

The role of “rescue saccades” in tracking objects through occlusions

Gregory J. Zelinsky

Andrei Todor

Department of Psychology, Stony Brook University,
Stony Brook, NY, USA, &
Department of Computer Science,
Stony Brook University,
Stony Brook, NY, USA



Department of Computer Science,
Stony Brook University,
Stony Brook, NY, USA



We hypothesize that our ability to track objects through occlusions is mediated by timely assistance from gaze in the form of “rescue saccades”—eye movements to tracked objects that are in danger of being lost due to impending occlusion. Observers tracked 2–4 target sharks (out of 9) for 20 s as they swam through a rendered 3D underwater scene. Targets were either allowed to enter into occlusions (occlusion trials) or not (no occlusion trials). Tracking accuracy with 2–3 targets was $\geq 92\%$ regardless of target occlusion but dropped to 74% on occlusion trials with four targets (no occlusion trials remained accurate; 83%). This pattern was mirrored in the frequency of rescue saccades. Rescue saccades accompanied $\sim 50\%$ of the Track 2–3 target occlusions, but only 34% of the Track 4 occlusions. Their frequency also decreased with increasing distance between a target and the nearest other object, suggesting that it is the potential for target confusion that summons a rescue saccade, not occlusion itself. These findings provide evidence for a tracking system that monitors for events that might cause track loss (e.g., occlusions) and requests help from the oculomotor system to resolve these momentary crises. As the number of crises increase with the number of targets, some requests for help go unsatisfied, resulting in degraded tracking.

Keywords: eye movements, attention, visual cognition, active vision

Citation: Zelinsky, G. J., & Todor, A. (2010). The role of “rescue saccades” in tracking objects through occlusions. *Journal of Vision*, 10(14):29, 1–13, <http://www.journalofvision.org/content/10/14/29>, doi:10.1167/10.14.29.

Introduction

Imagine being at a pond and seeing many visually similar ducks swimming about; how is it that we perceive this as a stable visual scene when the objects in it are moving? Attempts to answer questions of this sort have led to the development of the multiple object tracking (MOT) paradigm (Pylyshyn & Storm, 1988). Although there are many variants, most MOT tasks designate targets within an array of visually identical disks, then ask observers to track the movements of these disks and to indicate afterward the locations of the targets. The fact that they can do this accurately for up to 9 targets by some reports (Alvarez & Franconeri, 2007) is commonly interpreted as evidence against MOT being mediated by a single focus of attention moving rapidly from object to object (see also Alvarez & Cavanagh, 2005). In place of single-focus theory, dominant theories of MOT have instead posited the existence of either preattentive perceptual indices attached to each of the moving objects (e.g., Pylyshyn, 2001), or a multi-focal distribution of attention (Cavanagh & Alvarez, 2005) that allows each

object to have its own dedicated spotlight tracking it as it moves.

Important to the present discussion, the rejection of single-focus attention theory likely contributed to the widespread disregard of eye movements, another explicitly single-focus behavior, as a meaningful factor affecting tracking performance. Reinforcing this practice was an early analysis comparing observers who were free to make eye movements during tracking to those who were instructed to maintain central fixation (Pylyshyn & Storm, 1988). After excluding trials from the fixation condition in which eye movements were detected, tracking performance was found to be highly comparable between these two groups, leading the authors to conclude that the core tracking processes are mediated by attention and not changes in gaze. Following this demonstration, subsequent tracking studies either instructed observers to maintain fixation but did not explicitly monitor gaze (e.g., Pylyshyn, 2004; Scholl & Pylyshyn, 1999) or, more commonly, ignored eye movements altogether and allowed observers to freely shift their gaze during tracking (e.g., Intriligator & Cavanagh, 2001; Yantis, 1992). Since its inception, the MOT literature has therefore largely

discounted the contributions that eye movements might make to tracking performance, with some reports even going so far as to suggest that eye movements are irrelevant to the tracking task (Pylyshyn, 2004).

Recently, however, there has been a flurry of independent studies showing that MOT is accompanied by a rich repertoire of eye movement behavior. The first work on this topic had observers monitor for potential object collisions in the context of an air traffic control task (Landry, Sheridan, & Yufik, 2001), with the goal being to identify strategies for clustering targets. More recently, Fehd and Seiffert (2008) asked a similar question using a standard MOT paradigm. They found that observers tended to dynamically position their gaze on the moving centroid formed by the configuration of targets and demonstrated in follow-up work that this center-looking behavior is resilient to large changes in target speed and size (Fehd & Seiffert, 2010). Interestingly, however, targets were not ignored by gaze, as evidenced by frequent center–target gaze switching. Independently, Zelinsky and Neider (2008) also reported centroid looking in the context of multiple target sharks swimming in a three-dimensional (3D) underwater scene but found that this centroid-tracking strategy transitioned to a target-tracking strategy as tracking load increased from three to four targets. All of these findings are generally consistent with Yantis' (1992) suggestion that MOT can be accomplished by tracking the vertices of a virtual object formed by the target configuration. Eye movements during tracking have also been used to clarify the mechanism of concentration and amplification effects in MOT (Doran, Hoffman, & Scholl, 2009). Despite these studies, however, the fundamental question regarding the relationship between eye movements and MOT remains: Are these eye movements contributing somehow to tracking performance, or are they largely irrelevant to the task?

Recent work makes indisputable the fact that eye movements accompany natural MOT behavior. This gaze behavior has also proven to be useful in revealing the allocation of attention during a MOT task, similar to how eye movements have been used to study the allocation of attention over time in the context of static visual search (e.g., Zelinsky, Rao, Hayhoe, & Ballard, 1997). In the present study, we extend this role of eye movements during MOT, showing that these movements are not only a useful tool for understanding tracking but are actually *functional* in producing good tracking performance. We do this by linking eye movements to our ability to track objects successfully through occlusions.

Tracking objects as they undergo occlusion is a computationally challenging problem, one that causes most automated methods from computer vision to fail (e.g., Khan, Balch, & Dellaert, 2005; Yang, Li, Pan, & Li, 2005; Zhou & Tao, 2003). This is understandable; an occluded object disappears from view, leaving no visible part or property of the object to track. Yet despite the

inescapable complexity of the problem, the behavioral tracking literature tells us that humans have an exceptional ability to track objects through many forms of occlusion (Scholl & Pylyshyn, 1999). For example, Viswanathan and Mingolla (2002) studied MOT in the context of moving objects occluding each other and found that when such occlusions were accompanied by consistent accretion and deletion cues, tracking performance returned to non-occlusion levels. Flombaum, Scholl, and Pylyshyn (2008) combined MOT with a dot-probe task to study the allocation of attention to objects as they passed, or did not pass, behind stationary barrier occluders. Surprisingly, they found that probes were detected *more* accurately when they were on objects undergoing occlusion, compared to when they were on visible objects. They interpreted this finding as evidence for a burst of attention directed to occluded objects for the purpose of compensating for the added tracking difficulty created by the occlusion, a phenomenon they refer to as the attentional high-beams effect (see also Iordanescu, Grabowecky, & Suzuki, 2009, for a related idea applied to effects of crowding on tracking). The present study extends this idea to the eye movement domain. We hypothesize that our ability to track objects through occlusions is mediated by timely assistance from gaze in the form of *rescue saccades*—eye movements to tracked objects that are in danger of being lost due to occlusion.

Methods

Participants

Thirty observers from the Stony Brook University undergraduate psychology subject pool were tested. All were experimentally naive and had normal or corrected-to-normal vision, by self-report.

Stimuli and apparatus

Stimuli were nine sharks moving in a pseudorealistic manner throughout a 3D underwater scene (Zelinsky & Neider, 2008). Both the sharks and the scene were created, animated, and rendered using Autodesk's 3D Studio Max. Shark movement was confined to a virtual swim volume of 100 (width) \times 60 (height) \times 80 (depth) m, corresponding to a visual angle of $25^\circ \times 15^\circ$. The visual angles of individual sharks ranged from 0.48° to 1.13° , depending on their perceived depth in the scene. Trajectories were random, except for constraints to prevent collisions and abrupt bounces off of the boundaries of the swim volume. Stimuli were presented in color on a 19-inch CRT monitor, with DirectDraw used to display the 600 frames

comprising each motion sequence at 30 frames/s. Eye position was recorded throughout using an EyeLink II eye tracking system (SR Research), sampling at 500 Hz (with chin rest).

Procedure and design

A typical trial is illustrated in Figure 1.¹ The task was standard MOT, to indicate whether a probed shark was one of the targets designated at the start of the trial. There were three tracking load conditions; observers tracked 2, 3, or 4 target sharks, out of 9 total, as they swam for 20 s. We also manipulated target occlusion. In the no occlusion

condition, none of the target sharks occluded, or were occluded by, any other shark. In the occlusion condition, at least one target shark, and usually more, was involved in an occlusion. The occlusion manipulation was a within-observer variable, with the 112 trials per observer evenly divided into randomly interleaved occlusion and no occlusion trials. Tracking load was a between-observer manipulation; 10 observers participated in each of the Track 2–4 conditions.

Results and discussion

How accurately did people track?

Table 1 shows percent correct as a function of tracking load. Tracking accuracy declined with increasing load, $F(2,27) = 89.99, p < 0.001$, with this decline carried by the Track 4 condition differing from the rest ($p < 0.001$; all post-hoc tests Bonferroni corrected). Accuracy with two or three targets was high and did not reliably differ ($p > 0.05$). Tracking load also interacted with target occlusion, $F(2,27) = 12.89, p < 0.001$; target occlusions on Track 4 trials resulted in more errors compared to non-occluded targets. This observation is important. The tracking literature had assumed that people can track through occlusions with little problem (Flombaum et al., 2008; Scholl & Pylyshyn, 1999; Viswanathan & Mingolla, 2002), but in our task this was clearly not the case; occlusions resulting from intersecting objects degraded accuracy even at moderate tracking loads. Other factors potentially contributing to this load \times occlusion interaction may have been the three-dimensional movement of our objects, the fact that they changed size and direction unpredictably, and the relatively low contrast between the gray of the sharks and the blue of the sea.

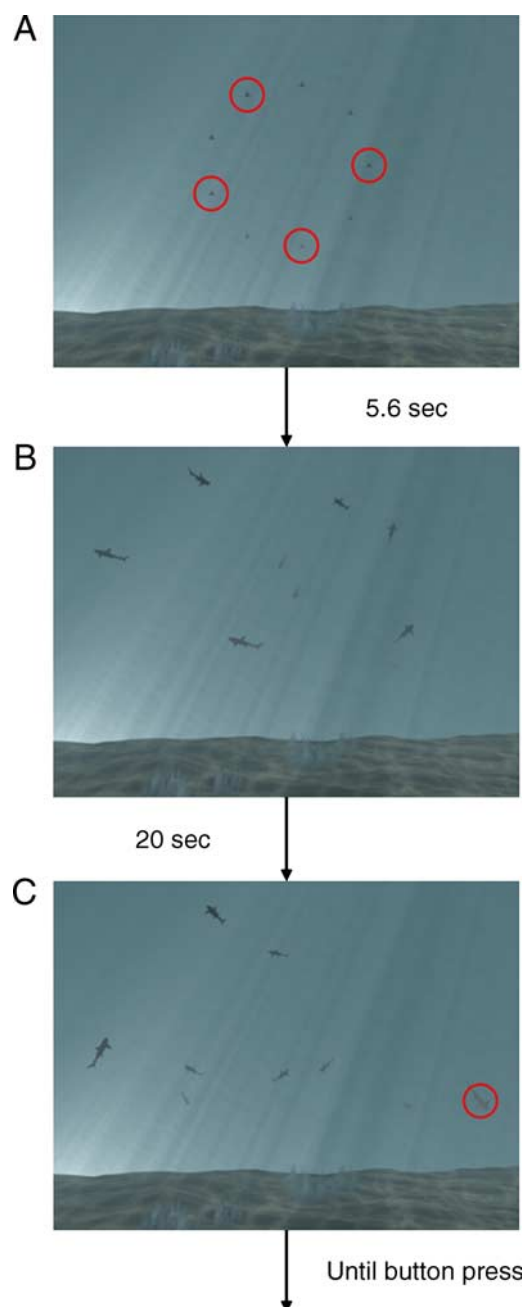


Figure 1. A representative trial from the Track 4 condition. (A) Each trial began with the static presentation of 9 sharks arranged into a circle and oriented so as to have a visually identical appearance. After 800 ms, two to four target sharks were designated by flashing red rings (4.8 s). Target positions were balanced, as much as possible, over the 9 positions in each load condition. (B) The sharks then swam throughout the scene for 20 s. (C) At the end of the motion sequence (frame 600), all objects froze in position and a red ring was displayed around one of the sharks. On half of the trials, this probed shark was one of the targets; on the other half of the trials, a non-target shark was probed. The observer's task was to indicate whether this probed object was one of the targets. The probe display remained visible until the judgment, and observers were instructed to respond as accurately as possible without regard for time. There was no accuracy feedback.

	Number of targets		
	2	3	4
No occlusion	97.7 (0.5)	94.8 (1.2)	83.2 (1.8)
Occlusion	96.2 (0.5)	91.5 (1.1)	74.0 (1.6)

Table 1. Percent correct tracking accuracy as a function of number of targets and occlusion condition. *Note:* Values in parentheses indicate one standard error of the mean (*SEM*).

How did occlusions vary with tracking load?

Given the above effect of occlusions on tracking, our next step was to better characterize the frequency of occlusions in our target conditions. This analysis is shown in Table 2 for occlusion trials. The average number of occlusions per trial increased roughly linearly with the number of targets. This was expected—the more targets, the more opportunities for these targets to become involved in occlusions. However, this relationship also suggests an explanation for the observed load \times occlusion interaction in the accuracy data. To the extent that each target occlusion increases the probability of a tracking error, then as the number of occlusions per trial increases with load, so does the potential for tracking failure. This also suggests that occlusions place some demand on a limited resource important for tracking; under low loads the system can deal with the detrimental consequences of occlusion, but not under high loads. As for what this limited resource might be, to answer this question we next turn our attention to the eye movements made during tracking.

Does looking at a target depend on its proximity to another object?

Our central premise is that people may look at targets in order to prevent occlusion-related track losses. A reasonable first step in establishing this relationship is to show that looking behavior depends on the minimum distance between a target and another object. More specifically, we expected that as the distance between a target and another object decreased, so too would the distance between gaze and that target. To look for this relationship, we computed for each stimulus frame the minimum distance between each target and the nearest other object, then found the scene-wide minimum distance from among these 2–4 (depending on the target number condition) minimum distance values. Doing this gave us a single value for that frame, TD_m , indicating the smallest distance between any target and another object. We then found the distance between the target used to compute TD_m and the observer's gaze position. Repeating this analysis for each of the 600 frames, 112 trials, and 30 observers gave us

over 2 million pairings of TD_m with gaze distance, GD . Finally, we grouped these data by target number condition and sorted them into 1° TD_m bins to create the three $GD \times TD_m$ functions shown in Figure 2.

Each Figure 2 function shows how the distance between gaze and the minimally separated target varies with that target's inter-item distance. The data points on each function represent averages from 10 observers. There are two patterns to note from this plot. First, and as predicted, gaze–target distance tended to decrease with the distance between that target and another object. Moreover, this decrease in gaze–target distance was largely confined to small values of TD_m . For the Track 2 condition, the mean gaze–target distance averaged across the $0\text{--}1^\circ$ TD_m bin was significantly smaller than the gaze–target distance averaged across the $1\text{--}2^\circ$ TD_m bin, $t(9) = 5.58$, $p < 0.001$ (paired group). For the Track 3 data, this dip shifted slightly to larger TD_m values; gaze–target distances differed between the $2\text{--}3^\circ$ and $1\text{--}2^\circ$ bins, $t(9) = 2.77$, $p < 0.05$ (paired group), but did not differ between the $1\text{--}2^\circ$ and $0\text{--}1^\circ$ bins, $t(9) = 0.76$, $p > 0.1$ (paired group). The Track 4 data followed a qualitatively different pattern, characterized by a linear decrease in gaze–target distance with decreasing TD_m and non-significant differences between successive bins (all $p > 0.05$; paired group). Second, gaze–target distances were relatively stable over larger values of TD_m , with none of the pairwise comparisons between bins across the $2\text{--}5^\circ$ range proving reliable. However, gaze–target distances averaged over this range were larger in the Track 2 data compared to the Track 3 data, $t(18) = 3.16$, $p < 0.01$, and the Track 4 data, $t(18) = 4.24$, $p < 0.001$. These differences may reflect the fact that observers in the Track 2 condition adopted more of a centroid-looking tracking strategy whereas observers in the Track 3 and Track 4 conditions tended to look more at individual targets (Zelinsky & Neider, 2008).

Rescue saccades: What are they and how did we define them?

The previous analysis demonstrated a tendency for our observers to look at targets when they approached (or were approached by) other objects. However, whereas this relationship would be expected if tracking were to use rescue saccades, it is lacking as a demonstration of this

Probed target	Number of targets		
	2	3	4
2.82 (0.52)	5.25 (0.68)	7.57 (0.75)	9.14 (0.90)

Table 2. Average number of target occlusions per trial. *Note:* The value under “Probed target” indicates the number of occlusions for the one target that was ultimately probed. Values in parentheses indicate one standard error of the mean (*SEM*).

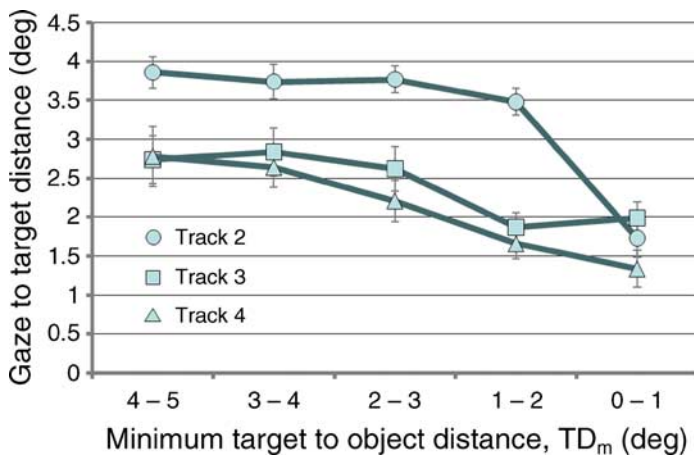


Figure 2. Mean gaze–target distance plotted as a function of the minimum target–object separation distance in the Track 2 (circles), Track 3 (squares), and Track 4 (triangles) conditions. Error bars indicate one standard error of the mean (SEM).

behavior in at least two respects. First, gaze–target distance might decrease for any number of reasons; perhaps a saccade is made to a target, perhaps gaze is pursuing an object that is moving toward a target, or perhaps gaze is fixated and a target is moving toward it. Central to the definition of a rescue saccade is the actual detection and rapid shift of gaze to a target that is about to be occluded, making it important to remove all other sources of decreasing gaze–target distance from estimates of rescue saccade frequency. Second, the Figure 2 functions do not specify how the alignment of gaze with a target is time-locked to its occlusion. This is because the frames grouped into each TD_m bin might have come from very different times during a trial. More generally, missing from these plots is *any* measure of time, making it impossible to know how soon before an occlusion gaze is directed to a target.

The first step in quantifying a relationship between a saccade and an occlusion is to define the occlusion event in time. Figure 3A shows our method for doing this. Illustrated are data from the motion of one target (t_i) on one trial for one observer in the Track 3 condition. The green function indicates the distance between t_i and the nearest other object. As this target swims through the scene, this distance will sometimes be large and sometimes small. However, during an occlusion this distance will necessarily drop to near zero. We define an occlusion as any overlap in the bounding boxes between a target and another object. The initial overlap tells us when an occlusion started; to find its end, we then determined the moment in time when the movement of these objects caused their bounding boxes to again separate. An *occlusion event* (OCE) is defined as the window of time extending 800 ms immediately preceding the end of the occlusion. For reasons to follow, we also required that the target–object distance function be monotonically decreasing

throughout the 800-ms window defining an OCE (excluding frames in which the bounding boxes actually overlapped). This resulted in the exclusion of 17.6% of events that would have been labeled as occlusions.

Having defined an occlusion event, we can now analyze gaze behavior relative to this event in order to establish a relationship between saccades and target occlusions. The red function in Figure 3A indicates the distance between target t_i and an observer’s gaze over time. As in the case of the target’s movement, sometimes this distance is large and sometimes it is small, with these distances depending on what the observer was choosing to look at in the display. The nearly vertical lines in this function indicate saccades, and the relatively flat periods immediately preceding and following the saccades indicate fixations. The stretches of this function that do not fall into these two categories are either smooth pursuit eye movements, periods during which the observer was tracking an object with their gaze, or some interaction or combination of events, such as when the target was moving toward gaze as the observer was pursuing a different object, or when pursuit was transferred from one object to another. Given the complexity of this function, all analyses were confined to gaze behavior during well-defined events—in this case, occlusions. Specifically, a rescue saccade was defined as a saccade landing 1 degree from a target during an OCE. In Figure 3A, we therefore find two rescue saccades: one occurring at ~ 11.5 s in which a large 6-degree saccade brought gaze to t_i , just within the critical time window (followed by a brief period of pursuit), and another occurring near the end of the trial. Note that the synchronous dip in the gaze–target and target–object distance functions occurring at ~ 6.5 s is not counted as a rescue saccade, for two reasons: the distance between t_i and another object, although small, did not meet the occlusion criterion (bounding box overlap), and gaze did not acquire the target via a saccade—to be counted as a rescue saccade, a saccade had to bring gaze to the target during the occlusion event.

How will we know if the frequency of rescue saccades is high or low, or if it is even different from chance? Answers to these questions require a baseline, and for this purpose, we defined the *before occlusion event* (BOE). Calculation of the BOE is illustrated in Figure 3B (left). Once an occlusion was detected (at ~ 4.9 s), we backed up the target–object distance function to find its preceding local maximum, which occurred at ~ 3.1 s in the illustrated data. The BOE is defined as the 800-ms window immediately prior to this maximum. Because the OCE required a monotonically decreasing target–object distance function over this portion of its time window, the BOE and the OCE could not overlap, and for every OCE, there was a corresponding BOE. With the BOE, we can directly compare the percentage of rescue saccades to the percentage of saccades made to within 1 degree of the same target during a comparably long 800-ms event, one that is divorced from an occlusion.

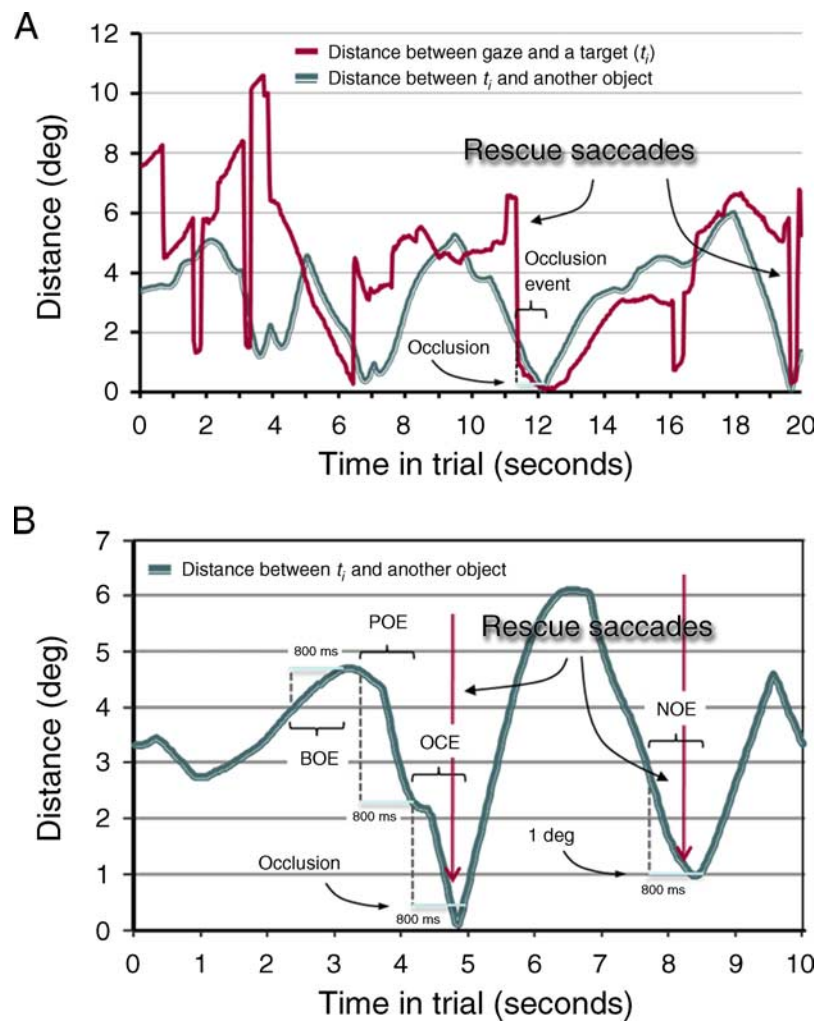


Figure 3. Representative data for one target (t_i) from one occlusion condition trial for one observer in the Track 3 condition. (A) Data in red indicate the distance in degrees between the observer's eye position and t_i as a function of time (gaze–target distance); data in green indicate the distance in degrees between t_i and another object, again as a function of time (target–object distance). The light blue line indicates the duration of the occlusion event. (B) A representative target–object distance function (one-half of a trial) illustrating a before occlusion event (BOE), a pre-occlusion event (POE), an occlusion event (OCE), and a near-occlusion event (NOE). The light blue lines indicate the event windows.

Figure 4A shows the mean percentage of OCEs (light bars, left) and BOEs (dark bars, right) that were accompanied by a saccade in the occlusion condition trials; data indicated by the middle bars will be discussed in the following section. Regardless of the number of targets, saccades were far more likely to be made to targets during an OCE compared to a BOE (all $p < 0.001$, paired group t -tests). Indeed, in the Track 2–3 conditions approximately half of all occlusion events were accompanied by a saccade. This is compelling evidence for the existence of rescue saccades; targets about to undergo an occlusion attracted far more saccades than when the same targets had not yet started to move toward other objects. This preferential looking behavior also varied with tracking load. The percentage of rescue saccades in the Track 4 condition was smaller than that in the Track 3

condition, $t(18) = 2.98$, $p < 0.01$, although comparison to the Track 2 condition was not significant, $t(18) = 1.69$, $p = 0.11$. An opposite trend was found for BOEs, with saccades to these events increasing slightly with tracking load, $t(18) \geq 3.01$, $p < 0.01$. These load effects likely reflect a system under stress; as load increased, observers made more saccades to targets generally (consistent with Zelinsky & Neider, 2008), but these looks to the target were less synchronized with occlusions.

Is it occlusion or proximity that signals rescue saccades?

Now that we know that rescue saccades exist, we can begin to explore the signal that calls them. Until now, we

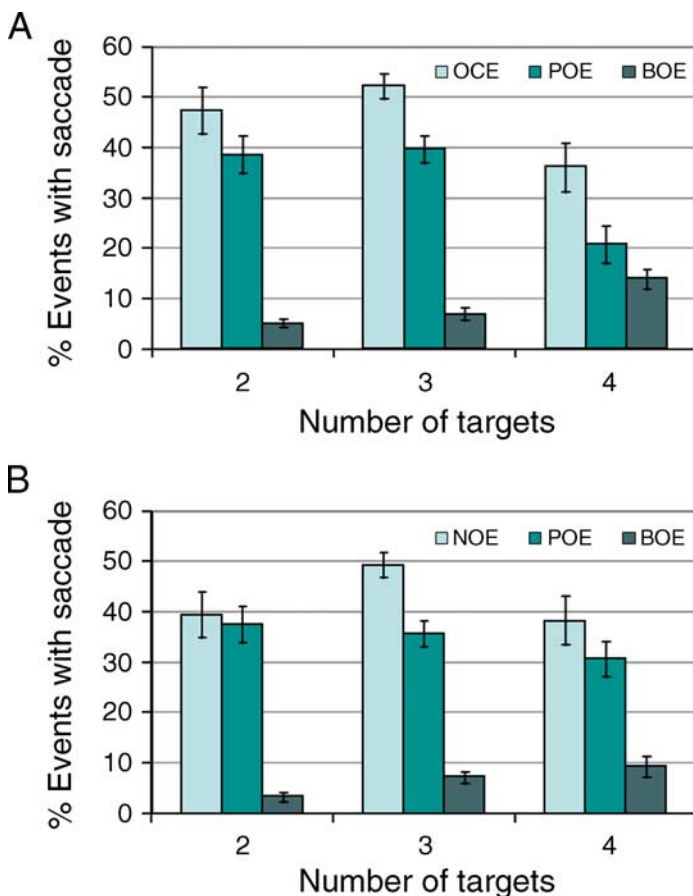


Figure 4. Mean percentage of events accompanied by a saccade. OCE = occlusion events; POE = pre-occlusion events; BOE = before occlusion events; NOE = near-occlusion events. (A) Occlusion trials. (B) No occlusion trials.

had assumed that it was the occlusion event itself that summoned a rescue saccade and that they occurred in response to an occlusion rather than in anticipation of one. This is highly unlikely; an actual occlusion occupied only about 200 ms (on average) of the OCE window, leaving very little opportunity for the detection of the occlusion and the programming and execution of a saccade to that target. Indeed, it was our observation during piloting of clearly anticipatory rescue saccades that motivated our backward-in-time placement of the OCE window; had we centered this window on the occlusion event, we would have greatly underestimated the occurrence of rescue saccades.

To quantify the dependency of rescue saccades on actual target occlusions, and to further explore the anticipatory nature of rescue saccades, we define two additional events, a *pre-occlusion event* or POE and a *near-occlusion event* or NOE (Figure 3B). A POE is an 800-ms window positioned relative to the start of the OCE (or NOE); it begins 800 ms before the OCE and ends when the OCE begins. As in the case of an OCE, we again required that the target–object distance function be monotonically decreasing throughout the POE window,

so as to prevent overlap with the BOE. As a result of this constraint, not every OCE could be paired with a POE. The NOE is a proxy for an OCE that can be applied to the no occlusion data. Although occlusions did not strictly occur on these trials, there were cases in which a target came close to another object before one or the other (or both) avoided an occlusion by veering away at the last moment. We defined an NOE based on a 1-degree distance threshold; the event was flagged when an object came within a degree of a target, and the event ended when this separation again exceeded the 1-degree criterion. The NOE event window extended 800 ms prior to the event’s end, and we again required that the target–object distance function be monotonically decreasing during this period (excluding frames in which the target was less than 1 degree from another object). If it is the proximity between a target and another object that is important for tracking, as has been recently suggested in the literature (Franconeri, Jonathan, & Scimeca, 2010; Franconeri, Lin, Pylyshyn, Fisher, & Enns, 2008; Intriligator & Cavanagh, 2001; Iordanescu et al., 2009; Tombu & Seiffert, 2008), then the conservative 1-degree threshold used to define an NOE would lead us to expect no meaningful difference in rescue saccade rate between NOEs and OCEs. However, if occlusions are somehow special irrespective of target–object distance, we should find more rescue saccades to OCEs compared to NOEs. To the extent that rescue saccades anticipate the occlusion event, we would also predict that the rate of rescue saccades to POEs would be greater than the BOE baseline rate.

Figures 4A and 4B show the percentage of saccades to each event type for the occlusion and no occlusion trials, respectively. The corresponding mean target–object separations for each event, averaged over its temporal window, are provided in Table 3. Consistent with our expectation, rescue saccade rates did not differ between OCEs (45.1%, averaged over load condition) and NOEs (42.3%, averaged over load condition). Separate comparisons conducted for the Track 2–4 conditions also failed to reveal any significant differences between these two event types (all $p \geq 0.19$). Analysis of mean target–object separations revealed reliably smaller separations for OCEs (by 95% confidence interval, CI), but the average separation difference between OCEs and NOEs was small (only 0.5°). We therefore conclude that the process responsible for triggering a rescue saccade does not differentiate between a near occlusion and an actual occlusion. Analysis of POEs revealed a saccade rate that was significantly lower than either OCEs, $t(9) = 6.25$, $p < 0.001$, or NOEs, $t(9) = 7.05$, $p < 0.001$. However, the rate of POE saccades was far higher than what was found for BOEs (occlusion trials: $t(9) = 12.75$, $p < 0.001$; no occlusion trials: $t(9) = 27.44$, $p < 0.001$; paired group). Averaged over all conditions, 33.7% of the POEs were accompanied by a saccade, compared to only 7.6% of the BOEs. This confirms that rescue saccades are highly

	Event type			
	OCE	NOE	POE	BOE
Occlusion	1.3 (± 0.05)	a	3.8 (± 0.17)	4.9 (± 0.11)
No occlusion	b	1.8 (± 0.05)	4.1 (± 0.22)	5.2 (± 0.12)

Table 3. Mean target–object separation (degrees) averaged over the event window for each event type and occlusion condition. *Notes:* (a) Too few NOEs existed in the Occlusion data for analysis. (b) OCEs are undefined in the No occlusion condition. Values in parentheses indicate a 95% confidence interval. OCE = occlusion event; NOE = near-occlusion event; POE = pre-occlusion event; BOE = before occlusion event.

anticipatory—commonly triggered by events (occlusions or near occlusions) that would not occur for another 1,600–800 ms. Also consistent with the patterns shown in Figure 2, target–object separations for POEs were substantially larger than separations for OCEs and NOEs and smaller than BOE separations (all by 95% CI within occlusion condition). Taken together, this relationship between rescue saccades and target–object separation, combined with the high overall rate of saccades to POEs and the lack of a difference between OCEs and NOEs, strongly supports the claim that target–object proximity, and not occlusion per se, triggers a rescue saccade. Our observers made more saccades to a target as the distance between that target and another object decreased, presumably in anticipation of these two objects ultimately intersecting.

How does the potential for target–distractor confusion affect rescue saccades?

We know that rescue saccades are made to occlusion and near-occlusion events, and that these saccades are likely triggered by the perceived immediacy of an occlusion, as evidenced by their dependence on target–object separation. However, what is it about the occlusion of one object by another that makes it deserve such a drastic intervention by overt attention? Our assumption has been that it is the potential for target–distractor confusion; that object individuation is threatened during an occlusion, thereby increasing the potential for the swapping of identities. In the case of an intersecting target and distractor, such identity swapping would indeed be highly detrimental to the task, resulting in the loss of that target and the consequent potential for an error. However, such swapping is far less consequential in the case of two targets intersecting, as each would still remain a target even after swapping identities.

If the tracking process responsible for signaling rescue saccades makes optimal use of information about potential target confusion, then rescue saccades to target–target occlusions should be less common than those to target–distractor occlusions. To test for this possibility, we collapsed the OCE and NOE data into one more general

occlusion event and grouped these events according to whether the target intersected a non-target object or another target. We then reanalyzed the rescue saccade data by group. We restricted this analysis to the Track 4 data because this condition maximized the number of target–target occlusion cases. The proportion of rescue saccades to target–distractor occlusions was 0.39; the proportion of rescue saccades to target–target occlusions was 0.4. This difference was not significant by paired group *t*-test, $t(9) = 0.78$, $p = 0.45$. Analysis of the combined POE data yielded a similarly non-significant result.

What can be made of our finding that rescue saccades are insensitive to the potential for target confusions? It might mean that this process is simply not optimal; rescue saccades may still be driven by the potential for confusion, but the tracking process does not distinguish between target–target and target–distractor intersections when signaling these saccades. We consider this possibility to be highly plausible; given that the anticipation of an occlusion is already a demanding computational feat, making this computation contingent on object identity may be asking too much of a just-in-time process. Certainly, it would be simpler for the system to assume that all intersections are potentially problematic, regardless of what the constituent objects may be. Relatedly, such identity checking may not even be part of the tracking process. MOT tasks are artificial in that object identity information is purposefully de-weighted, typically by having all of the objects be visually identical when the targets are designated. However, this is not the case for most tracking tasks in the real world, where maintaining object identity may be half the battle (e.g., Oksama & Hyönä, 2008). It is often not good enough to know that an object is a target, one must also know *which* target. Our tracking system may therefore seek to maintain object identities through an occlusion, regardless of whether the two objects are both targets or a target and a distractor. Of course, a final possibility is that the system does differentiate between target–target and target–distractor intersections, but that our experiment was simply unable to reveal this difference. Our argument against differentiation rests on a negative result and must therefore be weighted accordingly. Determining whether this absence of an

effect is real or not will be an important direction for future work, as the answer to this question will inform the computation used to signal rescue saccades, and ultimately the situations that threaten to reduce tracking efficiency.

General discussion

Our ability to track objects through occlusion is due in part to the timely assistance of gaze in the form of rescue saccades. We have a remarkable ability to track multiple moving objects, and to do so through events that currently cause computer vision tracking methods to fail miserably. How can a problem of such undeniable computational complexity be solved so thoroughly by humans? A takeaway message from this study is that this is accomplished through team effort. We developed no sophisticated computational solution to the occlusion problem; our solution relies instead on a highly proactive tracking process, one that monitors for events that might cause a track loss, such as occlusions, and requests help from the eye movement system to resolve these momentary crises when they are detected.²

For tracking purists, our proposal that the tracking system occasionally outsources help from the oculomotor system may seem unsettling, but this practice of relying on eye movements to help resolve moment-by-moment behavioral crises is very likely the perceptual rule rather than the exception. Real-world perception happens at a frenzied pace; problems arise, and solutions are found, perhaps many times each second. Information from the selective allocation of gaze is probably an important part of many of these solutions. Even the coordination of very simple perceptual-motor tasks, such as making a peanut butter and jelly sandwich (Land & Hayhoe, 2001), is sufficiently computationally demanding so as to require gaze to be directed to objects just moments before they are manually manipulated (see also Land, Mennie, & Rusted, 1999). Hayhoe and colleagues referred to these behaviors as “just in time” fixations, in recognition of the fact that these fixations acquire information to be used in the solution of an immediate perceptual problem (Aivar, Hayhoe, Chizk, & Mruzek, 2005; Ballard, Hayhoe, & Pelz, 1995; Droll, Hayhoe, Triesch, & Sullivan, 2005; Hayhoe, Bensinger, & Ballard, 1998; Hayhoe, Shrivastava, Mruzek, & Pelz, 2003; see Hayhoe & Ballard, 2005, for a review). In this sense, rescue saccades are just another example (albeit a very good one) of “just in time” gaze behavior, one tailored to a problem commonly encountered in a MOT task.

Given that rescue saccades assist in the tracking of objects through occlusion, one might expect more of them with increased tracking load. This is not what we found;

moving from three to four targets decreased both accuracy and the proportion of rescue saccades. Why did the proportion of rescue saccades not increase when they were needed the most—in the Track 4 condition? One reason is likely due to the fact that tracking four targets in this task was difficult. If one of the four targets was frequently lost, a possibility consistent with the drop in Track 4 accuracy, then rescue saccades would not be expected to these lost targets, resulting in their lower Track 4 frequency. Another reason may be that there were simply too many crises in this condition to devote a rescue saccade to each, thereby driving down their proportion. With four targets, an occlusion would occur, on average, about every other second, and when you factor in the high rescue saccade rates to pre-occlusion and near-occlusion events, very often events might have competed for a rescue saccade. Declines in tracking performance are often explained in terms of competition for limited attention resources; the analogous suggestion here is that rescue saccades may be another limited resource important for tracking success. Whereas the existence of a limited capacity pool of attention resources has been questioned (Allport, 1980; Logan, 1997; Neumann, 1987), foveal resource limitations are real and their expression is undeniably serial; we have only one functional eye and it can be pointing at only one place at a time. When two (or more) crises occur at about the same time, as might often have happened when there were four targets in our task, only one of these targets could be rescued, leaving the other to go through the crisis unassisted. To the extent that looking at targets helps prevent occlusion-related track losses, then *gaze* may be the limited resource that constrains tracking performance in these situations, not attention.

The above distinction between eye movements and attention belies the fact that the two behaviors are almost certainly related; the spatially selective properties of attention likely act as a pointer telling the eye movement system where to look next (Deubel & Schneider, 1996; Kowler, Anderson, Doshier, & Blaser, 1995; see Findlay, 2009; Findlay & Gilchrist, 2003, and Zelinsky, 2008, for reviews). Assuming this close relationship holds during a tracking task, the implication is that covert attention was also preferentially directed to targets during, or moments before, an occlusion. The tight coupling between these behaviors begs the question of which was actually responsible for rescuing occluded targets—the eye movement or the shift of attention that preceded it? Flombaum et al. (2008) suggested that it is attention that serves this function; that there is a supplementary pool of emergency resources that tracking can draw on in times of crises, much like switching on the high beams of a car in anticipation of a treacherous patch of road. Our suggestion is that it is the rapid allocation of foveal resources that assists tracking through occlusions and that the function previously attributed to attentional high beams is actually fulfilled by the darting movements of rescue saccades.³

From a theoretical perspective, the suggestion that tracking performance is tied to the availability of attention resources is controversial, with the very existence of limited capacity resource pools in doubt (Logan, 1997; Pashler, 1998). Indeed, past claims that performance decrements arise from depleted resource pools have been soundly criticized as being circular, or merely descriptive rather than explanatory (Allport, 1980; Neumann, 1987). How *exactly* are extra attention resources believed to benefit the tracking of objects through occlusions? In contrast, the existence of foveal resource limitations is incontrovertible, and the potential benefits of looking at a target during (or immediately before) an occlusion are many and tangible. For one, the quality of visual information extracted from the fixated target would be improved as a result of the higher foveal resolution, and this might help to prevent it from being confused with another object. The ability to track an object through occlusion also depends on the availability and discrimination of accretion and deletion cues (Scholl & Pylyshyn, 1999; Viswanathan & Mingolla, 2002) and the availability of precise information about the object's trajectory and appearance (Horowitz, Birnkrant, Fencsik, Tran, & Wolfe, 2006). The quality of both forms of information would likely be improved for objects viewed at or near the fovea. Another potential benefit of looking at a to-be-occluded object is that it may be possible to actually pursue the object with gaze through the occlusion event. This brute force solution to the occlusion problem essentially treats the multiple object tracking task as a single object tracking task, with all foveal resources momentarily brought to bear on the single problematic target.

Experimentally, it may be possible to distinguish between these hypothesized sources of tracking assistance by having two occlusions occur at the same time, thereby pitting one against the other. For example, a target shark on the left side of the display might be occluded at the same moment as a target shark on the right side of the display. If it is the dynamic allocation of attention from a supplemental resource pool that is responsible for good occlusion tracking, then it should be possible to funnel some resources to both crises simultaneously, resulting in no performance cost to either target. Of course, if occlusion crises are resolved using rescue saccades, then only one occluded target can be rescued at a time due to the strictly serial nature of eye movements. Costs might therefore be found for the occluded target that could not be rescued. It will also be informative to learn whether costs are observed for the other non-fixated (and non-occluded) targets as well, as this might indicate that the direction of a rescue saccade to one target is accompanied by the withdrawal of attention from the other targets.⁴ In times of crisis, the handful of metaphorical fingers that ordinarily each point to a different target (Pylyshyn, 2001) may very well clench into a single fist, embodied by gaze, that points to only the one target about to enter into an occlusion.

Another direction for future work will be to further specify the information used to summon a rescue saccade. Our data tell us that rescue saccades are triggered by the anticipation of an occlusion, and not by the occlusion itself. We know this because they were frequently observed well in advance of occlusions (POEs), and even on no occlusion trials (NOEs). The fact that rescue saccade rates for the OCE/NOE, POE, and BOE events varied with target–object separation also suggests that rescue saccades may be signaled by the proximity between a target and another object. The importance of target–object proximity for tracking has been highlighted in several recent studies, both as a general factor in determining tracking load effects (Franconeri et al., 2010, 2008) as well as a factor determining the dynamic allocation of attention to tracked targets (Iordanescu et al., 2009). The present study joins the ranks of these others in suggesting that the tracking process also computes and monitors target–object distances for the purpose of triggering rescue saccades. However, there remains one plausible alternative to this proximity signal hypothesis that warrants further testing. Perhaps object motion is represented during tracking (Fencsik, Klieger, & Horowitz, 2007; Iordanescu et al., 2009; Narasimhan, Tripathy, & Barrett, 2009), and the tracking process extrapolates from these motion representations whether the path of a target is likely to intersect that of another object. If so, then it may be the probability of intersection that signals a rescue saccade, and not proximity per se. Future work will attempt to dissociate the path of target motion from its proximity to another object so as to test these two hypotheses.

In conclusion, our data are important in showing that gaze behavior plays an active role in maintaining track on objects. Recent work looking at eye movements during MOT has demonstrated two incontrovertible facts: that there is rich eye movement behavior accompanying MOT, and that these eye movements can be useful in understanding how people accomplish the tracking task. However, the importance of an eye movement-dependent measure does not stop there. We believe that eye movements are not only useful as estimates of how attention is allocated during tracking, they are actually *instrumental* in the tracking process. Specifically, they are a crucial tool that we use to selectively process targets that are in danger of being lost due to momentary tracking crises, such as occlusions from intersecting objects. This proposal differs radically from the largely non-existent role that eye movements were thought to play in tracking just a few years ago (Pylyshyn, 2004; Pylyshyn & Storm, 1988). The operations mediating MOT are probably not confined to the maintenance of indices to targets. Successful tracking likely requires operations spanning a range of sub-tasks, such as the computation of distance relationships between objects (for related ideas, see Franconeri et al., 2010, 2008), and the prediction and detection of potential

tracking problems. These sub-tasks would in turn interact with other behavioral modules, one of which is the eye movement system. From this broader perspective, successful MOT requires the orchestrated contribution of potentially many operations, one of which is the selective signaling for help in the form of rescue saccades.

Acknowledgments

This work was supported by grants from the National Science Foundation (IIS-0527585) and the National Institutes of Health (2-R01-MH063748) to G.J.Z. We thank Todd Horowitz and Werner Schneider for many insightful comments on a draft of this article.

Commercial relationships: none.

Corresponding author: Gregory J. Zelinsky.

Email: Gregory.Zelinsky@stonybrook.edu.

Address: Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500, USA.

Footnotes

¹See <http://mysbfiles.stonybrook.edu/~gzelinsky/Track4-trial.avi> for a movie of a typical Track 4 trial.

²It is reasonable to ask whether the term “rescue” should even be applied to saccades in this context. A “rescue” implies that something has failed and is in need of assistance, whereas the clearly anticipatory saccades observed in the present study in some sense suggest exactly the opposite—that saccades are directed to targets based on a continuous analysis of target–object separation as part of a highly adaptive and proactive tracking process. On this point, we believe that it is useful to distinguish between the specific tracking *operation* and the broader tracking *task*. By our view, the core tracking operation, when considered in isolation from all other tracking-related processes (i.e., in the absence of rescue saccades), would indeed fail as a result of target occlusion. At the level of the tracking task, however, rescue saccades may be correctly viewed as one of many ancillary processes that, when fluidly coordinated with the tracking operation, produce some immunity to the otherwise deleterious effects of target occlusion on tracking behavior.

³Note that because Flombaum et al. (2008) manipulated occlusions using a stationary barrier whereas occlusions in the present study were limited to intersections between moving objects, it is conceivable that our evidence for rescue saccades may be specific to one type of occlusion event, but not the other.

⁴Or stated in less resource-laden language, that the continuous selection of targets for the purpose of tracking is interfered with by the selection of a target for the purpose of making a rescue saccade.

References

- Aivar, M. P., Hayhoe, M. M., Chizk, C. L., & Mruczek, R. E. B. (2005). Spatial memory and saccadic targeting in a natural task. *Journal of Vision*, 5(3):3, 177–193, <http://www.journalofvision.org/content/5/3/3>, doi:10.1167/5.3.3. [PubMed] [Article]
- Allport, D. A. (1980). Attention and performance. In G. Claxton (Ed.), *Cognitive psychology* (pp. 112–153). London: Routledge & Kegan Paul.
- 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. L. (2007). How many objects can you track?: Evidence for a resource limited attentive tracking mechanism. *Journal of Vision*, 7(13):14, 1–10, <http://www.journalofvision.org/content/7/13/14>, doi:10.1167/7.13.14. [PubMed] [Article]
- Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7, 66–80.
- Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, 9, 349–354.
- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 36, 1827–1837.
- Doran, M. M., Hoffman, J. E., & Scholl, B. J. (2009). The role of eye fixations in concentration and amplification effects during multiple object tracking. *Visual Cognition*, 17, 574–597.
- Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 1416–1438.
- Fehd, H., & Seiffert, A. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition*, 108, 201–209.
- Fehd, H., & Seiffert, A. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision*, 10(4):19, 1–13, <http://www.journalofvision.org/content/10/4/19>, doi:10.1167/10.4.19. [PubMed] [Article]

- Fencsik, D. E., Klieger, S. B., & Horowitz, T. S. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception & Psychophysics*, *69*, 567–577.
- Findlay, J. M. (2009). Saccadic eye movement programming: Sensory and attentional factors. *Psychological Research*, *73*, 127–135.
- Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford, UK: Oxford University Press.
- Flombaum, J. I., Scholl, B. J., & Pylyshyn, Z. W. (2008). Attentional resources in tracking through occlusion: The high-beams effect. *Cognition*, *107*, 904–931.
- Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not speed, time, or capacity. *Psychological Science*, *21*, 920–925.
- Franconeri, S. L., Lin, J., Pylyshyn, Z. W., Fisher, B., & Enns, J. T. (2008). Multiple object tracking is limited by crowding, but not speed. *Psychonomic Bulletin & Review*, *15*, 802–808.
- Hayhoe, M. M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*, 188–194.
- Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research*, *38*, 125–137.
- Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1):6, 49–63, <http://www.journalofvision.org/content/3/1/6>, doi:10.1167/3.1.6. [PubMed] [Article]
- 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.
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*, 171–216.
- Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision*, *9*(4):1, 1–12, <http://www.journalofvision.org/content/9/4/1>, doi:10.1167/9.4.1. [PubMed] [Article]
- Khan, Z., Balch, T., & Dellaert, F. (2005). Mcmc-based particle filtering for tracking a variable number of interacting targets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*, 1805–1918.
- Kowler, E., Anderson, E., Doshier, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*, 1897–1916.
- Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*, 3559–3565.
- Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living: Making a cup of tea. *Perception*, *28*, 1311–1328.
- Landry, S. J., Sheridan, T. B., & Yufik, Y. M. (2001). A methodology for studying cognitive groupings in a target-tracking task. *IEEE Transactions on Intelligent Transportation Systems*, *2*, 92–100.
- Logan, G. (1997). The automaticity of academic life: Unconscious applications of an implicit theory. In R. S. Wyer (Ed.), *Advances in social cognition* (vol. 10, pp. 157–179). Mahwah, NJ: Erlbaum.
- Narasimhan, S., Tripathy, S. P., & Barrett, B. T. (2009). Loss of positional information when tracking multiple moving dots: The role of visual memory. *Vision Research*, *49*, 10–27.
- Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 361–394). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology*, *56*, 237–283.
- Pashler, H. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, *80*, 127–158.
- 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., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.
- Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*, 259–290.
- Tombu, M., & Seiffert, A. E. (2008). Attentional costs in multiple object tracking. *Cognition*, *108*, 1–25.
- Viswanathan, L., & Mingolla, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking. *Perception*, *31*, 1415–1437.
- Yang, T., Li, S., Pan, Q., & Li, J. (2005). Real-time multiple objects tracking with occlusion handling in dynamic scenes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, *1*, 970–975.

- Yantis, S. (1992). Multi element visual tracking: Attention and perceptual organization. *Cognitive Psychology*, 24, 295–340.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115, 787–835.
- Zelinsky, G. J., & Neider, M. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition*, 16, 553–566.
- Zelinsky, G. J., Rao, R., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatio-temporal dynamics of visual search. *Psychological Science*, 8, 448–453.
- Zhou, Y., & Tao, H. (2003). A background layer model for object tracking through occlusion. *Proceedings of the IEEE International Conference on Computer Vision*, 2, 1079–1085.