

Multiple-Object Tracking Enhanced Visuospatial Representations as a Result of Experience

Kathryn Barker, Roy Allen, and Peter McGeorge

Vision Research Laboratories, School of Psychology, King's College, University of Aberdeen, UK

Abstract. Research has demonstrated that individuals who routinely engage in complex visuospatial tasks (e.g., radar operators) show an enhanced ability to track multiple randomly moving targets. This study examined tracking expertise using members of a University Officer Training Corps (OTCs) who regularly engage in tasks requiring good dynamic spatial cognition. As expected, the results show that OTCs have enhanced tracking ability relative to other undergraduates. More importantly, they support the idea that, while one set of executive processes are involved in the moment-by-moment updating of the visuospatial representations necessary for dynamic, multiple-object tracking, other processes are activated when whole object sets disappear simultaneously, to create a long-term memory trace of the objects' locations at the moment of their disappearance. Expertise only arose in the former processes, but was lost after a short decay period, such as occurred with a delayed response.

Keywords: multiple-object tracking, visuospatial representations, attention, expertise

The dynamic environment presents our cognitive and perceptual systems with a complex set of challenges, not least of which is the frequent need to monitor several moving objects simultaneously. Such “multiple-object tracking” (MOT) arises as part of a host of everyday activities ranging from engaging in sports through to navigating the aisles of a crowded supermarket. In order to examine the cognitive and perceptual processes underpinning MOT, Pylyshyn and Storm (1988) developed a laboratory paradigm in which participants were presented with a set of 10 identical objects. A number of these objects were then identified as a “target” subset before all the objects began to move randomly about the screen. Immediately following a variable duration period of movement, several objects were probed sequentially, and participants had to respond only when the probed item was a target. Using this and similar paradigms, research has demonstrated consistently that, in general, people have the capacity to track up to four targets (e.g., Allen, McGeorge, Pearson, & Milne, 2006; Pylyshyn & Storm, 1988). This can be explained in terms of attentional resources. Attentional resources include those executive functions, or “... general purpose control mechanisms that modulate the operation of various cognitive sub-processes and thereby regulate the dynamics of human cognition ...” (Miyake et al., 2000, p. 50). Miyake et al. demonstrated the separability of three distinct central executive functions, namely “shifting” (mental set/task shifting), “updating” (information updating and monitoring), and “inhibition” (inhibition of prepotent responses). Allen et al. (2006) concluded that the actual task of MOT was achieved via moment-by-moment updating of some form of memory trace by execu-

tive processes, potentially supported by strategies generated by modality-specific processes.

People are so good at the tracking task that they can continue to track successfully even when all the objects disappear from view for up to 900 ms. Performance is best when the objects reappear at the same location from which they disappeared, rather than where they might have been predicted to reappear, given their speed and trajectory (Keane & Pylyshyn, 2006). This suggests that when all the objects disappear, tracking ceases and target reacquisition, when the objects reappear, is based upon some recorded trace of the objects' last positions. Horowitz, Birnkrant, Fencsik, Tran, and Wolfe (2006) have suggested a “task-switching” account in which, when a whole stimulus set disappears, their last locations are written down into long-term memory and this record of the last state is used, when objects reappear, to resume tracking.

While in the general population an individual tracking capacity of up to four items appears the norm, a number of studies have demonstrated that this figure can be increased if individuals regularly engage in activities which have a high dynamic spatial component. For example, Allen, McGeorge, Pearson, and Milne (2004) examined tracking capacity in a group of radar operators. They found that the radar operators had an enhanced tracking capacity relative to controls who had not had any radar operating experience. Similarly, Green and Bavelier (2006a) showed enhanced tracking abilities as a result of people playing action-video games and Trick, Jaspers-Fayer, and Sethi (2005) demonstrated that engaging in action sports predicted improved tracking ability in children.

The work by Green and Bavelier (2003, 2006a, 2006b) on action-video game players has provided insight into the potential cognitive/perceptual changes that underpin the increases in tracking capacity observed in these “expert” groups. In a series of experiments, Green and Bavelier (2003) demonstrated how, compared to individuals without action-video game experience, useful field-of-view was enhanced in action-video gamers, and gamers showed faster recovery from attentional blink. In addition, gamers showed an enhanced ability to enumerate items in briefly presented displays. Together these changes indicate that action-video gamers have an increased availability of attentional resources and are also better able to allocate these resources within the visual field (Green & Bavelier, 2003, 2006b). Given the surface similarities in the activities of the gamers studied by Green and Bavelier and the radar operators studied by Allen et al. (2004), it seems parsimonious to expect that the same attentional changes have occurred in radar operators as can be found in gamers. Indeed, as might be expected, given increased attentional resources, the performance of radar operators in an MOT task is less adversely influenced by the addition of a concurrent, attentionally demanding task than is the performance of matched controls (Allen et al., 2004) and, recently, Allen and McGeorge (2008) have shown that radar operators also show enhanced enumeration abilities.

Green and Bavelier (2006a) have argued that the increases in attentional resources seen in gamers may result in enhanced fidelity in visual short-term memory, or that gamers may have a more durable visual short-term memory trace and/or they are able to cycle through active memory traces faster than nongamers. In respect of the latter possibility, Oksama and Hyona (2004) have shown that performance on an MOT task is correlated with task-switching ability. Task switching provides a measure of how effectively an individual can shift attentional focus.

In sum, engaging in activities with a high dynamic spatial component (action video gaming, air traffic control, and action sports) appears to facilitate performance on other tasks requiring multiple objects to be tracked simultaneously. The facilitative effects of these dynamic spatial activities may be the result of enhanced visuospatial memory traces, created as the result of an increased availability of attentional resources.

This study compares the performance of a group consisting of members of a University Officer Training Corps (OTCs) with matched controls on an MOT task that sought, specifically, to investigate the durability and fidelity of the requisite attentional representations. University OTCs are army cadet forces that offer an optional gateway into the British Army, whose goal is to develop the leadership potential of selected university students. They do this through a range of team-based battlefield and adventure exercises, overseas expeditions, and various action sports such as rugby, hockey, and cricket. Such activities engage dynamic spatial skills and can be expected to enhance their object-tracking skills relative to controls. Therefore, we would predict that, dependent upon where their “expertise” is manifested: (1) OTC members’ memory traces will be

more durable than those of controls, as reflected in greater accuracy in target/distracter probe identification or (2) the fidelity of the visuospatial memory traces created by members of the OTC will be higher than controls, as reflected in smaller location errors.

Methods

Participants

Sixteen members of the University of Aberdeen’s Officer Training Corp (nine men), membership ranging from 2 to 4 years, and 22 undergraduate controls (nine men) took part in the study. OTCs’ ages ranged from 19 to 24 years ($M = 21.6$, $SD = 1.5$) and for controls, ages ranged from 17 to 27 ($M = 19.5$, $SD = 2.3$). Four participants were subsequently dropped from the analyses (1 OTC) because they did not have a complete dataset.

Materials and Procedure

Participants were tested individually using Superlab 4.0 software (Cedrus Corp.) on a PC with a 17" CRT monitor. The procedure was explained and ethical consent obtained. Participants then completed 144 trials of an MOT task, lasting approximately 45 min. Each trial began with the presentation of 10 white circular objects located randomly on a black background (see Figure 1). On presentation, a subset of between two and five objects flashed red to indicate that they constituted the target set, the remaining items being distracters. After 5 s all the objects began to move in random directions for a variable time of approximately 6 s. For half the trials, a 1,000 ms blank screen was presented after the objects ceased to move (delay condition) and this was followed by the re-presentation of the final screen positions of 9 out of the 10 objects in the trial. Participants were instructed to use the mouse to indicate the position of the missing object, a target for half the trials, as quickly and accurately as they could, and then indicate with a key press whether the missing item was a target or a distracter. Both the screen coordinates indicated by the participant and the identity of the object were recorded for subsequent analysis. For the remaining trials there was no delay between the completion of the object movement phase and participants indicating the location and identity of the missing object (no-delay condition). Each participant undertook a practice trial prior to completing two blocks, each consisting of 72 trials.

The accuracy with which participants correctly identified the missing objects, converted to a tracking capacity estimate, and mean location error measures were calculated for each participant. Accuracy in the delay condition, relative to the no-delay condition, will give some measure of visuospatial representation durability while fidelity of the visuospatial representation will be reflected in location error scores.

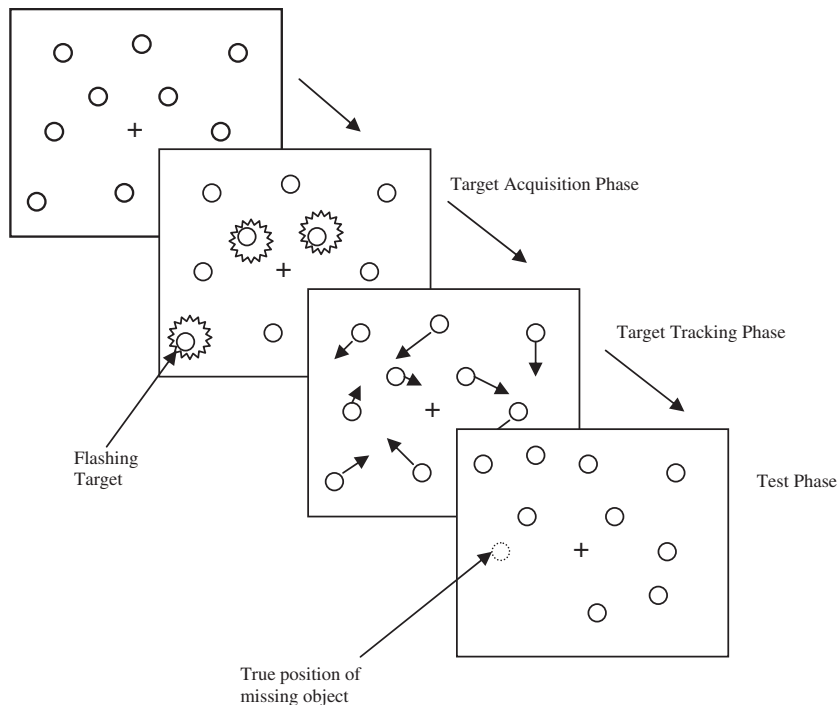


Figure 1. Example of an MOT trial.

Results

Measure of Tracking Performance

Following the procedure developed by Scholl, Pylyshyn, and Feldman (2001), “effective number of items tracked” (ENOT) scores were derived for each participant as a function of target set size (2, 3, 4, and 5) and delay condition (see Figure 2). ENOT scores were calculated using the formula $m = n(2P - 1)$, where m represents the ENOT, n the number of targets, and P the observed proportion of correct answers to probes. The ENOT measure assumes that observed individual performance is based on correctly tracking some items and guessing others (see Scholl et al., 2001, for a detailed exposition of the measure). However, the ENOT calculation does not account for any response bias, and there is clear evidence that experts and novices can differ in their response bias, with experts being consistently the more conservative (Allen et al., 2004). It is, therefore, important to control for this. Thus, the original accuracy scores, upon which the above formula is applied, were first adjusted to eliminate any such biases by subtracting the frequency with which participants responded “target” to a distracter probe (False alarms, FAs) from the times they successfully responded “target” to a target probe (i.e., Hits – FAs). Finally, all negative ENOT values were set to zero since, as estimates of tracking ability, negative ENOT values are illogical. The resulting data were subject to a three-way, repeated-measures ANOVA with delay (no delay, delay) and target set size (2, 3, 4, and 5) as the within-subjects variables, and expertise (OTC, control) the between-subjects factor. Wherever necessary, Greenhouse-Geisser corrected values are reported.

Significant main effects of delay and expertise were moderated by a significant interaction of Delay \times Expertise, $F(1, 32) = 8.11, p < .01, \eta_p^2 = .20$ (see Figure 2a), in that the OTCs significantly outperformed controls. However, independent-sample t tests revealed this was only in the no-delay condition (no-delay $t(32) = 2.75, p = .01$; delay $t(32) = 0.56, p = ns$). In addition, a significant main effect of target set size was moderated by a Delay \times Target set size interaction, $F(2.28, 73.06) = 5.82, p < .01, \eta_p^2 = .15$ (see Figure 2b). Paired-sample t tests (p set to .0125 for multiple comparisons) revealed that the effect of a delay was only significant for 3-target trials (2-target: $t(33) = 0.90, p = ns$; 3-target: $t(33) = 4.16, p < .01$; 4-target: $t(33) = 2.52, p = .017$; 5-target: $t(33) = 0.49, p = ns$).

Note, a repeated-measures ANOVA, with block (1st, 2nd) as the within-subjects variable and expertise (OTC, control) as the between-subjects variable, did not produce either significant main effects or a significant interaction, suggesting the absence of a practice effect in participants’ tracking performance.

Measure of Location Accuracy

For each participant mean location errors (being the displacement scalar between true and reported item position measured in pixels) were calculated as a function of the nature of the item being probed (target or distracter). These location error scores were then initially subject to a four-way, repeated-measures ANOVA with delay (no delay, delay), probe type (target, distracter), and target set size (2, 3, 4, and 5) as the within-subjects variables, and expertise (OTC, control) as the between-subjects factor.

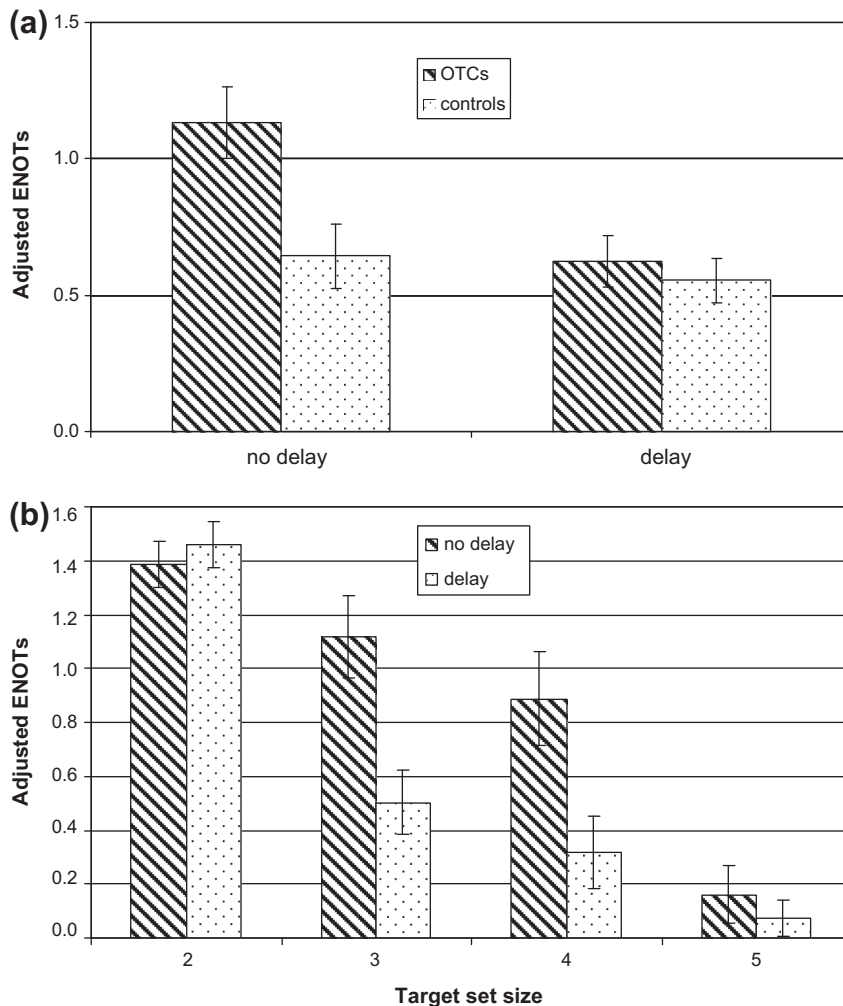


Figure 2. Adjusted ENOTs as a function of (a) expertise and delay (b) target set size and delay.

There was a significant interaction of Delay \times Expertise, $F(1, 32) = 4.42, p = .043, \eta_p^2 = .12$. Paired-sample t tests of OTCs and controls separately showed that a delay significantly reduced OTCs' location accuracy ($t(14) = 2.29, p = .038$) but had no effect on controls ($t(18) = 0.14, p = ns$) (see Figure 3).

In addition, there were main effects of probe, $F(1, 32) = 231.89, p < .01, \eta_p^2 = .88$, in that targets were located much more precisely than were distracters ($M = 123.75, SE = 10.50$ vs. $M = 294.48, SE = 12.01$), and target set size, $F(3, 96) = 12.53, p < .01, \eta_p^2 = .28$, in that accuracy declined significantly beyond a set size of four targets. However, these effects were moderated by a significant Probe \times Target set size interaction, $F(2.30, 73.50) = 5.42, p < .01, \eta_p^2 = .15$, in that, as target set size increased accuracy was decreasing for targets until at 5 targets there was a significant decline in both target and distracter localization; and Delay \times Target set size interactions, $F(2.28, 96) = 3.47, p = .031, \eta_p^2 = .10$, in that, while there was no effect of delay upon target set sizes of 4 or less, the effect of a delay on a set size of 5 targets was a significant reduction of accuracy over performance with no delay. Finally, these interactions were themselves moderated by a

three-way interaction of Delay \times Probe \times Target set size, $F(2.87, 91.87) = 5.92, p < .01, \eta_p^2 = .16$.

Note, a repeated-measures ANOVA, with block (1st, 2nd) as the within-subjects variable and expertise (OTC, control) as the between-subjects variable, did not produce either significant main effects or a significant interaction, suggesting the absence of a practice effect in participants' location accuracy.

Discussion

Individuals whose work or leisure activities contain a significant visuospatial component (e.g., action-video game players, radar operators, and action-sports players) all show an enhanced ability to track multiple, randomly moving targets compared with individuals who lack such extensive dynamic visuospatial experience (e.g., Allen et al., 2004; Green & Bavelier, 2006b; Trick et al., 2005). The results of this study confirm the expectation that this extended to Officer Training Cadets (OTCs), the selected "expert" group.

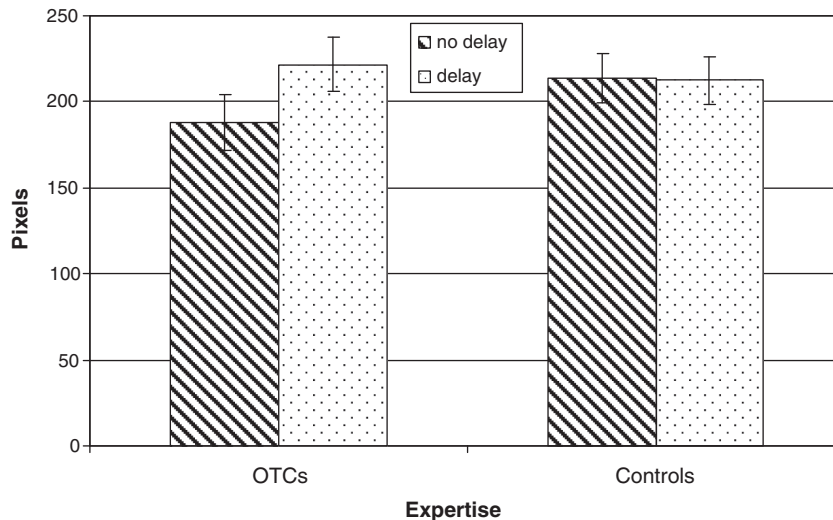


Figure 3. Mean pixel displacement as a function of expertise and delay.

The essential question for this research was whether, as suggested by Green and Bavelier's (2006a) observations of action-video game players, the increases in attentional resources seen among such experts result from either greater fidelity and/or more durable visuospatial memory traces than those of controls.

Compared to the control group, the OTCs were able to track significantly greater numbers of targets, based upon their overall accuracy at correctly identifying target and distracter probes. However, this enhanced tracking ability was only in evidence when participants were tested immediately after the objects had ceased to move. A delay of just 1,000 ms resulted in tracking performance that was only equivalent to that observed in the control group. Consequently, there is no evidence to indicate that the visuospatial representations that underlie target tracking are any more durable in OTC participants than in the controls. While this may appear counterintuitive, in fact, all previous evidence that might suggest an enhanced ability to track multiple objects were either from experiments consisting of trials with a single period of tracking or from those where a temporal gap appeared between two periods of active tracking, so that any visuospatial trace only needed to be of sufficient resolution to reacquire the targets (e.g., Alvarez, Wolfe, Horowitz, & Arsenio, 2001; Fencsik, Klieger, & Horowitz, 2007; Keane & Pylyshyn, 2006); the active process of tracking providing the enhanced sensitivity. In this study, with no post temporal gap period of tracking, the more sensitive executive processes associated with moment-to-moment tracking were disengaged when the objects disappeared and were never subsequently reengaged.

Measures of location error clearly demonstrated how much more accurately targets can be located, compared to distracters, ostensibly because of the greater attentional resources applied to the former. However, no significant expertise effect was reported. And though a delay by expertise interaction was revealed, this only indicated that experts' own performance varied significantly between the no-delay and delay conditions, not that they significantly exceeded those of the controls. Further, while the fact that targets were

located more accurately than distracters might be taken as support for the notion that distracters are actively inhibited during tracking (Pylyshyn, Haladjian, King, & Reilly, 2008), other explanations, such as a passive attentional gradient across targets and distracters, are equally supported. And since there was no significant Probe \times Expertise interaction it is unlikely that OTC's expertise is explained by a greater ability to inhibit distracters.

When we consider the results for both accuracy and location error together; that we only found an expertise effect in accuracy in the no-delay condition is significant because it clearly suggests that the underlying cognitive processes that support these two tasks differ. Further, that the expertise effect in accuracy did not carry over into the delay condition suggests that, whatever the nature of experts' "enhanced" visuospatial representations, they are vulnerable to decay.

Findings here support the idea that one set of central executive processes is involved in MOT by their dynamic, moment-by-moment updating of visuospatial representations of object locations, as suggested by Allen et al. (2006), with attentional resources being focused more significantly upon targets. It is here that expertise effects arise. When an entire object set disappears simultaneously, updating processes cease. The last memory trace begins to decay and a separate, set or task-switching process saves a visuospatial representation of the objects' last displacements into long-term memory (Horowitz et al., 2006), or into a capacity-limited episodic buffer (Baddeley, 2000; Oksama & Hyona, 2008). Note, however, that we are not saying that such savings *only* happen when active tracking ceases. In fact both processes may be occurring all the time, so long as there are objects in motion.

Thus, if we accept the idea that the decay profiles of the tracking processes' last memory traces do not differ between OTCs and control then, with no evidence of an expertise effect in the delay condition, we suggest that the memory traces had all but disappeared by the end of the delay period. If so, this would mean that only the long-term memory representations of the objects' last displacements were available

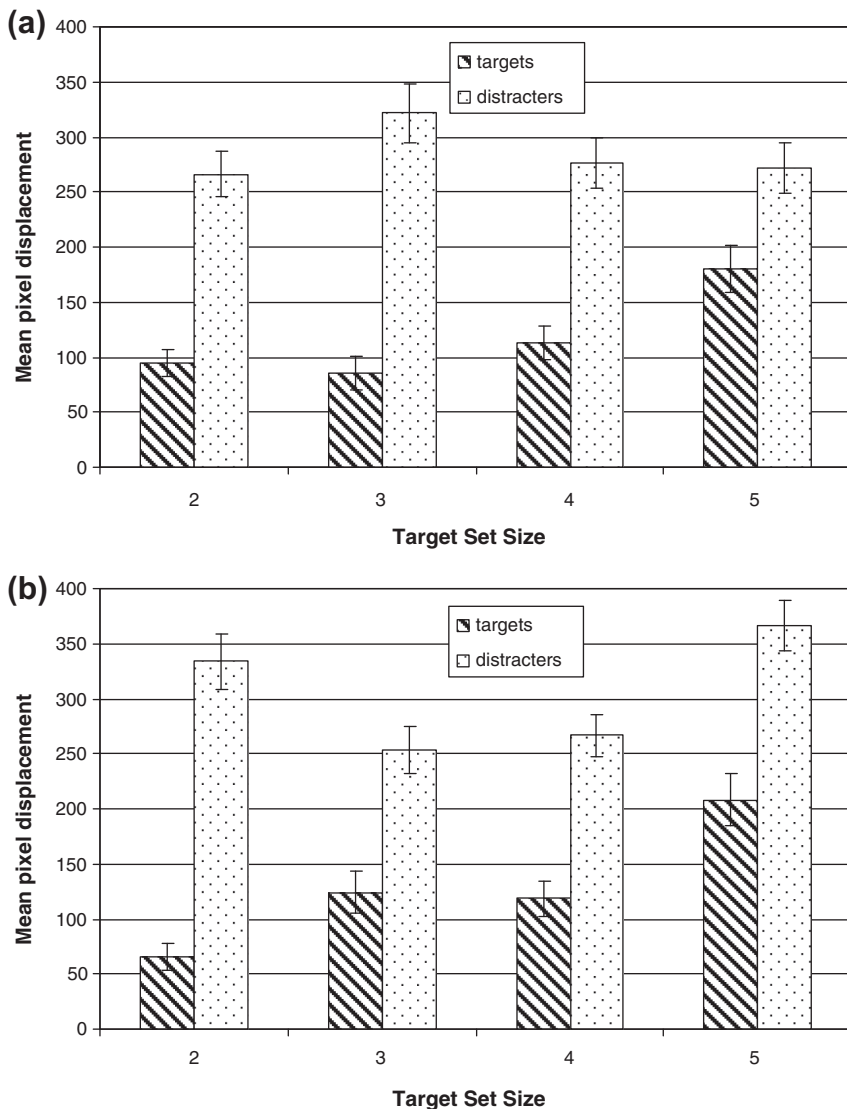


Figure 4. Mean pixel displacement as a function of target set size, delay and probe under (a) no-delay condition and (b) delay condition.

in order to determine target/distracter identification in the delay condition and to locate objects in both no-delay and delay conditions. And with no advantage of durability or fidelity, it may be that expertise effects are associated with an increased updating frequency during moment-to-moment tracking.

Finally, Alvarez and Franconeri (2007) have proposed a flexible-resource model to account for MOT (see also Allen et al., 2004). One of the implications of this model is that as the number of targets to be tracked increases the spatial resolution of targets should become coarser. The results of this study provide qualified support for this hypothesis. As expected, for correctly identified targets, the accuracy with which participants localize targets decreases as the number of targets increases (see Figure 4). However, the pattern suggests that this difference lies predominantly between trials where the target set size is greater than four and trials where the target set size is less than or equal to 4. Previously, it has been shown that

participants are able to track up to four targets accurately (e.g., Allen et al., 2006). The current results could indicate that it is only outside this range that the coarseness of coding becomes an issue and that it is this coarseness that may account for the difficulty in tracking higher numbers of targets.

Acknowledgment

This work was supported by the British Academy (ref. SG-44895).

References

Allen, R., & McGeorge, P. (2008). Enumeration: Shape information and expertise. *Acta Psychologica*, 129, 26–31.

- Allen, R., McGeorge, P., Pearson, D. G., & Milne, A. B. (2004). Attention and expertise in multiple target tracking. *Applied Cognitive Psychology, 18*, 337–347.
- Allen, R., McGeorge, P., Pearson, D. G., & Milne, A. B. (2006). Multiple-target tracking: A role for working memory? *Quarterly Journal of Experimental Psychology, 59*(6), 1101–1116.
- Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*(13), 1–10.
- Alvarez, G. A., Wolfe, J. M., Horowitz, T. S., & Arsenio, H. C. (2001). Limits on multielement tracking [Abstract]. *Journal of Vision, 1*(3), 347a.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423.
- Fencsik, D. E., Klieger, S. B., & Horowitz, T. S. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception & Psychophysics, 69*(4), 567–577.
- Green, C. S., & Bavelier, D. (2003). Action video game modifies selective attention. *Nature, 423*, 534–537.
- Green, C. S., & Bavelier, D. (2006a). Enumeration versus multiple object tracking: The case of action video game players. *Cognition, 101*, 217–245.
- Green, C. S., & Bavelier, D. (2006b). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1465–1478.
- Horowitz, T. S., Birnkrant, R. S., Fencsik, D. E., Tran, L., & Wolfe, J. M. (2006). How do we track invisible objects? *Psychonomic Bulletin and Review, 13*(3), 516–523.
- Keane, B. P., & Pylyshyn, A. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low level no-predictive function. *Cognitive Psychology, 52*, 346–368.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.
- Oksama, L., & Hyona, 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*(5), 631–671.
- Oksama, L., & Hyona, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology, 56*(4), 237–283.
- Pylyshyn, Z. W., Haladjian, H. H., King, C. E., & Reilly, J. E. (2008). Selective nontarget inhibition in multiple object tracking. *Visual Cognition, 16*(8), 1011–1021.
- Pylyshyn, Z., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179–197.
- 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.
- Trick, L. M., Jaspers-Fayer, F., & Sethi, N. (2005). Multiple-object tracking in children: The “catch the spies” task. *Cognitive Development, 20*, 373–387.

Received March 13, 2009

Revision received April 23, 2009

Accepted April 29, 2009

Published online: December 14, 2009

Roy Allen

School of Psychology
King's College
University of Aberdeen
Aberdeen AB24 2UB
Scotland
United Kingdom
Tel. +44 1224 272665
Fax +44 1224 273426
E-mail roy.allen@abdn.ac.uk
