Perception & Psychophysics
2007, 69 (2), 172-184

Tracking unique objects

Todd S. Horowitz
Brigham and Women's Hospital, Boston, Massachusetts
and Harvard Medical School, Boston, Massachusetts

Sarah B. Klieger
Brigham and Women's Hospital, Boston, Massachusetts

David E. Fencsik
Brigham and Women's Hospital, Boston, Massachusetts
and Harvard Medical School, Boston, Massachusetts

Kevin K. Yang
Boston College, Chestnut Hill, Massachusetts

George A. Alvarez
Harvard University, Cambridge, Massachusetts

and

Jeremy M. Wolfe
Brigham and Women's Hospital, Boston, Massachusetts
and Harvard Medical School, Boston, Massachusetts

Is content addressable in the representation that subserves performance in multiple-object-tracking (MOT) experiments? We devised an MOT variant that featured unique, nameable objects (cartoon animals) as stimuli. There were two possible response modes: standard, in which observers were asked to report the locations of all target items, and specific, in which observers had to report the location of a particular object (e.g., "Where is the zebra?"). A measure of capacity derived from accuracy allowed for comparisons of the results between conditions. We found that capacity in the specific condition (1.4 to 2.6 items across several experiments) was always reliably lower than capacity in the standard condition (2.3 to 3.4 items). Observers could locate specific objects, indicating a content-addressable representation. However, capacity differences between conditions, as well as differing responses to the experimental manipulations, suggest that there may be two separate systems involved in tracking, one carrying only positional information, and one carrying identity information as well.

We all live in a dynamic environment characterized by multiple, distinctive objects that move through the world. In the study of visual attention, however, when researchers have studied distinctive objects, those objects have not moved (e.g., Henderson & Hollingworth, 2003; Oliva, Torralba, Castelhano, & Henderson, 2003; Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002), and when they have instead examined the ability to attend to moving objects, those objects have not been distinctive. In this article, we will explore the role of attention in keeping track of identifiable objects as they move through the world.

In recent years, the multiple-object-tracking (MOT) task (Pylyshyn & Storm, 1988) has emerged as a primary tool for studying dynamic attention—that is, attention to objects over time. In the standard version of MOT, the observer is initially presented with a display of eight identical items, with four of the items blinking on and off to identify them as targets. All of the items then move for several seconds, during which the observer has to keep track of the targets. When the objects stop, the observer is asked to discriminate between targets and nontargets, either by pointing out all of the targets or by indicating whether a probe item is a target or a nontarget. The typical result is that humans can accurately track about four targets, although performance strongly depends on stimulus factors such as the speed of the items (Alvarez & Franconeri, 2004; Oksama & Hyönä, 2004).

The MOT paradigm has proved to be a versatile research tool. It has been used to study the nature of visual objects (Scholl & Pylyshyn, 1999; Scholl, Pylyshyn, & Feldman, 2001), the capacity for change detection (Bahrami, 2003; Saiki, 2002), and the allocation of attention in depth (Viswanathan & Mingolla, 2002). Furthermore, despite its

T. S. Horowitz, toddh@search.bwh.harvard.edu

artificial nature, the MOT paradigm seems to capture the demands of important real-world challenges. For instance, air traffic controllers need to track the changing positions of multiple planes in an airspace (see Allen, McGeorge, Pearson, & Milne, 2004), drivers need to pay attention to other cars in their immediate vicinity, athletes need to track teammates and opponents, and daycare providers need to be aware of the movements of several children at once.

Although the MOT task can be informative, it differs from such real-world tasks in critical and related ways. First, objects in the real world (cars, children, etc.) are rarely identical; they can differ in visual features as well as in characteristic motion. Second, the type of real-world answer that we demand of our visual system is typically different from the answer demanded of the observer during MOT. In the standard MOT task, observers are asked to differentiate between tracked targets and untracked nontargets. However, out in the world we rarely ask questions like "which children are you tracking now?" Instead, we often wish to know the whereabouts of a specific child, car, or plane.

In this article, we seek evidence for this ability in an MOT task. If an observer is tracking four unique targets, is he or she able to distinguish between those targets and report the location of a particular one? There are reasons to argue both for and against such "content-addressable" representations underlying MOT performance. We will address the question from the three dominant theoretical perspectives.

A prominent approach to MOT is Pylyshyn's theory of visual indexes, or FINSTs (Pylyshyn, 1989). Visual index theory assumes that MOT is mediated by a limited-capacity preattentive representation consisting of pointer-like indexes that can be attached to objects. These indexes convey the location of the object in question and provide priority access to attention (Sears & Pylyshyn, 2000). They can be likened to fingers (hence the term FINST, from "*f*ingers of *inst*antiation") that can be placed on the targets. At the end of the trial, the observer simply asks which objects his "fingers" point to. As with fingers, there are only a handful of indexes available.

Visual index theory seems to be ambivalent about the question of content addressability. One interpretation of the theory holds that the tracking representation is not content addressable, since the indexes only tell you where they are. They do not encode object identity or feature information. Another interpretation of the theory, however, is that the representation must be content addressable, since in order to know that item X is a target at any time t_i, you must know that it is the same item that was a target at time t_0. Pylyshyn (2004) refers to this as the *discrete reference principle*.

So far, data collected by Pylyshyn and his colleagues have supported the first interpretation. For instance, Scholl, Pylyshyn, and Franconeri (1999) showed that when items stopped moving, observers were able to accurately report the previous direction and speed of targets but not of nontargets. However, when the shape or color of the items was masked, observers were unable to accurately report the premask features of either targets or nontargets. From these results, Scholl et al. (1999) proposed a distinction between *spatiotemporal* properties of objects, such as speed, direction, and trajectory, and *featural* properties, such as color or shape. Although featural properties can change as objects move through different lighting environments and nonrigid transformations, spatiotemporal properties are the key to object continuity. Scholl et al. (1999) argued that visual indexes encode only spatiotemporal properties.

In order to explicitly test the discrete reference principle, Pylyshyn (2004) attached identities to targets during the target acquisition phase of the MOT trial, either by presenting a number inside the target disks or by having target disks start off in different corners of the display. To his surprise, he found that observers were rather poor at reporting the identities, particularly the numerical labels, even when they had successfully tracked the target objects. Part of this discrepancy apparently resulted from observers inadvertently swapping target identities when targets passed close to one another. Pylyshyn (2004) also speculated that the "internal name" might not be consciously available; in other words, although some part of the visual system was aware that a given disk was the same object as a target identified during the acquisition phase, there was no way to link this internal representation with an external label that the observer could report.

Instead of hypothesizing visual indexes, Yantis (1992) stressed the importance of perceptual grouping. According to Yantis, observers in an MOT task imagine that the target items form the vertices of a virtual polygon. Yantis has demonstrated that when the target items move into a configuration that collapses the polygon, tracking suffers. In contrast to the preattentive visual index account, Yantis's perceptual grouping account assumes that tracking is mediated by a spatial working memory representation. On this view, holding items in spatial working memory serves to make them items of spatial attention as well (Awh, Jonides, & Reuter-Lorenz, 1998). It is not clear whether such a representation could be content addressable. On the one hand, since this view is based on a top-down structure that adds information rather than a preattentive structure that discards it, one might expect the system to know which target is which. On the other hand, if the representation is purely spatial in nature, the system might not know.

Finally, Kahneman and Treisman's object file theory (Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992) has often been invoked to explain MOT results. The object file concept was developed to explain the perceptual continuity of objects over time. As we noted above, objects change their appearance over time as they move through space. The interpretation given to an object can also change as the featural information changes. Thus, as an object in the sky approaches, it can be perceived as a bird, then a plane, then Superman, all without losing the sense that a single object has been approaching. Kahneman and Treisman proposed that a single representation, the object file, is responsible for the perception of continuity. When an object is first attended, an object file is opened for it, and all subsequent information about that object goes into the object file. When the object is attended at some later time, a process of impletion recalls the appropriate object file.

In contrast to visual indexes, object files are clearly content addressable. In their seminal article presenting evidence for the object file concept, Kahneman, Treisman, and Gibbs (1992) developed the reviewing paradigm. In this paradigm, observers see a preview display that includes a number of objects (at least two). The objects are characterized by some feature, often a letter (see, e.g., Mitroff, Scholl, & Wynn, 2004). The featural information disappears, and then some sort of linking display, usually consisting of smooth or apparent motion, connects the objects in the preview display with objects in a target display. The observer then has to respond to the target display, for example by identifying a letter. Kahneman et al. found that if the letter was the same as the one presented in that object in the preview display, observers were faster to respond than if that letter had appeared in another object in the preview display, indicating that information about what was in that object had been preserved.

Here, we introduce a new variation on the basic MOT task that allows us to look at tracking of unique objects (a related paradigm was developed by Oksama & Hyönä, 2004; we will return to this in the Discussion section). In our task, the objects consisted of unique cartoon animals (see Figure 1). We chose these stimuli because they were visually distinct and easily nameable.

The principal technical problem with such stimuli is that observers could simply memorize the identities of the target animals at the beginning of the trial, then respond on the basis of those identities at the end of the trial, without actually tracking the targets while they moved. We made this strategy impossible by hiding the animals behind cartoon cactuses at the end of the trial. These cactuses were present throughout the trial, and the animals moved in front of them. At the end of the trial, the depth relationships were switched, and the animals moved behind the cactuses and remained there. In other work, we have found that the sudden appearance of previously invisible occluders, whether one at a time or all at once, did not disrupt tracking performance (Horowitz, Birnkrant, Fencsik, Tran, & Wolfe, 2006; see also Scholl & Pylyshyn, 1999). Thus, we consider it unlikely that a mere change in the depth relations between tracked animals and occluding cactuses should be problematic.

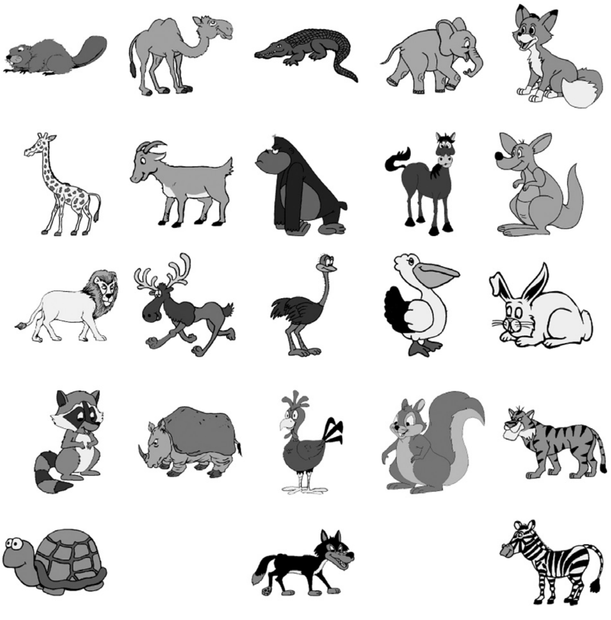

Figure 1. Cartoon animals used as stimuli. See Appendix A for a list of stimulus names.

A second technical problem with this arrangement is that, if the end of each trial were predictable, observers could anticipate this and locate the targets just prior to occlusion. To thwart such a strategy, we made the time from tracking onset to occlusion variable from trial to trial. The animals moved in such a fashion that they all approached the stationary cactuses every 1,333 msec. The number of such cycles in a trial varied randomly, so if observers were noting target locations only at times when they were likely to be occluded by cactuses, they would have to successfully search the display for the tracked animals in a manner that produced correct search results every 1,333 msec. This seems unlikely.

At the end of a trial, we asked observers one of two questions. These questions were blocked in Experiments 1–3 and mixed within blocks in Experiment 4. The *standard* question was “Where are all the targets?” Observers had to click on the cactuses in front of each target. The *specific* question was “Where is the _____?,” with the blank filled by one of the target animals. The observer then had to click on the cactus where that particular animal was “hiding.”

These two questions obviously have different rates of chance performance. In order to compare performance on these two different questions, we computed a common metric of *capacity*, based on the number of items tracked, corrected for guessing (the details are described in the Method section).

In Experiment 1, we found that capacity was substantially lower for the specific question than for the standard question in a blocked design. This “content deficit” was replicated in four more experiments. Experiment 2 ruled out the possibility that the content deficit is due to interference in short-term memory. In Experiment 3, we confirmed that the data do not change if we increase the variability of the tracking duration. Experiment 4 replicated the content deficit with a mixed design, using both visual and auditory postcues. The remaining experiments employed only the standard question, but took advantage of the possibilities opened by our method. Experiments 5 and 6 measured how much additional capacity is gained by using unique rather than identical stimuli. Finally, Experiment 7 measured capacity in a “full report” version of the standard task.

METHOD

Observers

Unless otherwise specified, each experiment involved 8 observers recruited from the Brigham and Women’s Hospital Visual Attention Laboratory volunteer pool. The observers ranged in age from 18 to 55 years (M = 28.3 years, SD = 9.7);[1] all had visual acuity of at least 20/25 when corrected and had passed the Ishihara color screen. The observers gave IRB-approved informed consent and were compensated $10/h for participation. Twelve of the observers participated in more than one experiment.

Apparatus and Stimuli

The visual stimuli were presented on 21-in. color CRT monitors (SuperScan Mc801 RasterOps and Mitsubishi Diamond Pro 91TXM) controlled by Power Macintosh G4 computers (Mac OS 9.2.2) running MATLAB 5.2.1 using the Psychophysics Toolbox routines (Brainard, 1997; Pelli, 1997). Monitor spatial resolution was set to 1,024 × 768 pixels. The display area subtended 36.1º × 27.1º of visual angle at a viewing distance of 57.4 cm. Monitor refresh rates were set to 75 Hz (13.3 msec per frame).

The tracking stimuli were 23 cartoon animals (Figure 1; see Appendix A for a list) in 8-bit color. Occluders were two types of cartoon cactuses. Cartoon images were scaled so that the maximum dimension (horizontal or vertical) was 3.5º for animals and 5.3º for cactuses. The background was yellow.

Cactuses were placed in a 5 × 4 grid. Grid cells were 6.2º on a side, so cactuses were separated by 0.9º edge to edge. The two types of cactuses were assigned to cells at random on each trial, so the background cactus array was different from trial to trial.

Auditory probes (Experiments 4, 5, and 6) were recorded in a female voice and played through a pair of Sony MDR V150 headphones.

Procedure

On each trial, eight animals were placed in locations randomly selected from the grid (Figure 2A). The selection of the animals varied by experiment. At the start of a trial, the target animals blinked on and off six times at 1 Hz. There were four targets in Experiments 1–6 and eight in Experiment 7. The animals then began to move. For each animal, a destination cactus was selected at random without replacement. Animals moved toward their destinations in straight lines at different speeds, such that each animal reached its destination at the same time (Figure 2B). Each cycle consisted of 100 monitor refreshes, or 1,333 msec. The maximum speed for an animal was 29.7º sec^{-1} (corner-to-corner movement), and the minimum speed

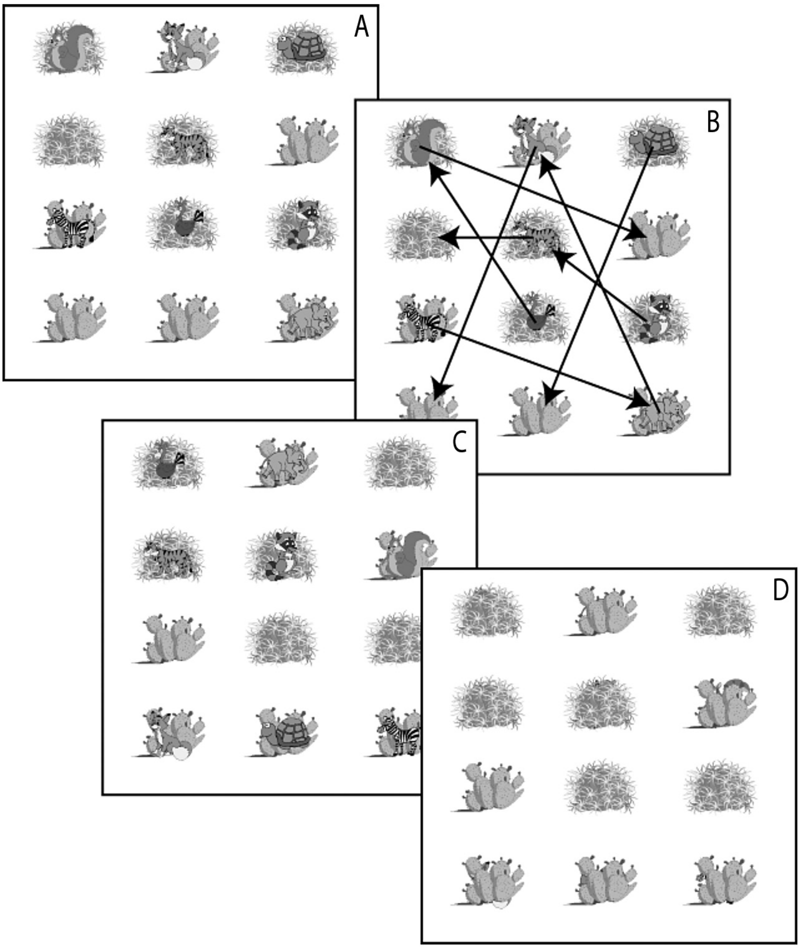

Figure 2. Critical frames from the procedure. Note that although we depict 3 × 4 grids in order to save space, the grids actually used in the experiments were 5 × 4. At the beginning of a cycle, each animal was in front of a cactus (A). Each animal was then assigned a different cactus as a destination (B). Each trajectory was completed in the same amount of time (one 1,333-msec cycle), leading to different speeds for each animal on each cycle. Each animal ended the cycle in front of the destination cactus (C), except on the last cycle, on which the animal was occluded by the cactus (D) for 200 msec. Animals were then removed for a 133-msec interval, followed by the response phase. To see an example of a trial, go to search.bwh.harvard.edu/new/Movies/noah.mov.

was 4.6º sec^{-1} (movement to an adjacent cactus). Once the destinations were reached, another set of destinations was selected. A trial consisted of 10–15 of these cycles (except for Experiment 3, which used 5–15). The number of cycles was selected randomly on each trial. Animals could occlude one another, and moved in front of the background cactus array.

At the end of the last cycle, the cactus array was moved to the front, such that the animals were instantly occluded. Parts of some animals would be visible through transparent portions of the cactus images (Figure 2D). After 200 msec, the occluded animals were erased, eliminating any cues to their identities. After an additional 133 msec, the computer cursor appeared and a probe instruction was presented at the top of the screen.

In the standard condition, the instruction was "Please click on the bushes where all of the target animals are hiding." Observers responded by moving the cursor to click on the appropriate cactuses. The cactus that the cursor was currently over was highlighted with a red border. Once the observer had made a number of responses equal to the number of targets, feedback was presented. If all targets were correctly identified, the message "You got all of them" was printed at the top of the screen. Otherwise, the observer was told how many targets were missed, and the missed targets blinked on and off four times at the appropriate locations.

In the specific condition, the probe instruction was "Where is the <target>?," where <target> was randomly selected from the names of the targets for that trial. The target image was also presented at the upper right of the screen. The observer's task in this condition was to move the cursor to the cactus where the selected target had been at the end of the trial and click. The animal behind the selected cactus was then revealed. Feedback indicated whether the observer had selected the correct cactus, or whether there was no animal behind that cactus.

Observers typically participated in two blocks of 50 trials, each preceded by 5 practice trials. When conditions were blocked, block order was counterbalanced across observers. The experiments typically took between 1 and 1-1/2 h to complete.

Data Analysis

Raw accuracy data were transformed according to high-threshold guessing models devised for each condition. In the standard condition, the guessing model (see Equation 1) was derived from the one presented by Scholl et al. (2001). We assumed that performance P, in terms of the number of targets correctly identified, was determined by the number of objects actually tracked (k, for capacity) plus guessing. In Equation 1, t is the number of targets and a the number of possible response options, in this case, cactuses. The observer makes k informed responses and then guesses on the remaining $(t - k)$ targets, leaving $(t - k)$ targets and $(a - k)$ possible responses. Performance is therefore given by

$$P = k + \frac{(t-k)^2}{a-k}; \tag{1}$$

solving for k gives

$$k = \frac{aP - t^2}{a + P - 2t}. \tag{2}$$

Performance in the specific condition was modeled according to similar logic. Here, performance p is the probability of correctly selecting the cactus where the specific target is hiding. The probability that the probe animal (selected at random from the t targets) will be one of the k objects being tracked is k/t. So on k/t trials, performance is 1.0. On the remaining $(1 - k/t)$ trials, the observer guesses. In this case, the observer can eliminate the k locations with tracked animals because she knows they contain an animal that was not probed, so she guesses from among the remaining $(a - k)$ locations. Hence, performance is given by Equation 3:

$$p = \frac{k}{t} + \left(1 - \frac{k}{t}\right)\left(\frac{1}{a-k}\right). \tag{3}$$

When we again solve for k,

$$k = \frac{a + pt - 1 - \sqrt{(1-a-pt)^2 - 4(apt - t)}}{2}. \tag{4}$$

Equations 2 and 4 convert performance in both conditions to a common metric, k.

The raw data underlying the capacity computations for all experiments are reported in Appendix B. The majority of our analyses were single-degree-of-freedom planned comparisons carried out through paired t tests.

EXPERIMENT 1

In this experiment, we introduce the paradigm and the basic result, the content deficit. In this version of the experiment, the set of 8 unique animals to be presented on each trial was selected at random from the 23 available animals. A subset of 4 target animals was then selected at random from among these 8. No attempt was made to control overlap in sets from trial to trial. The standard and specific conditions were run in separate blocks.

Estimated capacity was 1.6 items ($SEM = 0.3$ items) in the specific condition and 2.9 items ($SEM = 0.3$) in the standard condition. There were two notable results in this experiment. First, observers were reliably able to report on the identity of more than one item in the specific condition [$t(7) = 5.0$, $p < .001$, one-tailed], indicating a content-addressable representation. Second, estimated capacity was lower in the specific condition than in the standard condition [$t(7) = 8.0, p < .0001$, two-tailed]. That is, when asked to report the location of a specific target, observers appeared to have tracked fewer items than when they were asked to simply click on all of the target locations. We will call this result the *content deficit*. Subsequent experiments will test possible explanations for this deficit.

One puzzling aspect of these data was that performance in the standard condition seemed low in this experiment, in comparison with typical MOT experiments with identical items. Observers are typically able to track more than three items (see Oksama & Hyönä, 2004; Scholl et al., 2001). Our task may have been more difficult simply because the animals often occluded one another, which can reduce performance (Klieger, Horowitz, & Wolfe, 2004). There is also a more serious possibility: Since the stimulus set changed from trial to trial, observers may have experienced interference in working memory. In particular, items that were targets on one trial could have served as nontargets on subsequent trials, and vice versa; such conditions have led to interference in both the visual search (e.g., "variable mapping"; Schneider & Shiffrin, 1977) and negative priming (Milliken, Tipper, & Weaver, 1994; Tipper, 1985) paradigms. Such interference might be particularly problematic in the specific condition. Thus, the changing stimulus sets might be responsible for both the content addressability and the content deficit from the data. Experiment 2 was designed to remedy this problem.

EXPERIMENT 2

In this experiment, we simply fixed the target and nontarget sets so that the same stimuli were used throughout a block of trials. That is, an observer might track the lion, the rabbit, the turtle, and the goat on all the trials in a block. This made it easier for observers to remember what they were supposed to track. In addition, it eliminated the possibility of interference or negative priming arising from targets and nontargets switching roles. We expected that performance in the standard condition would improve to above 3.0 items; the question was whether the content deficit would still manifest itself under these conditions.

The procedure was identical to that of Experiment 1, except in the selection of animals. For each observer, one set of eight animals was selected randomly at the beginning of each block. Four of these animals were designated as targets and the rest as nontargets. The set of animals and the target/nontarget assignments were held constant for the duration of a block.

As expected, performance in the standard condition improved to a creditable 3.4 (*SEM* = 0.2) items. However, the content deficit was still present. Observers tracked only 2.1 (*SEM* = 0.1) items when asked which target was where. This deficit was reliable [$t(7) = 8.3$, $p < .00001$] and similar in magnitude to that observed in Experiment 1, of around 1.3 items. A mixed ANOVA comparing the two experiments[2] revealed a large effect of condition [$F(1,8) = 115.2$, $p < .000001$] but none of experiment [$F(1,8) = 1.0$, $p > .10$]. The interaction term in the between-experiments analysis was near zero [$F(1,8) < 1$].

The roughly half-item increase in capacity when we changed from a variable to a fixed target set suggests that switching target sets from trial to trial created some sort of interference. This was true not only in the specific condition, when we probed target identities, but also in the standard condition. This result is important, because it suggests a role for object identity (either at the perceptual or semantic level), irrespective of task demands. According to Leonard and Pylyshyn (2003), blinking the targets on and off should automatically assign visual indexes to them, and these indexes should stick equally well whether or not these particular stimuli were targets on the previous trial. Since the increase in capacity was not statistically reliable, this conclusion should not be taken too seriously. However, in subsequent experiments, we used a fixed target set.

Just as in Experiment 1, the observers here could reliably locate multiple targets according to their identities, which we take as evidence for content addressability. Experiment 2 also replicated the content deficit, with reliably better capacity in the standard report condition than in the specific report condition.

These results raise a number of methodological issues. First, our experiments differ from most MOT experiments in their temporal structure; our trials were both longer than usual and included temporal uncertainty. How did these factors affect performance? Experiment 3 was designed to address this issue.

A second important issue is whether blocking response modes induced observers to use different encoding strategies in the two conditions, leading to differential performance. This issue will be addressed in Experiment 4.

A third issue is whether the use of unique, identifiable objects helps or hinders MOT performance in comparison with the use of identical objects in most previous studies of MOT. We will address this issue in Experiments 5 and 6.

Finally, Experiment 7 will address the role of nontargets in hindering tracking performance.

EXPERIMENT 3

As we noted in the introduction, it is important to ensure that observers are attending to a task throughout a trial. One of our strategies we used to ensure this was to randomly vary trial duration. Observers did not know when each trial would end, between the 10th and 15th cycles. Was this enough temporal uncertainty, however? In Experiment 3, we increased the uncertainty so that a trial could end anywhere between the 5th and 15th cycles (a duration from 6.7 to 20.0 sec). Otherwise, the method was identical to that of Experiment 2. If observers were actively tracking the targets only after Cycle 9 in the previous experiments, we should see a decrease in performance in this experiment, where they potentially have to sustain attention for a longer period of time.

A wider array of trial end points also allowed us to examine a possible explanation for the content deficit. We used trial durations that were long relative to previous MOT studies. Perhaps location–identity bindings decay over time, and we would observe a smaller difference between conditions if we probed earlier in the trial.

Figure 3 shows k in Experiment 3 as a function of cycle, with data from Experiment 2 shown for comparison. Mean capacity for the standard condition was 3.4 items (*SEM* = 0.1), versus 2.6 items (*SEM* = 0.2) for the specific condition, representing a significant advantage in the standard condition [$t(7) = 3.9$, $p < .01$]. We also conducted an analysis of covariance on the results, with condition as a categorical variable and cycle as a continuous variable (covariate). Neither the main effect of cycle [$F(1,7) < 1$] nor the cycle × condition interaction [$F(1,7) = 2.49$, $p = .16$] was significant.

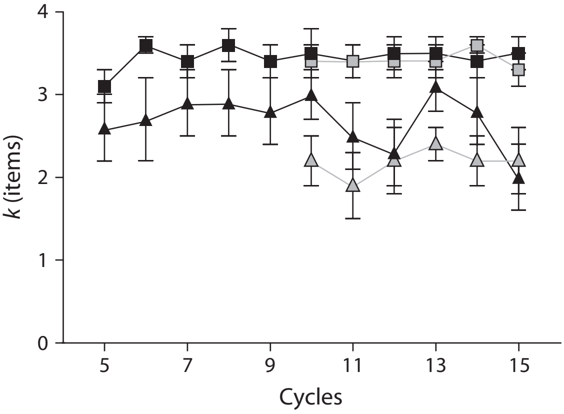

Figure 3. Capacity as a function of cycle. Squares denote the standard and triangles the specific target condition. Data from Experiment 3 are plotted in black, and those from Experiment 2 in gray. Error bars indicate standard errors of the means.

When comparing the two experiments, it seems clear that increasing the temporal variability in Experiment 3 did not change the results notably. We conducted a cross-experiment ANOVA, using data only from Cycles 10–15 in Experiment 3. Standard versus specific target condition and cycle were within-subjects factors, and experiment was a between-subjects factor. We found no main effect of experiment [$F(1,14) = 1.1, p = .31$]. The condition × experiment interaction approached significance [$F(1,14) = 4.1, p = .06$], probably because the specific target capacity was somewhat higher in Experiment 3. Adding temporal uncertainty clearly did not disrupt performance in this experiment in relation to Experiment 2.

There is a hint in Figure 3 that performance, especially in the specific target condition, dropped as trial duration increased. This would be compatible with data from Oksama and Hyönä (2004), possibly reflecting a vigilance decrement (Parasuraman, 1986). If the drop over time were greater in the specific condition, this would support the idea that location–identity bindings were decaying. However, there was no evidence for this in an ANOVA conducted on data from all of the conditions of Experiment 3, in which neither the main effect of cycle [$F(10,70) = 1.1, p = .38$] nor the cycle × condition interaction [$F(10,70) < 1$] was significant. If there was decay, it had approached asymptote within the first 6–7 sec of tracking.

EXPERIMENT 4

One obvious hypothesis to explain the content deficit is that observers encode information differently in the two conditions. In the standard condition, only spatiotemporal properties are encoded. In the specific target condition, observers know that they will be asked about featural or identity information, so they encode that as well. Featural information (or perhaps the binding of identity and location) takes up more resources (Bahrami, 2003; Saiki, 2002), reducing tracking capacity. Thus, information about features or identities requires more capacity.

In Experiment 4, we tested this strategic encoding hypothesis by mixing the questions within a block of trials. Observers did not know on a given trial whether they would be asked to locate all of the targets or just a specific target. Therefore, they had to adopt the same strategies on both types of trials. If the strategic encoding hypothesis is correct, then the content deficit should be reduced in this experiment. Performance should also be reduced overall, since observers would not be optimally encoding information on each trial. In other respects, the method was identical to that of Experiment 2.

Pilot work indicated that when questions were mixed within a block of trials, having to look away from the cactus array to read the question at the top of the screen impaired performance. Therefore, in this experiment, we used auditory rather than visual probes. Instead of printing the question at the top of the screen, observers heard the probe questions over headphones.

Mean capacity for the standard condition was 3.4 items ($SEM = 0.1$), as opposed to 2.2 items ($SEM = 0.2$) for the specific condition, representing a significant advantage for the standard condition [$t(7) = 9.9, p < .00005$]. Performance in this experiment, with questions mixed within blocks, was nearly identical to that in Experiment 2, with blocked questions. Even when observers could not have adopted different encoding strategies for the different conditions, we still observed a sizable (1.2-item) content deficit. This suggests that the content deficit is not a result of the strategic demands of the conditions we devised.

EXPERIMENT 5

The previous experiments established a paradigm for studying MOT with unique stimuli. One question that we had not yet addressed is whether unique objects help or hinder MOT performance in comparison with identical objects. For example, if the objects are identical, an observer might accidentally swap a target and a distractor if they approach too closely. With unique objects, this is less likely to happen. Similarly, if at some point the observer realized that he was only tracking three targets and had lost one, he could theoretically recover the lost target if it were unique. Of course, such advantages assume a sort of "ideal observer" who always knows exactly what he is tracking. If the tracking system was completely insensitive to target features or identity, unique objects would provide no advantage. Similarly, the observer could only recover a lost target if information were available about how many targets were currently being tracked and which one was missing.

In this experiment, we compared tracking unique objects with tracking identical objects. Since "Where is the zebra?" is not a diagnostic question when all objects are zebras, we only used the standard response mode. There were three conditions. The *unique* condition was identical to the standard response condition in Experiment 1. In the *identical* condition, eight identical animals were presented; animal identity was selected at random on each trial. Finally, in the *paired* condition, there were four unique targets, again selected at random. For each target, there was an identical distractor. Probes were presented auditorily, as in Experiment 4. A total of 10 observers participated in this experiment.

We can compare performance in the unique and identical object conditions to determine whether there is any advantage for unique objects. If the advantage for unique objects derives from the ability to recover lost targets, this advantage should be abolished in the paired condition.

Capacity values for the three conditions are plotted in Figure 4. Observers tracked 3.1 items ($SEM = 0.2$) in the unique condition, 2.3 ($SEM = 0.2$) in the paired condition, and 2.5 ($SEM = 0.1$) in the identical condition. Planned t tests showed that the paired and identical conditions did not differ [$t(9) = 1.3, p > .10$], but that performance was significantly worse in both than in the unique condition [for paired, $t(9) = 2.7, p < .05$; for identical, $t(9) = 2.7, p < .01$].

These results indicate that observers can take advantage of the additional information provided by unique objects. The unique object advantage (approximately 0.6 items

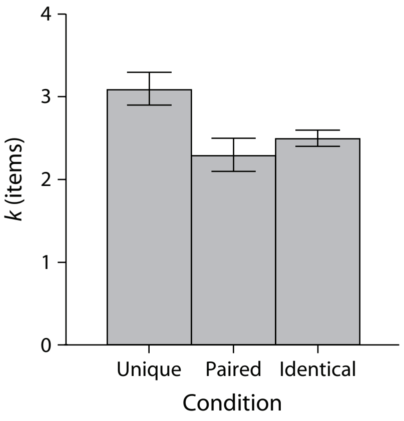

Figure 4. Capacity as a function of condition in Experiment 5.

under these conditions) is entirely eliminated if targets are identical to distractors, suggesting that some sort of error recovery process is responsible.

EXPERIMENT 6

In the previous experiment, pairing targets and distractors reduced performance in the standard condition relative to that in the fully unique condition. Would this manipulation affect the specific condition as well? In Experiment 6, we measured the content deficit with both paired and unique stimuli. By comparing these two stimulus conditions, we can determine whether eliminating featural distinctions between target and nontarget sets impairs the ability to track specific objects as well as the ability to hold on to target locations. Probes were presented auditorily, as in Experiment 5. There were 6 observers in this experiment.

As shown in Figure 5, we once again replicated the content deficit in this experiment, since observers could track 1.1 more items in the standard condition than in the specific condition [$F(1,9) = 119.0, p < .001$]. Pairing targets and nontargets reduced capacity by roughly the same amount as in Experiment 5 [0.9 items; $F(1,9) = 37.3$, $p < .001$]. Critically, the pairing manipulation appeared to have less of a deleterious effect in the specific condition than in the standard condition [interaction $F(1,9) = 18.5$, $p < .005$]. In the standard condition, the pairing manipulation cost observers 1.2 items, versus a loss of only 0.7 items in the specific condition. This interaction suggests that different representations may underlie performance in the two conditions.

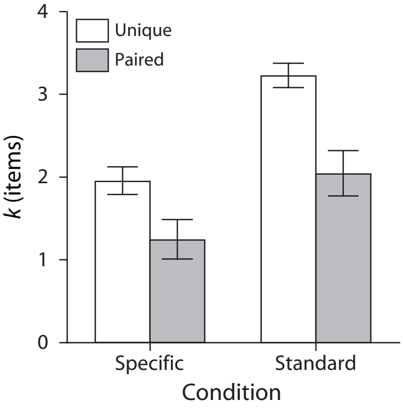

Figure 5. Effect of pairing target and nontarget objects on capacity in Experiment 6. Data from the paired conditions are plotted as gray bars, and those from the unique conditions as open bars.

On the other hand, capacity was already lower in the specific condition. If the effect of pairing stimuli is multiplicative, rather than additive, we would then expect that the impact of paired stimuli would be proportionally smaller in the specific condition. When we log-transformed the capacity values, converting proportional differences into additive ones, the question × pairing interaction was eliminated [$F(1,9) < 1$].

Since we cannot reject the hypothesis that the effect of pairing targets with nontargets is proportional to capacity, this experiment provides at best weak evidence for different representations.

EXPERIMENT 7

One advantage of our method is that we can test MOT performance in a case with only targets. Why would we want to do this? There are reasons to believe that suppressing nontargets consumes some resources that would otherwise be available for tracking targets. For example, there is evidence that observers track better on trials in which nontarget trajectories are repeated from earlier trials (Ogawa & Yagi, 2003). If nontargets are processed, they may have to be suppressed in order to reduce confusion with targets, and Pylyshyn and Leonard (2003) showed evidence for such inhibition. Moreover, unpublished data from our laboratory indicate that MOT performance decreases as nontargets are added to the display, even when the number of targets remains constant.

Accordingly, we designed Experiment 7 to test the hypothesis that tracking capacity would increase if there were no nontargets. This experiment was identical to Experiment 2, except that all items were targets. Capacity in the standard condition was 5.8 items ($SEM = 0.2$), significantly greater than the 2.6 items ($SEM = 0.1$) obtained in the specific condition [$t(7) = 14.3, p < .000005$].

In agreement with our hypothesis, capacity improved markedly in the standard condition when there were no distractors. Standard capacity in Experiment 7 was significantly greater than in Experiments 2, 3, and 4, which were methodologically comparable except for the number of targets and nontargets (all ps $< .00005$).

One explanation is that previous experiments underestimated capacity because of a ceiling effect. These experiments were based on the hypothesis that capacity varies across trials, with no underlying differences across experiments. However, the measured capacity has been limited by the number of targets. For example, if there are only four targets (e.g., Experiments 2–4), observers could track all of them on trials in which their capacity was 4 or greater; any trials with an actual capacity greater than 4 would thus appear to have a capacity of 4. In this case, measured mean capacity would underestimate the true mean. In Experiment 7, which included eight targets, measured capacity could reflect actual capacity up to eight items, leading to a greater (and more accurate) estimate of mean capacity.

We can test this ceiling hypothesis against our hypothesis of increased mean capacity by comparing the capacity distributions in Experiment 7 with those in Experiments 2–4. According to the ceiling hypothesis, the proportion of trials with a capacity less than 4 should be identical across experiments. Furthermore, the proportion of trials with a capacity of 4 or greater in Experiment 7 should be the same as the proportion of trials with a capacity equal to 4 in Experiments 2–4. In contrast, the hypothesis that tracking capacity increases when no nontargets are present predicts that the entire distribution should shift to the right, leading to fewer trials with capacity less than 4 in Experiment 7 than in the earlier experiments.

We tested these hypotheses by computing capacity for each trial, using Equation 2, and generating a distribution across trials for each observer. Figure 6 plots the distribution of k for Experiment 7 and the corresponding distribution for Experiments 2, 3, and 4 combined (distributions from each of these experiments were averaged across observers). Clearly, the proportion of trials with an observed capacity less than 4 was lower with eight targets than with four targets. This is consistent with the capacity distribution shifting to the right in Experiment 7, rather than mere spreading of the $k = 4$ trials over a wider range. We suggest that the superior performance in this experiment may reflect additional capacity freed up when there is no need to suppress nontargets. However, the shape of the distribution for four targets does suggest that capacity was underestimated in the previous experiments.

The conclusions are quite different if we look at the specific target condition. Specific target capacity in this experiment was better than in Experiment 2 (uncorrected t test, $p < .05$), but not significantly different from the capacities observed in Experiments 3 and 4 ($p > .05$). Thus, the tracking of target identities in this condition does not seem to be limited by the need to suppress nontarget identities.

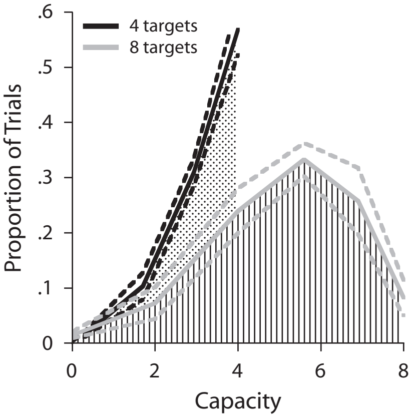

Figure 6. Capacity distributions for the standard condition in Experiment 7 (gray lines, striped area) and in Experiments 2, 3, and 4 (black lines, stippled area). Dashed lines indicate standard errors of the means.

DISCUSSION

We have developed a new MOT paradigm that allows us to test the tracking of unique objects. Four main findings have emerged from these experiments. First, there is a content-addressable representation of targets during MOT (specific capacity was >1 in all cases). Second, we observed a content deficit: The content-addressable representation has a lower capacity (measured in items) than the location-addressable representation (i.e., specific capacity was lower than standard in all cases). Third, the visual system is capable of taking advantage of differences between unique objects to improve tracking performance over the level seen with identical objects (Experiment 5). This advantage appears to be due primarily to the ability to recover lost targets (Experiment 5). Fourth, tracking capacity appears to be constrained by the need to suppress nontargets; when there are no nontargets, capacity increases markedly (Experiment 7).

A Content-Addressable Representation in MOT

Our data represent an existence proof of a content-addressable representation in MOT. Similar data have been reported by Oksama and Hyönä (2004). In their multiple-identity-tracking (MIT) task, observers tracked moving line drawings (either objects or "pseudo-objects") for several seconds. At the end of the trial, the stimuli were masked and a single item was marked with a black frame. In a subsequent response screen, observers had to select the picture that corresponded to the marked object. Oksama and Hyönä obtained a median capacity of four items.

Although it is difficult to compare capacity measures directly across experimental tasks,[3] it is interesting that both we and Oksama and Hyönä (2004) demonstrated that observers do know, to some extent, which target is which. This finding is in contrast to those of Pylyshyn (2004), whose observers had difficulty identifying targets. Two key differences between our paradigm and MIT, on the one hand, and Pylyshyn's (2004) paradigm, on the other, seem relevant to this discrepancy. First, in our study and that of Oksama and Hyönä, object identities were defined by intrinsic features, such as shape and (in our experiment) color. In Pylyshyn's (2004) experiment, "identity" was based on a tangential feature of an object, such as in which corner of the display it began the trial. It seems likely that the shape and color of an object may be more tightly bound markers of "identity" than initial position or a briefly presented letter.

Second, in our experiments and those of Oksama and Hyönä (2004), the visual differences between objects were continually available throughout the tracking phase of the trial, only masked during the response phase. In

Pylyshyn's (2004) experiments, the object identities were presented only briefly at the start of the trial; during the majority of the tracking phase, the objects were physically identical. It may be that the link between object identity and spatiotemporal position decayed during Pylyshyn's (2004) task but was constantly refreshed in both ours and Oksama and Hyönä's.

Our experiments also provided some indirect evidence for a content-addressable representation. For example, performance was consistently superior in experiments in which the same objects were used in the same roles from trial to trial (Experiments 2, 3, and 4) than in experiments in which the stimulus set changed from trial to trial (Experiments 1, 5, and 6). This occurred not only in the specific target conditions, in which object identity was relevant, but also in the standard conditions, in which object identity was not.

One System or Two?

Is tracking served by a single representation that encodes both position and identity, or are there two representations, one for position and one for identity? This question arises because of the striking capacity differences between our response conditions. In all of our experiments, capacity was significantly lower for specific targets than for targets in the standard MOT response mode. In Experiment 4, we disconfirmed the hypothesis that capacity differences resulted from varied encoding strategies (see, e.g., Davis, Welch, Holmes, & Shepherd, 2001), since we obtained the same result when response modes were mixed within a block as we did when they were run in separate blocks.

This capacity difference is not conclusive evidence that separate representations or systems are involved. There are several ways in which a unified account could explain the lower capacity in the specific target condition. For example, one might propose a system with a slot-like structure (Vogel, Woodman, & Luck, 2001) in which all slots are not created equal. If there were two slots with sufficient resolution to encode both position and identity, and two slots that could only encode position, we would also obtain different capacity estimates in our two response conditions. This representation could also be described in terms of Kahneman and Treisman's (1984) object files, if we assume that some of the object files are empty. The impletion process might in that case return the information that a given object was spatiotemporally continuous with a target object, but fail to retrieve the semantic or perceptual content associated with that object file.

In the MOT literature, *capacity* is generally assumed to refer to a fixed number of available representations in the visual system, either visual indexes or object files. However, there is some evidence that capacity in MOT might be better thought of as an undifferentiated resource (Alvarez & Cavanagh, 2005). Instead of a set of slots or pointers, we might assume that there is a fixed amount of information that can be encoded. Observers might encode position and identity for all targets, but the binding between positions and identity may be noisier than the position information itself (or may decay more quickly; Experiment 3 indicated that any such decay would have to approach its asymptote value within 6–7 sec), leading to apparently lower capacity when responses depend on binding.

Instead of one system, there also might be two. Although we tend to assume that a single experimental paradigm probes the performance of a single corresponding functional system, it may be that MOT naturally recruits multiple overlapping systems. The simplest two-system account might be a proposal that tracking is served by a position-tracking system, such as visual indexes (FINSTs), but that accessing identity–position bindings requires focal attention (Wheeler & Treisman, 2002). However, since specific capacity is reliably greater than one item, we would have to assume that multiple items can simultaneously be the focus of attention (Awh & Pashler, 2000; Davis et al., 2001).

If we wish to preserve a single focus of attention, we might instead propose that both visual indexes and object files can be used. Visual indexes would provide only position information, and object files could support more complex queries.

It will obviously be difficult to definitively decide between one and two systems. However, the evidence of our final three experiments points to two independent systems. In Experiments 5 and 6, pairing targets and nontargets reduced capacity in the standard but not in the specific target condition (although we could not reject the hypothesis of a proportional decrease). In Experiment 7, standard capacity increased when no nontargets were presented, but specific capacity was unaffected.

MOT has been likened to a visual working memory task at time zero (Cavanagh & Alvarez, 2005), so it is natural to look to studies of working memory for analogs of the content-addressable and location-only systems. The classic two-pathway scheme in vision is the distinction between ventral and dorsal streams, associated with object recognition and spatial processing, respectively (Ungerleider & Mishkin, 1982). This distinction has been extended into the domain of working memory by Wilson, Ó Scalaidhe, and Goldman-Rakic (1993), who provided evidence that object information and spatial information are neurally segregated in monkey prefrontal cortex, widely believed to be the neural substrate of working memory. A similar dissociation in humans was reported using PET imaging by Smith et al. (1995).

The division of working memory into multiple subsystems goes back at least to the multiple-component model of Baddeley and Hitch (1974), which proposed separate components for different modalities: a "phonological loop" for auditory information and a "visuospatial sketch pad" for visual information. In more recent versions of the model (Baddeley & Logie, 1999; Logie, 1995), the sketch pad has been divided into visual and spatial subcomponents, roughly corresponding to the object and spatial systems described by Wilson et al. (1993). Interestingly, the spatial subcomponent (the "inner scribe") is specifically tasked with holding sequence or path information (see Parmentier, Elford, & Mayberry, 2005), making it a likely substrate for MOT performance.

The widespread insistence on a separation between visual or object information and spatial information would seem problematic for our findings. Standard MOT is clearly a job for the spatial subsystem (Postle, D'Esposito, & Corkin, 2005, for example, used a version of standard MOT to tie up dorsal processing in a memory interference paradigm). However, there does not seem to be room in that account for a content-addressable representation. Our specific target task (as well as the MIT task of Oksama & Hyönä, 2004) is not a pure object recognition task; it requires location–identity bindings, which would seem to involve integrating information across streams. The problem disappears if we take into account recent behavioral (Olson & Marshuetz, 2005) and fMRI (Postle & D'Esposito, 1999) investigations demonstrating that object working memory representations also carry some location information.

Neuroimaging studies might prove useful here. Standard MOT has been shown to activate parietal lobe areas in the dorsal stream (Culham et al., 1998; Jovicich et al., 2001; Seiffert, 2003). It would be informative to know whether the pattern of activation changes when unique objects are introduced or when the task potentially involves object identities.

Unique Object Features in MOT

In addition to our findings about content addressability in tracking, we were able to demonstrate that having visually unique objects made tracking easier; observers were able to track roughly 0.6 more items when objects were unique than when objects were identical. In Experiments 5 and 6, we attributed this result to an error recovery strategy: If I lose track of the zebra, I can search for the zebra on the basis of its physical features and continue to track it. This strategy would obviously be useless if all items are zebras. This strategy is of limited utility if there are a zebra target and a zebra distractor; once I lose the zebra, I then have only a 50% chance of picking up the target zebra, as opposed to the distractor.

The error recovery hypothesis is not trivial. It has several important implications. First, it implies that observers know what they are *not* tracking. The phenomenology of MOT is such that an observer has the feeling of knowing how many targets he or she is successfully tracking. The error recovery hypothesis implies that when a target is lost, the observer knows which one (given unique targets). Second, the error recovery hypothesis also assumes that an observer can search for a missing target while continuing to track the remaining targets. This is in agreement with recent evidence that observers can successfully perform an MOT task and a visual search task on the same trial (Alvarez, Horowitz, Arsenio, DiMase, & Wolfe, 2005). Finally, the hypothesis implies the ability to add to the tracking set on the fly, while tracking, without any physical cue. These implications need to be confirmed through further research.

However, if observers use unique physical features of objects to recover lost targets in this fashion, it is curious that there is no such benefit in the specific target condition. If observers do recover lost targets during the course of a trial, they appear to recover only their locations, not their identities.

An alternative to the error recovery hypothesis, though not a mutually exclusive one, is that the results of Experiment 5 can be explained by the perceptual similarity of target and nontarget items. Yantis's (1992) approach to MOT highlights the importance of perceptual grouping factors that allow the targets to be treated as a unit by the visual system. Most of the research in this line has focused on spatiotemporal grouping factors, such as the spatial relationships between the objects (Sears & Pylyshyn, 2000; Viswanathan & Mingolla, 2002; Yantis, 1992).

Can grouping by (nonspatial) featural properties improve tracking? A number of studies have demonstrated that observers have limited explicit access to featural properties (Bahrami, 2003; Saiki, 2002; Scholl et al., 1999), but this finding does not address the direct role that features may play. We could explain the results of Experiment 5 by proposing that observers were able to use accidental differences in the shape and color statistics of targets and nontargets to help segregate them visually. In the paired condition, targets and nontargets had identical feature statistics, so that advantage was lost. This hypothesis could also explain why performance was better with a fixed stimulus set than with a variable stimulus set; over a number of trials with the same stimuli, observers learn subtle differences between the two sets.

Conclusions

In the introduction, we noted that real-world dynamic attention tasks differ from traditional MOT in two ways. First, observers in the everyday world usually deal with multiple unique objects. Second, they often wish to know where a given target might be, rather than which objects are the targets. The experiments presented here represent an attempt to bridge laboratory MOT research and these everyday cognitive tasks. We have shown that it is easier to track unique objects than to track identical objects, assuming that the objects one wishes to track are also uniquely different from nontarget objects. Furthermore, observers do know what they are tracking. To take one of our examples of ecologically valid tracking tasks, if you watch your children and their friends on the playground, you can be aware of where your kids are at any given moment; you may also be aware of the locations of their friends, without being able to tell one friend from the other. Furthermore, if you lose track of a child, you will probably be aware of which one you have lost.

AUTHOR NOTE

This research was supported by NIMH Grant R01 65576 to T.S.H. We thank Kristin Michod and Skyler Place for assistance with data collection, as well as Steve Luck and an anonymous reviewer for constructive criticism. Correspondence relating to this article may be sent to T. S. Horowitz, Brigham and Women's Hospital, Visual Attention Laboratory, 64 Sidney Street, Suite 170, Boston, MA 02139 (e-mail: toddh@search.bwh.harvard.edu).

REFERENCES

Allen, R., McGeorge, P., Pearson, D., & Milne, A. B. (2004). Attention and expertise in multiple target tracking. *Applied Cognitive Psychology*, **18**, 337-347.

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., & Franconeri, S. (2004, November). *How many objects can you track?* Paper presented at the 12th Annual Workshop on Object Perception & Memory, Minneapolis, MN.

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 & Performance*, **31**, 643-667.

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 780-790.

Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 834-846.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-90). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). Cambridge: Cambridge University Press.

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition*, **10**, 949-963.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, **9**, 349-354.

Culham, J. C., Brandt, S. A., 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., Welch, V. L., Holmes, A., & Shepherd, A. (2001). Can attention select only a fixed number of objects at a time? *Perception*, **30**, 1227-1248.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics*, **65**, 58-71.

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.

Jovicich, J., Peters, R. J., Koch, C., Braun, J., Chang, L., & Ernst, T. (2001). Brain areas specific for attentional load in a motion-tracking task. *Journal of Cognitive Neuroscience*, **13**, 1048-1058.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29-61). Orlando: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, **24**, 175-219.

Klieger, S. B., Horowitz, T. S., & Wolfe, J. M. (2004). Is multiple object tracking colorblind? [Abstract]. *Journal of Vision*, **4**(8), 363a.

Leonard, C., & Pylyshyn, Z. W. (2003). Measuring the attentional demand of multiple object tracking (MOT) [Abstract]. *Journal of Vision*, **3**(9), 582a.

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, U.K.: Erlbaum.

Milliken, B., Tipper, S. P., & Weaver, B. (1994). Negative priming in a spatial localization task: Feature mismatching and distractor inhibition. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 624-646.

Mitroff, S. R., Scholl, B. J., & Wynn, K. (2004). Divide and conquer: How object files adapt when a persisting object splits into two. *Psychological Science*, **15**, 420-425.

Ogawa, H., & Yagi, A. (2003). Priming effects in multiple object tracking: An implicit encoding based on global spatiotemporal information [Abstract]. *Journal of Vision*, **3**(9), 339a.

Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition*, **11**, 631-671.

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in real world scenes [Abstract]. *Journal of Vision*, **3**(9), 3a.

Olson, I. R., & Marshuetz, C. (2005). Remembering "what" brings along "where" in visual working memory. *Perception & Psychophysics*, **67**, 185-194.

Parasuraman, R. (1986). Vigilance, monitoring, and search. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 1-39). New York: Wiley.

Parmentier, F. B. R., Elford, G., & Mayberry, M. (2005). Transitional information in spatial serial memory: Path characteristics affect recall performance. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 412-427.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442.

Postle, B. R., & D'Esposito, M. (1999). "What"–then–"where" in visual working memory: An event-related fMRI study. *Journal of Cognitive Neuroscience*, **11**, 585-597.

Postle, B. R., D'Esposito, M., & Corkin, S. (2005). Effects of verbal and nonverbal interference on spatial and object visual working memory. *Memory & Cognition*, **33**, 203-212.

Pylyshyn, Z. [W.] (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, **32**, 65-97.

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, **11**, 801-822.

Pylyshyn, Z. W., & Leonard, C. (2003). Inhibition of nontargets during multiple object tracking (MOT) [Abstract]. *Journal of Vision*, **3**(9), 585a.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, **3**, 179-197.

Saiki, J. (2002). Multiple-object permanence tracking: Limitation in maintenance and transformation of perceptual objects. *Progress in Brain Research*, **140**, 133-148.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, **84**, 1-66.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, **38**, 259-290.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, **80**, 159-177.

Scholl, B. J., Pylyshyn, Z. W., & Franconeri, S. L. (1999, May). *When are spatiotemporal and featural properties encoded as a result of attentional allocation?* Paper presented at the conference of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL.

Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology*, **54**, 1-14.

Seiffert, A. E. (2003). Dissociating neural correlates of attentional tracking and attention to visual motion [Abstract]. *Journal of Vision*, **3**(9), 868a.

Smith, E. E., Jonides, J., Koeppe, R. A., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial versus object working memory: PET investigations. *Journal of Cognitive Neuroscience*, **7**, 337-356.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology*, **37A**, 571-590.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press.

Viswanathan, L., & Mingolla, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking. *Perception*, **31**, 1415-1437.

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 & Performance*, **27**, 92-114.

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

Wilson, F. A., Ó Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, **260**, 1955-1958.

Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research*, **42**, 2985-3004.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, **24**, 295-340.

NOTES

1. We are missing age information for 1 observer.

2. Three observers who participated in both experiments were eliminated from this analysis.

3. Measured capacity can vary based on tracking duration, the range of movement speeds, and whether or not objects can occlude one another (as in our experiment) or bounce off one another (as in Oksama & Hyönä, 2004).

APPENDIX A
Stimulus List

beaver
camel
crocodile
elephant
fox
giraffe
goat
gorilla
horse
kangaroo
lion
moose
ostrich
pelican
rabbit
rhino
rooster
squirrel
tiger
turtle
wolf
zebra

APPENDIX B
Raw Data

Table B1
Raw Numbers of Items Reported in the Standard Condition

		Targets	Nontargets	Total Filled
Experiment 1		3.08		
Experiment 2		3.44	0.14	3.59
Experiment 3		3.47	0.15	3.62
Experiment 4		3.39	0.15	3.54
Experiment 5	Unique	3.12	0.37	3.49
	Paired	2.49	0.80	3.29
	Identical	2.61	0.86	3.47
Experiment 6		2.48	0.78	3.26
Experiment 7		5.99		5.99

Note—*Targets* indicates the average number of correctly reported targets per trial, *Nontargets* the average number of times observers clicked on a nontarget location per trial, and *Total Filled* the sum of these two categories. Nontarget click counts were not collected for Experiment 1, and there were no nontargets in Experiment 7.

Table B2
Raw Proportions for the Specific Condition, by Experiment

		Errors				
	Correct Responses	Transposition	Nontarget	Empty Space	*N*	Trials per Observer
Experiment 1	.44	.14	.16	.26	8	50.00
Experiment 2	.56	.18	.10	.16	8	50.00
Experiment 3	.68	.12	.10	.10	8	50.00
Experiment 4	.58	.25	.09	.08	8	50.63
Experiment 6	.41	.14	.21	.23	7	50.00
Experiment 7	.33	.39	.00	.28	8	50.00

Note—Transposition errors refer to trials in which the observer clicked on the wrong target location. Trials per observer varied in Experiment 4 because condition was selected randomly from trial to trial.

(Manuscript received March 31, 2005;
revision accepted for publication February 13, 2006.)