the dispersed condition.

Stimuli

The MOT displays in Experiment 2 were identical to those in Experiment 1 except that the number of targets could be one, two, or four.

Each item in the WM tasks subtended 1.0° . The sample display was three coloured squares. To match the difficulty of our colour memory tasks with that of Fougne and Marois (2006), we used 10 candidate colours. Seven of these were the colours used in Experiment 1 and the three new (and RGB values) were blue violet (128, 0, 255), orange (255, 128, 0), and olive (128, 128, 64). The colours of the items were randomly selected without replacement from these colours. The spatial settings of sample displays are illustrated in Figure 5. For the centred group, coloured squares appeared at three fixed locations 1.6° from the fixation cross. For the dispersed group,

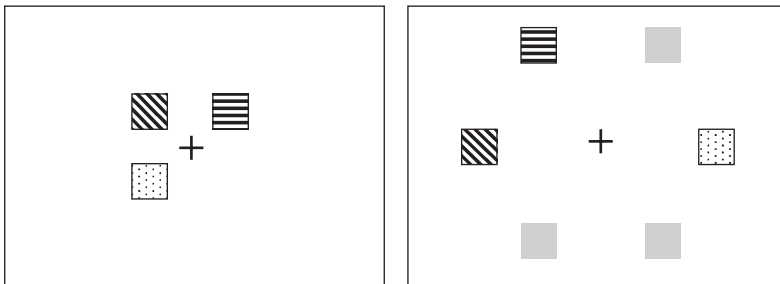


Figure 5. Sample displays used in Experiment 2. Each different fill pattern represents a different solid colour. For the centred group (the left panel), three squares are at three fixed locations near the fixation cross. For the dispersed group (the right panel), three squares are at three randomly selected locations. The light grey squares without fill patterns represent other possible locations and are invisible in the real display.

colour squares appeared at three randomly selected locations among six possible ones, which were 3.4° from the fixation cross and dispersed equally on an imaginary circle. The probe display was a single colour square at the centre of the screen.

Procedure

Like Experiment 1, there were three kinds of tasks, MOT-only, WM-only, and WM-MOT.

MOT-only. The procedure was the same as that of Experiment 1 except that the number of targets could be one, two, or four.

WM-only and WM-MOT. The basic procedures were like those of Experiment 1 with the following exceptions. The WM task was to remember colours of squares and subsequently decide whether the probe colour had occurred before. The timings were adjusted to match those of Fougny and Marois (2006), with the WM sample display lasting for 400 ms and followed by a 1200 ms blank.

Similar to Experiment 1, articulatory suppression was used to prevent the phonological rehearsal in WM.

Each participant was allocated to either the centred group or the dispersed group and completed three blocks of tasks, i.e., MOT-only, WM-only, WM-MOT. Number of targets in MOT (one, two, or four) was intermixed within blocks. Each condition was repeated for 16 times. Trials within a block were randomized for each participant and the sequence of the blocks was counterbalanced across participants. Before each block, there were five practice trials. Each participant completed all the 127 trials in a 40-minute session.

Results

As in Experiment 1, effective number of objects tracked (m) was computed as the measure of MOT performance to correct for guessing. For WM, number of colours remembered (K) was computed with the method developed by Pashler (1987). The method assumes that if a participant holds K colours in memory out of an array of N colours, then when the probe is an old colour, in K/N of the trials, the participant can recognize it and in the rest of the trials, the participant cannot recognize it but makes a false alarm (yes response). These response rates are governed by the relation

$$K = N \frac{H - F}{1 - F},$$

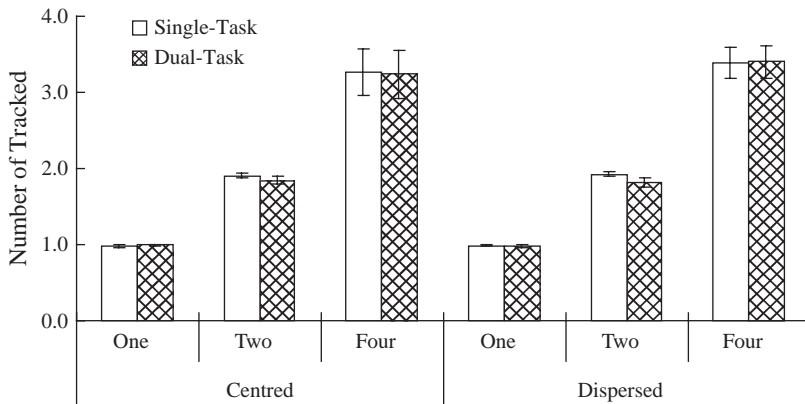


Figure 6. Mean effective number of objects tracked for the single-task and dual-task conditions of Experiment 2, as a function of layout group (centred or dispersed) and number of targets (one, two, or four). Error bars show standard errors.

where N is the set size of colours, H is the hit rate when the probe is an old colour, and F is the false alarm rate when the probe is a new colour.

MOT

The mean effective number of objects tracked in the single-task and dual-task conditions is plotted in Figure 6 as a function of layout group and number of targets. The dual-task costs of MOT were obtained for each group by subtracting the effective number of objects tracked in the dual-task conditions from the effective number of objects tracked in the single-task conditions. A 2 (layout group: Centred, dispersed) \times 3 (number of targets: One, two, four) mixed-design ANOVA analysis on the tracking performance data showed that no main effects or interaction were significant. The mean dual-task cost of MOT collapsed across all the conditions was not different from zero, suggesting that the performance of MOT was not impaired by the memory load.

WM

The mean number of colours remembered at each MOT load in each layout group condition is plotted in Figure 7. Dual-task costs of WM were obtained for each group by subtracting the number of colours remembered in the dual-task conditions from the number of colours remembered in the single-task condition. The dual-task costs of WM were submitted to a 2 (layout group: Centred, dispersed) \times 3 (number of targets: One, two, four) mixed-design ANOVA analysis. The interaction was significant, $F(2, 52) =$

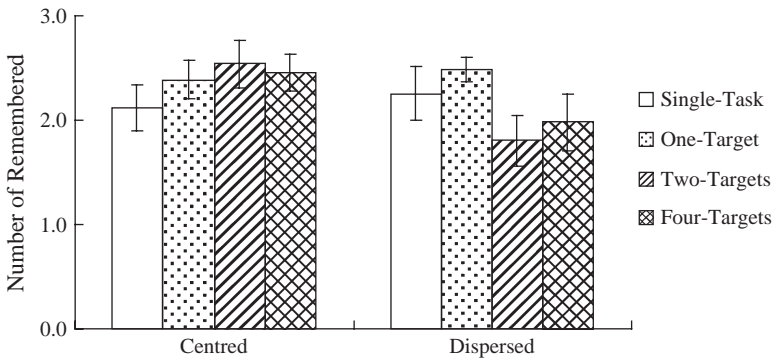


Figure 7. Mean number of colours remembered of Experiment 2, as a function of layout group (centred or dispersed) and MOT load (single-task, one target, two targets, or four targets). Error bars show standard errors.

4.70, $p = .01$. No main effects were significant. Simple effect analysis revealed that number of targets had no effect on the centred group, but had a significant effect on the dispersed group, $F(2, 26) = 7.75$, $p < .01$. According to a post hoc comparison, for the dispersed group, tracking two or four targets impaired WM to a larger extent than tracking one target did.

To preclude the effect of baseline differences, we compared the baseline WM scores of the centred group (mean = 2.12) and the dispersed group (mean = 2.25) by a t -test and found no significant differences.

Discussion

The results of Experiment 2 confirm our speculation that the spatial layout difference in WM sample displays is the cause of the inconsistency between our Experiment 1 and Experiment 6 of Fougne and Marois (2006). The WM of the dispersed group but not that of the centred group is subject to MOT interference. Therefore, it is tenable to conclude that the maintenance of nonspatial WM is not influenced by performing MOT.

This difference between the two layout conditions may be an effect of automatic representation of spatial information, which is consistent with several recent studies (Hollingworth, 2007; Treisman & Zhang, 2006). It has long been known that task-irrelevant spatial information can be automatically represented, as shown by the compatibility effect of the spatial property of a stimulus and its response (Lu & Proctor, 1994), or by the effect of WM for shapes on the pattern of eye movements, in which where the shapes originally appeared makes a difference (Meegan & Honsberger, 2005). In the centred condition, the sample colours stay at the same locations in each trial, which may lead participants to neglect the spatial properties of them. In

contrast, in the dispersed condition the locations of the sample colours vary with trials, which may be noticed and induce a representation of colours based on locations. When the memory of locations is disrupted by MOT, the representation of colours accompanied with locations is also impaired. In this way, only the retention of the dispersed condition but not that of the centred condition is subject to MOT. Another possible explanation is that the items in the dispersed condition cannot be so richly processed as those in the centred condition, for the attentional resolution drops rapidly with eccentricity (Intriligator & Cavanagh, 2001), or for the density of processing resources decreases as the size of the attentional field increases (Eriksen & St. James, 1986), and thus are more difficult to enter WM. A third possibility is that maintaining colours is more resource demanding in the dispersed condition than in the centred condition, and therefore the WM in the dispersed condition is more fragile in the face of interference. The latter two possibilities can be excluded by the fact that single-task memory performances of the two layout conditions are alike but far from perfect.

GENERAL DISCUSSION

In the present study, with a WM-MOT dual task paradigm, we found that spatial WM but not nonspatial WM is impaired by MOT and task-irrelevant spatial factors can determine whether WM is susceptible to interference by a concurrent MOT task. To be more specific, in Experiment 1, WM of colour-location conjunction but not of colour-shape conjunction was showed to be impaired by a concurrent MOT task; in Experiment 2, WM for coloured patches were demonstrated to be unaffected by MOT when their locations were fixed and constrained in a more compact layout. Furthermore, participants' performances of the MOT task seemed to be insensitive to WM load, no matter it is empty or full. These results imply that objects in WM and objects that are selected in MOT do not compete for the same cognitive resources in a general sense, demonstrating that previous evidences favouring a common object-based resource for visual WM and perception might be contaminated by spatially specific effects.

One's experience in doing the MOT task leaves the clear impression that MOT is attentionally demanding and effortful (e.g., Culham et al., 1998; He, Cavanagh, & Intriligator, 1997; Scholl, 2001; Yantis, 1992). Yet Pylyshyn's visual index theory (1989, 2001, 2003, 2007) assumes that the mechanism underlying tracking is automatic and preattentive. According to the theory the indexes automatically remain attached to each selected individual target. Thus, at this stage memory load would have no effect on the MOT performance, even though it might have an effect at a subsequent stage of using the indexes to generate a judgement and make a response.

The present finding that a WM task does not impair MOT is consistent with the Visual Index theory, yet the theory does not predict the unidirectional interference of the MOT task on the WM task nor does it naturally explain the fact that this interference is confined to spatial information. We suggest the unidirectional interference may be due to the insensitivity nature of the MOT task to a secondary task and the specificity of the spatial interference may result from the spatial property in the response stage of the MOT task.

The insensitivity of MOT performance to interference has also been reported in other studies. For example, no decrement in performance is observed when a search task (Alvarez et al., 2005) or a monitoring task (Leonard & Pylyshyn, 2003) are carried out at the same time as an MOT task. A series of studies by Zhang (2008) also showed that spatial memory load had no effects on the MOT performance but the reverse did not hold, i.e., spatial memory was impaired when the number of targets that had to be tracked was increased. Using a similar procedure as the present study, Zhang examined the dual-task interference between a Corsi Blocks Task (CBT) and an MOT task. Even when up to six serial positions had to be held in WM, the concurrent MOT performance remained consistent. In six of the seven experiments, the unidirectional interference effect of the MOT task was replicated.

It should be noted that in the present study and in Zhang (2008), participants made the MOT response before the WM response. When two tasks share a common resource, and one task (the interpolated task) occurs during the execution of another task (the background task), evidence suggests that the background task is more likely to be impaired (Sigman & Dehaene, 2006).

Note that Leonard, Pylyshyn, Dennis, and Cohen (2002) reported that a secondary task of monitoring a colour change in stationary objects interfered with MOT performance (much as shown by Fougny & Marois, 2006). But then in a subsequent study (Leonard & Pylyshyn, 2003), it was shown that this only occurs if the monitoring task requires a response *during tracking*, rather than after the tracking response had been made. Such an asymmetry has been observed in other tasks as well (Jolicoeur, 1999) and points to a general finding that making a response may be more disruptive than the task itself. Thus, in our WM-MOT paradigm, when subjects completed the MOT task and were waiting to make a response after the WM task, the performance on the MOT task would be more likely to be impaired. That was exactly the case in Fougny and Marois (2006).

In order to judge correctly that a particular object in the MOT task is a target, observers must focus attention on the cued object and also must have selected all the targets in order to judge that the cued one is a target. This in itself is a spatial task since in order to decide whether the object at $\{x,y\}$ is

one of the targets the observer has to compare the cued objects with a selected object. Consequently making a response in an MOT task logically requires using the response-selection stage. The WM also requires both a comparison operation as well as focusing on particular objects. This may explain why there is an interference effect only when the WM requires taking into account the location of symbols.

The present findings seem consistent with what has been found when search and visual working memory tasks are combined. Searching through objects does not interfere with working memory for objects but does for location memory in a set size dependent manner (Oh & Kim, 2004; Woodman & Luck, 2004; Woodman et al., 2001). Taken together, it seems that objects kept in visual WM and visually selected perceptual objects are not so closely linked by focus of attention as proposed by Cowan (2001), their interactions appear to be mediated by the processing of spatial information involved.

REFERENCES

- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111.
- Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*, 637–643.
- Alvarez, G. A., Horowitz, T. S., Arsenio, H. C., DiMase, J. S., & Wolfe, J. M. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology: Human Perception and Performance*, *31*(4), 643–667.
- Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition*, *10*(8), 949–963.
- Brockmole, J. R., Parra, M. A., Della Sala, S., & Logie, R. H. (2008). Do binding deficits account for age-related decline in visual working memory? *Psychonomic Bulletin and Review*, *15*, 543–547.
- Coull, J. T., & Frith, C. D. (1998). Differential activation of right superior parietal cortex and intraparietal sulcus by spatial and nonspatial attention. *NeuroImage*, *8*(2), 176–187.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–185.
- Culham, J. C., Brandt, S., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. H. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, *80*, 2657–2670.
- Davis, G. (2001). There is no four-object limit on attention. *Behavioral and Brain Sciences*, *24*(1), 119–120.
- Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwinding visuo-spatial memory. *Neuropsychologia*, *37*(10), 1189–1199.
- Delvenne, J. F., & Bruyer, R. (2004). Does visual short-term memory store bound features? *Visual Cognition*, *11*(1), 1–27.
- Epstein, R., & Kanwisher, N. (1999). Repetition blindness for locations: Evidence for automatic spatial coding in an RSVP task. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1855–1866.

- Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics*, *40*, 225–240.
- 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 and Psychophysics*, *69*(4), 567–577.
- Fougnie, D., & Marois, R. (2006). Distinct capacity limits for attention and working memory: Evidence from attentive tracking and visual working memory paradigms. *Psychological Science*, *17*(6), 526–534.
- Haladjian, H., & Pylyshyn, Z. W. (2008). Object-specific preview benefit enhanced during explicit multiple object tracking. *Journal of Vision*, *8*(6), 497.
- He, S., Cavanagh, P., & Intriligator, J. (1997). Attentional resolution. *Trends in Cognitive Sciences*, *1*, 115–121.
- Hecker, R., & Mapperson, B. (1997). Dissociation of visual and spatial processing in working memory. *Neuropsychologia*, *35*(5), 599–603.
- Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 31–47.
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*(3), 171–216.
- Jolicoeur, P. (1999). Concurrent response-selection demands modulate the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 1097–1113.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219.
- 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*(4), 346–368.
- Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General*, *133*(3), 355–381.
- Leonard, C., & Pylyshyn, Z. W. (2003). Measuring the attentional demand of multiple object tracking (MOT) [Abstract]. *Journal of Vision*, *3*(9), 582a.
- Leonard, C., Pylyshyn, Z. W., Dennis, J. L., & Cohen, E. H. (2002). *The effect of a secondary monitoring task on multiple object tracking*. Paper presented at the Vision Sciences annual conference, Sarasota, FL.
- Lu, C.-H., & Proctor, R. W. (1994). Processing of an irrelevant location dimension as a function of the relevant stimulus dimension. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 286–298.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.
- Meegan, D. V., & Honsberger, M. J. M. (2005). Spatial information is processed even when it is task-irrelevant: Implications for neuroimaging task design. *NeuroImage*, *25*(4), 1043–1055.
- Noles, N. S., Scholl, B. J., & Mitroff, S. R. (2005). The persistence of object file representations. *Perception and Psychophysics*, *67*(2), 324–334.
- Oh, S. H., & Kim, M. S. (2004). The role of spatial working memory in visual search efficiency. *Psychonomic Bulletin and Review*, *11*(2), 275–281.
- Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1243–1265.
- Pashler, H. (1987). Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Perception and Psychophysics*, *41*(3), 191–201.
- Postma, A., & de Haan, E. H. F. (1996). What was where? Memory for object locations. *Quarterly Journal of Experimental Psychology*, *49*(A), 178–199.

- Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32(1), 65–97.
- Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80(1–2), 127–158.
- Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*. Cambridge, MA: MIT Press/Bradford Books.
- Pylyshyn, Z. W. (2007). *Things and places: How the mind connects with the world (Jean Nicod Lectures Series)*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. W., & Annan, V. (2006). Dynamics of target selection in multiple object tracking (MOT). *Spatial Vision*, 19(6), 485–504.
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3), 179–197.
- Quinn, J. G., & McConnell, J. (1996). Irrelevant pictures in visual working memory. *Quarterly Journal of Experimental Psychology*, 49A(1), 200–215.
- Rensink, R. A. (2001). Four-sight in hindsight: The existence of magical numbers in vision. *Behavioral and Brain Sciences*, 24(01), 141–142.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Pittsburgh: Psychology Software Inc.
- Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, 80(1/2), 1–46.
- Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2000). What is a visual object? Evidence from “target merging” in multiple object tracking. *Investigative Ophthalmology and Visual Science*, 41(4), S759–S759.
- Scholl, B. J., Pylyshyn, Z. W., & Franconeri, S. L. (1999). When are featural and spatiotemporal properties encoded as a result of attentional allocation? *Investigative Ophthalmology and Visual Science*, 40(4), 4195.
- Sigman, M., & Dehaene, S. (2006). Dynamics of the central bottleneck: Dual-task and task uncertainty. *PLoS Biology*, 4(7), 1227–1238.
- Suganuma, M., & Yokosawa, K. (2006). Grouping and trajectory storage in multiple object tracking: Impairments due to common item motions. *Perception*, 35(4), 483–495.
- Treisman, A. (2006). Object tokens, binding and visual memory. In H. Zimmer, A. Mecklinger, & U. Lindenberger (Eds.), *Handbook of binding in human memory: Perspectives from cognitive neuroscience* (pp. 315–338). New York: Oxford University Press.
- Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory and Cognition*, 34(8), 1704–1719.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 92–114.
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131(1), 48–64.
- Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, 37(1), 25–43.
- Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychonomic Bulletin and Review*, 11(2), 269–274.
- Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science*, 12(3), 219–224.
- Xu, Y., & Chun, M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, 440, 91–95.
- Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, 24(3), 295–340.
- Yantis, S., & Johnson, D. N. (1990). Mechanisms of attentional priority. *Journal of Experimental Psychology: Human Perception and Performance*, 16(4), 812–825.

Zhang, H. (2008). *Spatial working memory and scope of attention*. Unpublished doctoral dissertation, Institute of Psychology, Chinese Academy of Sciences. (In Chinese)

Manuscript received July 2008

Manuscript accepted July 2009

First published online October 2009

APPENDIX A: COMPUTING THE EFFECTIVE NUMBER OF OBJECTS TRACKED (M)

Let n = total number of targets, H = observed hit rate (proportion correct when a target is probed), CR = observed correct rejection rate (proportion correct when a distractor is probed), and m = number of targets that is actually tracked.

Assume the following strategy: Track m objects, and guess on the others. When an unknown object is probed, the observer has a probability of x to identify it as a target.

Assuming this strategy,

$$H = \frac{m}{n} + x \left(1 - \frac{m}{n} \right)$$

$$CR = 1 - x$$

Solving for m :

$$m = n \frac{H + CR - 1}{CR}$$

Appendix B: COMPUTING THE NUMBER OF CONJUNCTIONS REMEMBERED (K)

Let N = total number of conjunctions, H = observed hit rate (proportion correct when the probe is an old conjunction, i.e., its two features are drawn from the same item previously presented), CR = observed correct rejection rate (proportion correct when the probe is a new conjunction, i.e., its two features are drawn separately from two items), and K = number of conjunctions that is actually remembered.

Assume the following strategy: The observer remembers K conjunctions, that is, $2K$ features and the match between them. For a probe, when both of the features are known, the observer recognizes the probe as an old or new

conjunction; when one feature is known and the other is unknown, the observer correctly identifies the probe as a new conjunction; when neither of the features is known, suppose the observer has a probability of x to judge the probe as an old conjunction.

Assuming this strategy, if the probe is an old conjunction,

$$H = \frac{K}{N} + x \left(1 - \frac{K}{N} \right)$$

If the probe is a new conjunction, assume the observer has a proportion of E to know at least one feature,

$$E = 1 - \frac{C_{N-K}^2}{C_N^2} = 1 - \frac{(N-K)(N-K-1)}{N(N-1)}$$

$$CR = E + E(1-x) = 1 - x \frac{(N-K)(N-K-1)}{N(N-1)}$$

Solving for K :

$$K = \frac{N*H + N - 1 - \sqrt{(N*H + N - 1)^2 - 4*N*(N - 1)*(H + CR - 1)}}{2}$$