Psychonomic Bulletin & Review
2007, 14 (2), 344-349

Multiple object juggling: Changing what is tracked during extended multiple object tracking

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

Skyler S. Place
Brigham and Women's Hospital, Boston, Massachusetts

and

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

The multiple object tracking (MOT) task has been a useful tool for studying the deployment of limited-capacity visual resources over time. Since it involves sustained attention to multiple objects, this task is a promising model for real-world visual cognition. However, real-world tasks differ in two critical ways from standard laboratory MOT designs. First, in real-world tracking, it is unusual for the set of tracked items to be identified all at once and to remain unchanged over time. Second, real-world tracking tasks may need to be sustained over a period of minutes, and not mere seconds. How well is MOT performance maintained over extended periods of time? In four experiments, we demonstrate that observers can dynamically "juggle" objects in and out of the tracked set with little apparent cost, and can sustain this performance for up to 10 min at a time. This performance requires implicit or explicit feedback. In the absence of feedback, performance tracking drops steadily over the course of several minutes.

Since its introduction, almost 20 years ago (Pylyshyn & Storm, 1988), the multiple object tracking (MOT) task has been a useful tool for studying the deployment of limited-capacity visual resources over time (Scholl, 2001). In a typical MOT task, several identical items are present in a visual display. A subset of these is cued in some fashion (e.g., by blinking on and off). Then, when all of the items are identical again, the items begin to move around in the display. The observer's task is to track the cued set. Since cued and uncued items are indistinguishable, tracking must be done by somehow "paying attention" to the cued items. After a few seconds of tracking, the objects stop, and the observers are asked either to indicate the entire tracked set or to state whether a specific item was or was not part of that set. Many variables influence performance (speed, trajectory, distance between objects, and so forth). Under standard laboratory conditions, a typical observer can track four out of eight objects with high accuracy.

When called upon to explain why MOT is a phenomenon worthy of study, researchers reach for examples in the real world. The world, we note, is filled with multiple objects of interest. Under conditions such as navigating in traffic, playing soccer, or monitoring children on a playground, several objects of interest may be moving on independent tracks. Of course, there are many differences between a collection of children at recess and a collection of identical disks on the screen of a computer monitor. For the present purposes, we are interested in two of these.

First, in real-world tracking, it is unusual for the set of tracked items to be identified all at once and to remain unchanged for the duration of the task. Say you are driving down the highway, tracking a couple of seemingly relevant nearby cars. A cat on the side of the road looks as if it might be thinking about entering traffic. One car slows down and falls back, relative to your position. A truck merges in from the right. Can MOT performance be maintained if the tracked set changes over the course of a single tracking episode? Second, real-world tracking tasks may need to be sustained over a period of minutes, and not mere seconds. How well is MOT performance maintained over extended periods?

This article presents four experiments addressing these issues. In Experiment 1, we show that observers are capable of *multiple object juggling*, a version of MOT in which objects are added and subtracted from the tracked set over the course of a 20-sec tracking episode. In Experiment 2, we extended the tracking episode to 10 min, and demon-

J. M. Wolfe, wolfe@search.bwh.harvard.edu

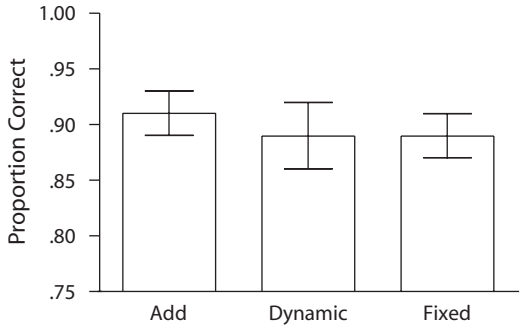

Figure 1. Average accuracy for the three conditions of Experiment 1. Error bars are ±1 *SEM*.

strated that observers could maintain their performance over that period. Experiment 3 shows that although performance degrades over time if feedback is eliminated, it does not fall to chance, even after 10 min. Finally, Experiment 4 varied the difficulty of the tracking task to show that the conclusions of Experiments 1–3 were not based on excessively easy versions of MOT.

EXPERIMENT 1
Tracking a Changing Set of Items

General Method

Displays consisted of eight identical gray disks (5.4 cd/m^2, CIE coordinates $x = .340$, $y = .356$). These measured 2º of visual angle in diameter and had a 0.25º-thick yellow border. The objects were contained within a 27.25º outlined square. The objects began each tracking episode arranged on a circle of 15º diameter. Luminance of the black background was .62 cd/m^2. Objects moved at 12º/sec. They followed straight-line motion paths, bouncing off of the walls and each other.

Across all of the experiments, the participants' average age was 29, with a standard deviation of 11. Each participant passed Ishihara's Tests for Colour-Blindness (Ishihara, 1989) and had 20/25 corrected vision or better. All of the observers gave informed consent before participating and were paid for their time. The observers were recruited from the community at large, and were not limited to an undergraduate population. Twelve observers were tested in Experiment 1.

There were three conditions in Experiment 1. The order of the conditions was counterbalanced between observers. For each condition, each observer completed 5 practice and 40 experimental tracking episodes. Each tracking episode consisted of 20 sec of tracking. At the end of a tracking episode, the objects stopped moving and turned solid yellow. The participants were told to click on the target objects. The observers clicked until all four targets were found. When a target object was selected, its yellow border turned green. If a distractor item was selected, its border turned red. Accuracy information was displayed following each tracking episode. The observers were told (for example): "It took you six guesses to find four targets. You were 66% accurate."

In the *fixed* condition, observers performed a standard tracking task. We marked four of the eight objects as targets by having them blink on and off six times: 250 msec on, 250 msec off. All eight objects then began to move.

In the *add* condition, targets were not identified at the beginning of each tracking episode, but instead were marked as targets sequentially, after all of the objects started moving. The objects started in the same circle as in the fixed condition, but began moving immediately. After a disk started moving, we designated it for tracking by briefly changing the yellow border to red. Targets were identified one at a time over the course of 6 sec. The duration of the tracking episode was timed from the designation of the first target.

In the *dynamic* condition, as in the add condition, the tracking episode started with no targets, and we added targets by briefly highlighting an object's border. In this condition, disks could be subtracted from the tracked set, as well as added. The subtract instruction was a large red X, placed over a member of the tracked set as it moved. No objects were ever physically removed from the screen, nor were new objects added. The same 8 objects were merely relabeled as targets or distractors. An object could be added to the set, deleted, and then added again. In this condition, a variable number of targets were in the tracked set at the end of the tracking episode, so observers might have 2, 3, or 4 targets to click on. Over the course of 20 sec, the average number of target disks was 3.0. The number was never allowed to exceed 4.

In both the add and dynamic conditions, addition or subtraction cues lasted for 400 msec. There was always an interval of at least 1,000 msec between any changes in the tracked set. The average interval was 2,000 msec (*SD* 750 msec).

For analysis purposes, we computed accuracy as the proportion of targets identified in the first t clicks, where t denotes the number of targets at the end of that tracking episode. Accuracy values were arc-sine transformed before analysis. We also converted accuracy to an estimate of the number of items actually tracked, using a high-threshold guessing model (Hulleman, 2005). If k is capacity (the number of targets tracked),[1] n the total number of items, t the number of targets in the tracked set, and c the number of targets correctly identified in the first t clicks, then:

$$k = \frac{nc - t^2}{n + c - 2t} \quad (1)$$

We report both accuracy and capacity estimates for all of the experiments.

A special note is required for the *dynamic* condition. In this condition, the number of targets the observer was asked to track varied from tracking episode to tracking episode. Obviously, accuracy for a tracking episode with only two targets is not comparable to accuracy for one with four targets. Thus, it is difficult to directly compare overall accuracy in the *dynamic* conditions to the *fixed* and *add* conditions. Therefore, we have restricted our analysis to those tracking episodes in the *dynamic* condition on which the observer was supposed to be tracking four targets at the end of the episode.

Although capacity estimates are assumed to correct for guessing, they do not allow us to compare performance across different numbers of targets. This is because we can never measure a capacity greater than the number of targets the observer is asked to track. Imagine an observer with a theoretical capacity of six items. In a condition with 50% two-target episodes and 50% four-target episodes, this observer would have an apparent average capacity of less than three items. As with accuracy, therefore, we report capacity only for those episodes in which the observer was asked to track four items. Capacity is provided as an estimate of the effective number of items tracked, and is intended to carry no theoretical baggage.

Results

Mean accuracy is displayed in Figure 1. Accuracy was overall quite high, and did not differ reliably between conditions [ANOVA on arc-sine transformed accuracy: $F(2,22) = 1.733$, $p = .200$]. We also computed tracking capacity (see the General Method section). Capacity was 3.6 (.11) in the *add* condition, 3.0 (.12) in the *fixed* condition, and 3.5 (.14) in the *dynamic* condition.

Discussion

Experiment 1 demonstrated that the tracked set can be defined and modified while the disks are moving. It is not necessary to establish the target set at the start of the tracking episode. The rough equality in performance between

the three conditions of Experiment 1 should not be overemphasized, since the demands of the tasks are somewhat different. In the *fixed* condition, the target set starts and remains as a fixed set of four objects. In the *add* condition, the target set increases to four objects and then those four are tracked for a shorter period of time. In the *dynamic* condition, the average number of items is lower than in the other conditions, though the maximum number of tracked items is the same four items as in the other conditions. Nevertheless, the results of Experiment 1 show that adding and subtracting items from the tracked set does not pose a dramatic challenge to the processes supporting MOT.

EXPERIMENT 2
Sustained Tracking

Experiment 1 established that observers can track a set of targets that change over time. In the real world, it seems likely that tracking would continue for longer than the few seconds typically involved in an MOT task. Experiment 2 investigated the robustness of MOT performance by asking observers to track continuously for 10 min.

The following changes were made from the method of Experiment 1. Disk speed was reduced to 6º/sec. Instead of having observers respond only at the end of each tracking episode, we probed individual disks at intervals; observers had to make a two-alternative forced-choice response as to whether or not a given disk was a target. A probe event, or trial, consisted of one disk turning blue for 400 msec, accompanied by a 650-Hz beep. The observer pressed the "a" key to indicate a target, and the quotation mark key to indicate a nontarget. Auditory feedback was provided, with a low tone (450 Hz) for a correct response and a high tone (1000 Hz) for an incorrect response. Observers were probed 65 times over the course of the 10-min tracking episode. Targets and distractors were probed equally often. In the *dynamic* condition, the number of tracked items never exceeded 4, and averaged 3.2 disks at any given time.

The time between probe events was normally distributed, with a mean of 9.0 sec and a standard deviation of 7.0 sec. In the *dynamic* condition, the interval between add events was normally distributed, with a mean of 9.0 sec and a standard deviation of 9.0 sec; subtract events were similarly distributed. Thus, the mean time between successive add or subtract events was 4.5 sec, with a standard deviation of 4.5 sec. Combined with the probe events (trials), the mean time between any two events in the *dynamic* condition was 3.0 sec, with a standard deviation of 2.8 sec.

Twelve observers each completed two 10-min sessions of each condition in counterbalanced order. Eight observers were new to the paradigm, whereas 4 had been tested in Experiment 1.

As in Experiment 1, we report accuracy as well as capacity, which represents a guess-corrected estimate as to how many disks the observer might have actually been tracking. In the *dynamic* condition, we only computed capacity for trials in which observers were supposed to be tracking four items. This allowed us to use a simplified equation for computing capacity (k) from the proportion of correct responses (p) in two-alternative forced-choice data (Hulleman, 2005) for the case when the number of targets (T) equals the number of distractors:

$$K = T(2p - 1) \quad (2)$$

Results and Discussion

For each observer, we averaged accuracy over three 20-trial bins covering the beginning (Trials 6–25), middle (Trials 26–45), and end (Trials 46–65) of the tracking episode (Figure 2). In Figure 2, all three conditions fluctuate

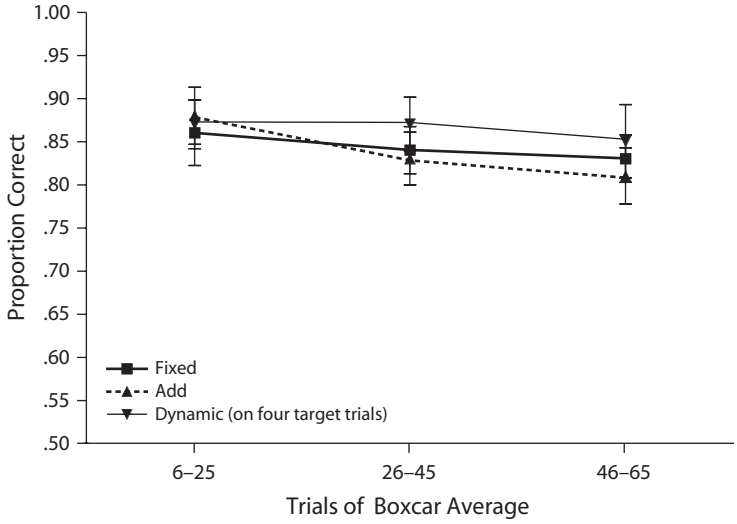

Figure 2. Accuracy as a function of time/trial for the three conditions of Experiment 2. Accuracy is computed for three nonoverlapping ranges of trials. Error bars show ±1 *SEM*.

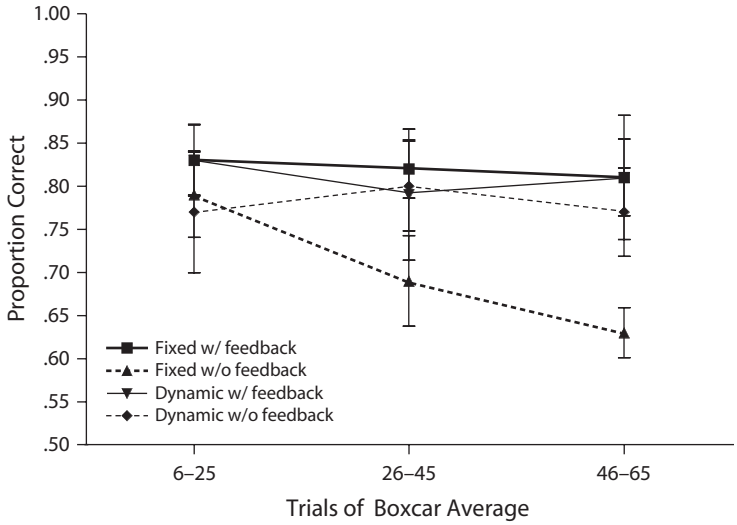

Figure 3. Accuracy as a function of time/trial for the four conditions of Experiment 3 (see Figure 2 for details).

between 80% and 90% accuracy. We performed a two-way ANOVA with three values of condition (*fixed*, *add*, and *dynamic*) and three values of time (the averages for nonoverlapping sets of Trials 6–25, 26–45, and 46–65). Using arc-sine corrected values for accuracy, there was no significant effect of condition [$F(2,22) = 1.121$, $p = .344$] or time [$F(2,22) = 2.460$, $p = .109$]. There was no interaction [$F(4,44) = 1.093$, $p = .372$]. The average capacities were 2.73 (.20) objects in the *add* condition, 2.87 (.21) in the *fixed* condition, and 2.94 (.20) in the *dynamic* condition.

The important conclusion from Experiment 2 is that observers can maintain performance over 10 min of sustained tracking. In the *dynamic* condition, observers appeared to pay a modest cost for the continuous changes in the tracked set. It might seem remarkable that observers could track even a fixed set for 10 min without loss, until one realizes that the auditory feedback gave observers some opportunity to correct errors during the course of the tracking episode (note that this is possible only because observers could update the tracked set without much difficulty, as shown in the *dynamic* conditions). Accordingly, in Experiment 3, we assessed the ability of observers to sustain performance in the absence of feedback.

EXPERIMENT 3
Sustained Tracking Without Feedback

In Experiment 3, observers performed the *fixed* and *dynamic* conditions for 10 min, with and without feedback. The with-feedback conditions replicated Experiment 2. In the without-feedback conditions, the observers heard a tone after each response, but that tone was the same whether the response was correct or incorrect, and served only to let the observers know that the response had been recorded. A total of 12 observers were tested, 2 of whom had participated in previous versions. Each observer performed each condition twice, in counterbalanced order.

Results and Discussion

Accuracy, averaged over the three 20-trial epochs, is plotted in Figure 3. As can be seen in the figure, accuracy remained relatively stable in the *dynamic* conditions and the *fixed* condition with feedback. In the *fixed* condition with no feedback, performance declined steadily over the 10-min session. We analyzed accuracy averaged over three independent epochs, as in Experiment 2. Feedback improved accuracy [$F(1,11) = 8.942$, $p < .05$]. The effect of time on accuracy was not significant [$F(2,22) = 1.711$, $p = .204$], nor was the distinction between dynamic versus fixed displays [$F(2,22) = 1.700$, $p = .219$]. None of the interactions were significant. If we restrict analysis to the *fixed* no-feedback condition, the decline in performance was reliable [$F(2,22) = 10.457$, $p < .001$].

Clearly, the central effect in this experiment is that in the absence of feedback, fixed set-tracking performance declines markedly over time, from a capacity of about 3.0 to 1.5 items over the course of 10 min. This decline in performance could be an interesting dependent measure for those studying sustained operations. For example, one could imagine that this measure might be very sensitive to an observer's state of sleep deprivation or circadian phase (Dinges et al., 1994; Kribbs & Dinges, 1994). It might seem surprising that performance holds relatively steady in the dynamic condition, even with no feedback. However, the dynamic case necessarily supplies a form of feedback, since observers are being asked to add and subtract items from the tracked set every few seconds. Even if an observer gave up on the task for some period of time, he could assemble a new tracked set from the ongoing add and subtract cues when he returned to the task. Thus, Experiment 3 extends the finding from Experiment 2 that

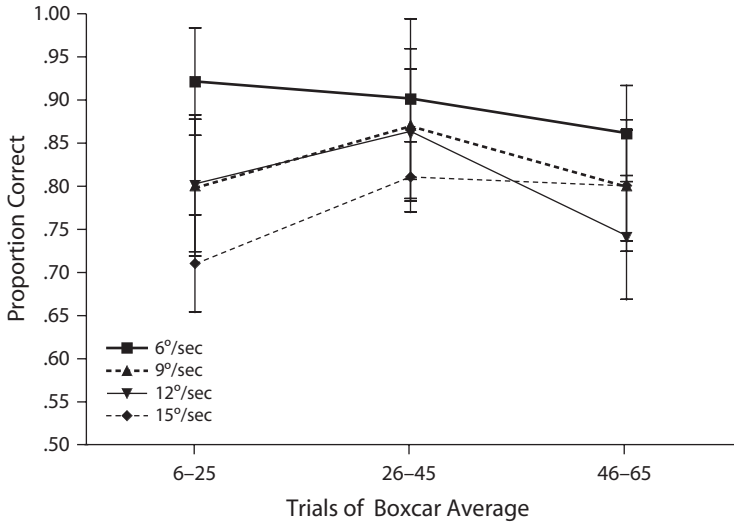

Figure 4. Accuracy as a function of time/trial for the four speeds of Experiment 4 (see Figure 2 for details).

it is possible to continue to add and subtract items over an extended period of time.

EXPERIMENT 4
Tracking at Higher Speeds

Having shown in Experiments 1–3 that observers can juggle items in and out of the tracked set and that they can track for 10 min at a time, in Experiment 4 we attempted to push observers closer to the limits of their capabilities. There are many ways to make a tracking task more difficult. In this case, we increased the speed of the items (Alvarez & Franconeri, 2007).

Experiment 4 replicated the *dynamic* (with-feedback) conditions of Experiments 2 and 3. The only difference was the speed with which the objects moved. In Experiments 2 and 3, objects moved at 6º/sec. In Experiment 4, observers were tested with speeds of 6º, 9º, 12º, and 15º/sec. There were 8 observers, 3 of whom had participated in previous experiments in this set. Each observer ran each condition once, with order counterbalanced across observers.

Accuracy is shown in Figure 4. As expected, accuracy decreased as the speed of the objects increased [$F(3,21) = 3.264$, $p < .05$]. There were no other significant effects with either dependent variable (all ps $> .10$). Average capacities were 2.5 items for 6º/sec, 2.2 items for 9º/sec, 2.0 items for 12º/sec, and 1.8 items for 15º/sec.

Although the task became more difficult with increasing speeds, the observers were still able to track about 1.5 objects at 15º/sec. These data indicate that prolonged multiple object juggling is possible over a range of speeds. Since the average number of tracked items in the *dynamic* conditions was around 3.0, one might imagine that there might be an even greater capacity waiting to be tapped by larger set sizes. However, when we repeated this experiment with larger numbers of tracked and untracked disks, we found either the same or somewhat lower capacities.

GENERAL DISCUSSION

These results show that it is possible to track a dynamically evolving set of targets over many minutes, as might be the case in a real-world tracking task. With appropriate feedback, observers can consistently track multiple items for at least 10 min. In the absence of feedback, however, targets appear to slip from the observer's attentional grasp until, after about 6 min, under our conditions, only a single item appears to be tracked. The feedback that keeps observers "on track" can be explicit information about the accuracy of response or it can be feedback that is inferred from the continual updating of the tracked set in the *dynamic* conditions.

Of course, there are numerous differences between multiple object juggling and the tracking tasks that might face us in the real world. For instance, under most conditions, the items that are being tracked are not identical. Object identity, therefore, can serve as feedback to allow us to dynamically maintain a tracked set (Horowitz, Birnkrant, Fencsik, Tran, & Wolfe, 2006).

In other work, we have shown that observers can suspend tracking for several hundred milliseconds and then return successfully to the task (Alvarez, Horowitz, Arsenio, DiMase, & Wolfe, 2005; Horowitz et al., 2006). Combined with the present findings, these results paint a picture of an ability to deal with the dynamic world in a flexible, if limited, manner. As we make our way through a world containing multiple, independently moving objects, it appears that we can keep track of about three of them. We can update the set of tracked items to suit our current needs and, if needed, we can devote our attention to a dif-

ferent task altogether (e.g., reading the speedometer) and still return to track without having to start anew.

AUTHOR NOTE

Correspondence concerning this article should be addressed to J. M. Wolfe, Visual Attention Lab, 64 Sidney Street, Suite 170, Cambridge, MA 02139 (wolfe@search.bwh.harvard.edu).

REFERENCES

ALVAREZ, G. A., & FRANCONERI, S. L. (2007). *The allocation of visual short-term memory capacity: Evidence for a flexible storage mechanism*. Manuscript submitted for publication.

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.

DINGES, D. F., GILLEN, K. A., POWELL, J. W., CARLIN, M., OTT, G. E., ORNE, E. C., & ORNE, M. T. (1994). Discriminating sleepiness by fatigueability on a psychomotor vigilance task. *Sleep Research*, **23**, 407.

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.

HULLEMAN, J. (2005). The mathematics of multiple object tracking: From proportions correct to number of objects tracked. *Vision Research,* **45**, 2298-2309.

ISHIHARA, S. (1989). *Tests for colour-blindness*. Tokyo: Kanehara.

KRIBBS, N. B., & DINGES, D. [F.] (1994). Vigilance decrement and sleepiness. In R. D. Ogilvie & J. R. Harsh (Eds.), *Sleep onset: Normal and abnormal processes* (pp. 113-125). Washington, D.C.: American Psychological Association.

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

SCHOLL, B. J. (2001). Objects and attention: The state of the art. *Cognition,* **80**, 1-46.

NOTE

1. Here we assume that observers are only tracking targets, not distractors.

(Manuscript received January 14, 2006;
revision accepted for publication June 12, 2006.)