Atten Percept Psychophys (2012) 74:703–715
DOI 10.3758/s13414-011-0260-1

Close encounters of the distracting kind: Identifying the cause of visual tracking errors

Gi Yeul Bae · Jonathan I. Flombaum

Abstract Why can we track only so many objects? We addressed this question by asking when and how tracking errors emerge. To test the hypothesis that many tracking errors are target/nontarget confusions emerging from close encounters, we compared standard multiple-object tracking trials with trials on which a nontarget turned a random color whenever it approached within 4° of a target. This manipulation significantly improved performance by alleviating the correspondence challenge of a close encounter. Two control experiments showed that color change benefits were not merely due to target recovery. Follow-up experiments demonstrated that color change benefits did not accrue monotonically with distance but, instead, seemed to obey a step function; and an additional experiment demonstrated that, without color changes, the frequency of close encounters predicts tracking performance. Taken together, these experiments suggest that uncertainty about target location imposes the primary constraint on tracking, at times causing errors by leading to confusions between targets and nontargets.

Keywords Multiple-object tracking · Attention · Resolution · Capacity limits

Introduction

While it can be hard to define exactly what attention is, it is often defined operationally as involving selection, capacity limits, and mental effort (Kahneman, 1973; Pashler, 1998). Many laboratory tasks evidence characteristic attentional capacity limits, including cuing tasks, demonstrations of change blindness, and visual search. But perhaps none of these tasks evidence the limited capacity of attention as effectively as multiple-object tracking (MOT; Pylyshyn & Storm, 1988). On a typical MOT trial, several identical objects are presented on a screen and move haphazardly for some duration. Observers attempt to mentally track a subset of these objects, a task that requires sustained effort because of the featural similarity among the objects (Fig. 1a). Capacity limits are salient when performing MOT, even subjectively salient; most observers can track no more than three or four objects at once, and in some instances—when objects move quickly, for example—observers appear able to track no more than one object (Alvarez & Franconeri, 2007; Holcombe & Chen, in press). Because of these seemingly severe capacity limits, MOT has become a testing ground for theories of why visual attention is limited in the first place. Why is tracking capacity limited?

The earliest attempt to answer this question coincided with the development of the MOT paradigm by Pylyshyn and Storm (1988), and Pylyshyn's subsequent FINST theory (e.g., 1989, 2001). This line of research began with a theoretical position, arguing that vision requires pointer-like representations that can be used to address features in the world. Critically, these pointers—FINSTs—must be "sticky," able to maintain their contact with an object over time and through motion. In virtue of this stickiness, FINSTs conveniently supply a mechanism for tracking objects. As for capacity limits, Pylyshyn stipulated that limited tracking abilities likely arise from a limited number of FINSTs. In other words, FINSTs are an early version of what has now become known as a *fixed-resource* theory of visual capacity limits. On such views, the visual system possesses a small number of discrete slot, buffer, or FINST-like representations, and capacity limits become

---

G. Y. Bae · J. I. Flombaum (✉)
Department of Psychological and Brain Sciences,
Johns Hopkins University,
Ames Hall / 3400 N. Charles Street,
Baltimore, MD 21218, USA
e-mail: flombaum@jhu.edu

G. Y. Bae
e-mail: gbae1@jhu.edu

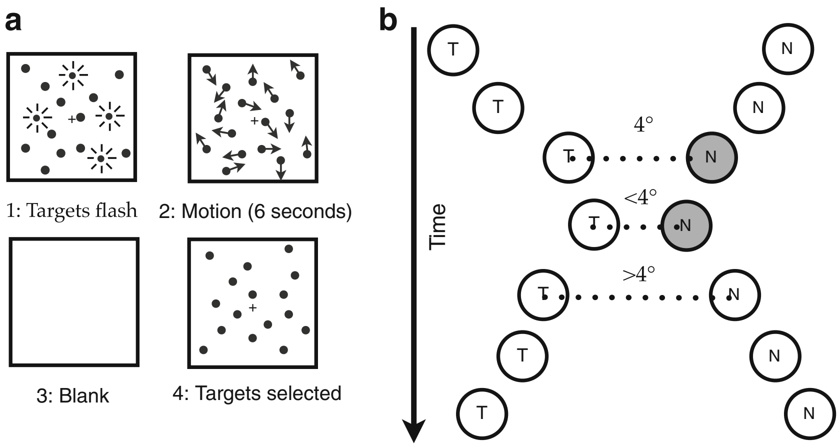

Fig. 1 **a** General multiple-object tracking procedures. First targets flash (1). Next they move for 6 s (2). Finally, all the items disappear (3), and when they reappear, the participants select the items they believe were the targets (4). **b** Schematic depiction of color change interactions in Experiment 1. Nontargets (designated by an "N") changed to a random color whenever they approached within 4° of a target (designated by a "T")

apparent when we try to process more objects than we have representations available (e.g., Drew & Vogel, 2008; Luck & Vogel, 1997). It is worth noting that according to Pylyshyn's original proposal, FINSTs are preattentive and, accordingly, capacity limits may in fact arise outside of attention. While the role of attention in tracking has generated some debate (Scholl, 2009), a discrete capacity limit reconciles easily with attentive (as opposed to preattentive) theories of how tracking works (Cavanagh & Alvarez, 2005).

Recently, many researchers have taken a more task-based approach to understand why tracking is limited, operationalizing the question by asking what kinds of display factors make tracking easier or harder. Alvarez and Franconeri (2007), for example, manipulated a variety of display features, including number of targets, object speed, and the minimum allowed distance between objects. Cleverly, in one experiment, they allowed participants to dynamically set their top tracking speeds for different numbers of targets, leading them to discover that participants could track one or just a handful of objects at high speeds but as many as seven or eight at very slow speeds. Similarly, they found that participants could track more items in relatively uncrowded displays and fewer items in dense ones. These results suggest that, contrary to many fixed-resource theories, there may not be an upper bound on tracking limits and that, in some circumstances, the lower bound can be quite low (Holcombe & Chen, in press). Tracking capacity appears to be a dynamically determined limit that depends on display factors such as object speed and density.

Why do these display factors limit tracking performance? Intriligator and Cavanagh (2001) supplied at least one reason, pointing out that attention possesses limited spatial resolution. Attention cannot select two items at an arbitrarily close distance to one another; instead, there must be some limit on how finely attention can discriminate between and select nearby objects. To measure the resolution of spatial attention and to demonstrate its role in limiting MOT abilities, they carefully evaluated how display density affects tracking performance. To vary density without varying other display factors, they simply varied participants' viewing distances. Longer viewing distances automatically yielded smaller visual angles for the display and the objects, in turn, producing denser displays. (Of course, increasing viewing distance also reduced perceived object size, which may have played a role in their effects.) Results showed that performance declined as a function of viewing distance, suggesting that objects perceived in greater proximity to one another were harder to select and track individually. These results connected the influence of a display factor (density) to tracking performance via a well-defined mental limit, the resolution of spatial attention.

More recently, Franconeri, Jonathan, and Scimeca (2010) have suggested that inter-item proximity may be the only display factor that independently limits tracking abilities. In their view, object speed, for instance, has no independent impact, but it does increase the likelihood that a target and nontarget (or a target and another target) will move in close enough proximity to one another to challenge the resolution of spatial attention. Supporting this theory, they tested participants on the same tracking trials played at different speeds. At faster speeds, the number of close interactions between targets and nontargets remained the same. Performance on faster and slower trials was indistinguishable, suggesting that speed impacts performance only through its influence on inter-item proximity. Similarly, they were able to show that tracking duration—the length of a trial—impairs performance by increasing the likelihood of a close spatial interaction.

Some debate surrounds the interpretation of these and similar results. Specifically, some researchers have argued that a *flexible resource* determines the resolution of spatial attention and that certain display factors, such as speed, independently consume this resource, leading to worse spatial resolution for faster moving than for slower moving objects (Alvarez & Franconeri, 2007; Holcombe & Chen, in press). Some empirical work has also suggested that the resource may best be conceived as working memory (Allen, McGeorge, Pearson, & Milne, 2006). A recent model suggests that it is flexibly allocated and that its allocation is dependent on display factors, including speed (Vul, Frank, Tenenbaum, & Alvarez, 2009). While it remains an open question whether display factors beyond inter-item proximity independently limit tracking, a consensus seems to have emerged that the most proximate cause of tracking limits is the spatial resolution of spatial attention (or working memory).

One consequence of this consensus is that we conceive of tracking errors as events that take place at various times during a trial. Recent work in a related context—how we track through occlusion—has demonstrated how it is possible to study specific kinds of events that take place during a tracking trial and unique aspects of those events that may not be true at all moments during tracking (Flombaum, Scholl, & Pylyshyn, 2008; Zelinsky & Todor, 2010). In the current context, limited resolution means that when targets and nontargets approach very near one another, the targets may become difficult to select independently and, therefore, that nontargets will sometimes become confused with targets. Indeed, this is an explicit component of a recent Bayesian model of multiple-object tracking. In the model, object positions are represented probabilistically, and the model successively compares stored positions with new observations to infer the current positions of targets. Errors arise when a nontarget is mistakenly inferred to be a target and, subsequent to that, the nontarget is tracked throughout the trial (unbeknownst to the participant; Vul et al., 2009). This perspective suggests that with the right manipulations or methods, one should be able to observe target/nontarget confusions as they take place and even to prevent them.

Some prior work has suggested that target and nontarget confusions really do result in nontarget tracking, as opposed to simply the selection of some nontargets at test due to guessing. For example, Sears and Pylyshyn (2000) combined MOT with a probe detection task and found that probes were generally more likely to be detected on targets than on nontargets. When they increased the number of nontargets in a trial, leaving the number of targets the same, the difference in detection rates was reduced, a result that they attributed to participants mistakenly tracking nontargets. Lending further support, on trials with perfect tracking performance, the difference in detection rates remained high, suggesting that, on those trials, nontargets were not tracked, whereas they were on error trials. These results are consistent with the idea that errors involve incorrectly tracking a nontarget for some portion of a trial. O'Hearn, Landau, and Hoffman (2005) added further support to this conclusion and related it to spatial proximity between targets and nontargets. By analyzing error data, they found that incorrectly chosen nontargets were closer to a missed target, on average, than were successfully rejected nontargets. (This was true in a group of Williams syndrome participants, as well as in age-matched controls.) These results suggest that targets and nontargets are more likely to be confused when they are in close proximity.

In the present study, we sought to investigate the specific moments during which tracking errors emerge, the moments when targets and nontargets are close enough to one another to be confusable. According to several current theories, all mental limits that impact tracking ultimately have their impact during these moments (Franconeri et al., 2010; Vul et al., 2009). More specifically, we sought to improve tracking performance via an intervention that would make targets and nontargets less confusable only at those critical moments, as opposed to a global manipulation. We reasoned that if tracking is limited by confusions during close encounters, better performance should be obtained if confusions are somehow prevented. In Experiment 1, we attempted to prevent confusions by supplying distinguishing surface properties to nontargets during the moments that we thought a confusable close encounter might take place. The logic of the experiment was that although the spatial resolution of attention would not allow targets and nontargets to be selected separately, when in close proximity, featural selection could help to keep them separate (Egeth, Virzi, & Garbart, 1984).

Experiment 1: Reducing the hazard in a close encounter

If attentional resolution places the most proximate constraint on object tracking abilities, errors should take place mainly when targets and nontargets approach close enough to one another to make it possible for a nontarget to be selected and then tracked in lieu of a target. Can the hazard created by such close encounters be reduced—in other words, can the frequency of tracking errors be reduced—by providing participants with an ability to favor target selection during close spatial encounters? We reasoned that supplying targets and nontargets with distinguishable surface features might facilitate target selection. To this end, we caused nontargets to change into a distinct random color during any moments that they approached within a critical distance of a target. The critical distance was 4°, chosen on the basis of pilot testing and an assessment of related literature (Franconeri, Alvarez, & Enns, 2007). We chose to change nontarget colors because changing target colors could facilitate recovery of lost targets, as opposed to maintained tracking of a currently selected target (Makovski & Jiang, 2009a, 2009b; we discuss this issue further in Experiments 2 and 3). Overall, the logic of the experiment was straightforward. During most moments of a trial, targets and nontargets were identical, requiring sustained tracking of targets by updating knowledge of their locations over time. At the moments that we expected errors might take place—that is, when a nontarget approached within a close encounter distance of a target—the relevant nontarget took on a random, distinguishable color that was different from all the other identically colored items in the display (Fig. 1b). We called these *color change* trials, and we compared performance on color change trials with that on *standard* trials, wherein targets and nontargets always possessed identical colors. Better performance on color change, as compared with standard trials would constitute evidence that close encounters cause errors by leading to tracking confusions.

Method

Participants A group of 11 Johns Hopkins University undergraduates participated in exchange for course credit. Each had normal or corrected-to-normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Apparatus The experiment took place in a dimly lit soundproof room. All displays were presented on a Macintosh iMAC computer at a viewing distance of approximately 60 cm, such that the display subtended 39.56° × 25.35° of visual angle.

Stimuli and procedure At the beginning of each trial, 16 green disks (diameter, 1.24°) were presented on a black background in randomly selected positions. A white fixation cross (0.5° × 0.5°) was present in the center of the display and remained present throughout each trial. Four of the green disks flashed on and off 4 times over the course of 2 s, identifying them as targets. All the disks then moved haphazardly through the display for 6 s. Motion speed was manipulated across trials such that one third of all trials were "fast" (6.69°/s), "medium" (5.43°/s), and "slow" (4.18°/s). Objects mostly moved in a fixed direction, but on each frame, each object had a 2% chance of turning. When an object did turn, the angle was selected randomly from between 1° and 359°. Additionally, if any two objects came within 2° of one another, they were repelled, forcing turns so that two objects could never be closer than 2°, measured center to center.

Half of all trials were standard trials, wherein all the moving disks remained green. In contrast, half of all trials were color change trials. On color change trials, nontargets turned a distinct, random color whenever they approached within 4° of a target. These colors were assigned by randomly choosing R, G, B values between 0 and 255. When a nontarget moved more than 4° away from a target, it changed back to green. When more than one nontarget was within the critical distance of any targets, those nontargets did not take on the same colors. In both trial types, all items stopped moving after 6 s, at which point all the items disappeared from the screen for 1 s to prevent observers from using color information at the last moment of motion. All circles then reappeared, colored in green, and observers used the mouse to select all the items they thought were targets. Selected disks turned yellow to prevent participants from choosing the same item twice. After they selected a total of four items, the true targets turn red, providing feedback. Each observer completed a total of 60 trials. The experiment lasted approximately 25 min. Although eye movements were not monitored, observers were instructed to fixate on the white fixation cross.

Instructions Participants were explicitly made aware of the fact that on half of the trials, nontargets would change random colors whenever they came within a fixed distance of targets. The following text was included within a longer set of instructions that were read by participants: ". . . then all the circles will turn green, and begin moving randomly around the display for 6 seconds. In some trials, only nontargets will change color whenever the distance to any target is less than some fixed amount so that you can easily distinguish between targets and nontargets. Your job is to do your best to track all the targets for 6s."

Results and discussion

Figure 2 displays mean tracking accuracy as a function of speed and color change condition. A two-way ANOVA confirmed significant main effects of both speed and color change. Performance decreased as speed increased, $F(2, 20) = 14.22$, $p < .001$, $\eta^2 = .468$, and it was improved when nontargets changed color, $F(1, 10) = 21.885$, $p = .001$, $\eta^2 = .531$.

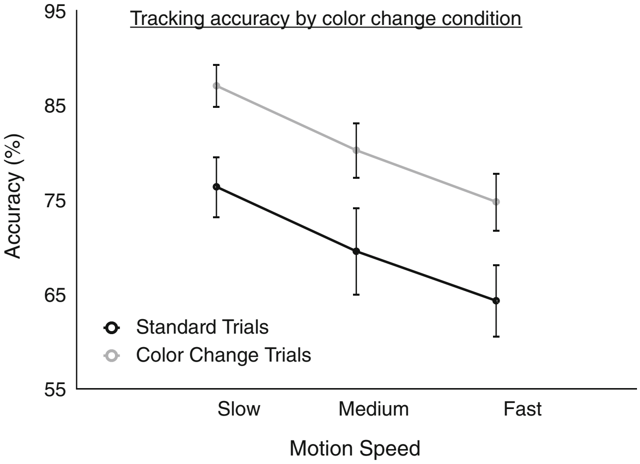

Fig. 2 Mean tracking accuracy as a function of speed and color change condition in Experiment 1. Error bars indicate ±1 standard error of the mean

There was no significant interaction between color change condition and speed, with a constant advantage of about 10% for color change trials at all speeds, *F*(2, 20) < 1.

These results supply clear evidence that tracking errors are caused by close spatial encounters between targets and nontargets. In other words, under standard tracking conditions, a correspondence challenge arises when targets and nontargets come too close to one another: With limited spatial resolution, participants are unable to distinguish targets from nontargets, and they may end up selecting and tracking a nontarget by mistake. Changing nontarget colors reduced the impact of these interactions by supplying a basis upon which participants could preferentially maintain targets, despite limited spatial resolution.

If tracking is most proximately limited by spatial resolution and our color change manipulation alleviated the challenges that emanate from close spatial encounters, one may wonder why performance was not perfect on color change trials. In Fig. 3, we display the proportions of trials on which participants successfully identified between one and four targets. This is to emphasize the size of the effect induced by our color change manipulation. On standard trials, participants tracked perfectly—identifying four targets—on only 26% of all trials, whereas on color change trials, this number nearly doubled to 47%. It is also likely that at least some poorly performed trials reflected general inattention, either at the start of a trial (i.e., when targets were flashing) or for extended durations during a trial. Obviously, errors induced by general inattention would not be eliminated by our color change manipulation, meaning that we would expect some erroneous trials to remain no matter how well our color change manipulation prevented item confusions.

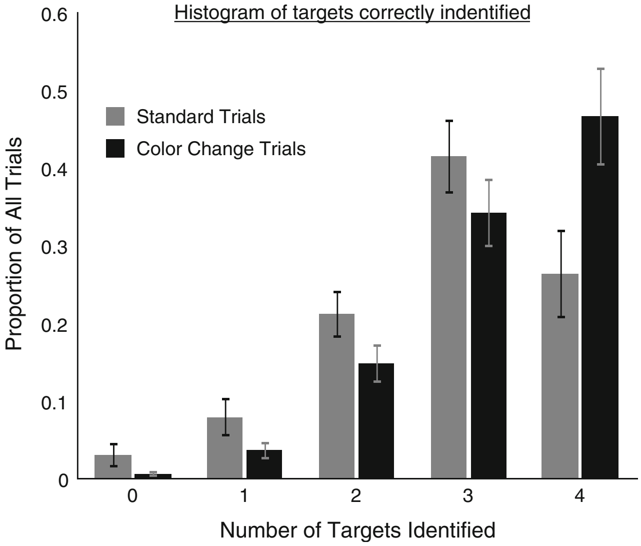

Fig. 3 Proportions of trials on which zero to four targets were correctly identified as a function of color change condition in Experiment 1. Error bars indicate ±1 standard error of the mean

Finally, we expect that color changes do not always prevent confusion during a close encounter, just as not all close encounters result in confusions. Instead, each of these happens probabilistically (Vul et al., 2009). This perspective is consistent with the general fact of declining performance as a function of speed, since higher speeds should produce more close encounters (Franconeri et al., 2010). Thus, the observed advantage of 10% across speeds is a consequence of both more individual close encounters taking place and more individual confusions prevented as speed increases.

Experiment 2: Are color change benefits caused by target recovery?

Experiment 1 evidenced a tracking benefit when nontargets took on unique colors within a 4° distance from a target. We argued that this demonstrates how correspondence problems during close spatial encounters typically lead to tracking errors. But previous work might suggest an alternative interpretation, one based on the recovery of already lost targets (Horowitz et al., 2007; Makovski & Jiang, 2009a, 2009b; St.Clair, Huff, & Seiffert, 2010). Perhaps this is all that our color change manipulation achieved? In one previous study, for instance, targets and nontargets each possessed distinguishing surface features as groups (e.g.. targets were all green, and nontargets were all red; Makovski & Jiang, 2009a), and the benefit of unique surface features was larger when the minimum distance allowed was shorter (Makovski & Jiang, 2009b). Broadly speaking, when these group-defining colors were relatively permanent, participants' performance was significantly better. But, because benefits often did not manifest when these group-defining features changed rapidly and unpredictably or when they failed to define groups exhaustively, the results suggested that participants could use the surface properties to realize that they were tracking a nontarget and then find, or recover, the target(s) that they had lost (on the basis of knowledge of their surface properties). It has also been demonstrated that participants can switch to tracking a new set of items during the course of a trial (Wolfe, Place, & Horowitz, 2007). In the present experiments, one might be concerned that our color change manipulation did not facilitate online tracking mechanisms by disambiguating targets and nontargets. Instead, perhaps participants just figured out that if they had lost a target, they could recover it by looking for a green item next to a uniquely colored item.

To rule out this possibility, we directly investigated how well participants could use changing nontarget features to find a missing target. Experiment 1 was different from prior experiments on target recovery because, among other things, only nontargets changed colors, they changed colors at different times from one another, and each instance of a color change involved taking on a new and unique color. Therefore, to use nontarget features in our experiment, participants would have had to (1) realize that they had lost a target and then (2) use the fact that nontargets changed color close to targets in order to infer the identity of the target that they had lost. In the present experiment, we investigated the possibility of achieving step 2—that is, the possibility of inferring a missing target's identity—by revealing three targets at the start of the trial and then instructing participants both to track those three and to find and track the fourth, "hidden" target. The dynamics of nontarget color changes in this experiment were the same as in Experiment 1, such that nontargets changed colors at the appropriate times when they moved close enough to the hidden target (or to any other target). If color change benefits are due solely to target recovery, participants should be able to use the color changes to find and track the missing target on these *hidden target* trials.

Method

Participants A new group of 11 Johns Hopkins University undergraduates participated in exchange for course credit. Each had normal or corrected-to-normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Apparatus, stimuli, and procedure This experiment was identical to the medium speed condition of Experiment 1, except as follows. We employed three tracking conditions. Standard and color change conditions were identical to those employed in Experiment 1. We also included a new, third condition, the *hidden target* condition. This was identical to the color change condition in Experiment 1, except that only three targets flashed at the beginning of each trial, so that observers did not know which of the remaining 13 items was the fourth target. (Nontargets changed color during motion when they were close enough to this "hidden" target, as in Experiment 1.) The experiment was conducted in two blocks. In the first block, standard and color change trials were presented randomly intermixed. The second block included only hidden target trials. Participants completed 40 trials plus 10 practice trials in the first block and 20 trials plus 5 practice trials in the second block. The total duration of this experiment was around 30 min.

Results and discussion

Figure 4 displays mean tracking accuracy as a function of tracking condition. A one-way ANOVA confirmed a significant main effect of tracking condition, $F(2, 20) = 25$, $p < .001$, $\eta^2 = .71$. Planned contrasts showed that tracking performance was reliably better for the color change condition than for the standard condition, $F(1, 10) = 12.83$, $p < .01$, replicating the color change benefit observed in Experiment 1. Most importantly, performance in the hidden target condition was reliably worse than performance in the color change condition, $F(1, 10) = 33.08$, $p < .01$, and crucially, worse than that in the standard condition, $F(1, 10) = 19.33$, $p < .01$.

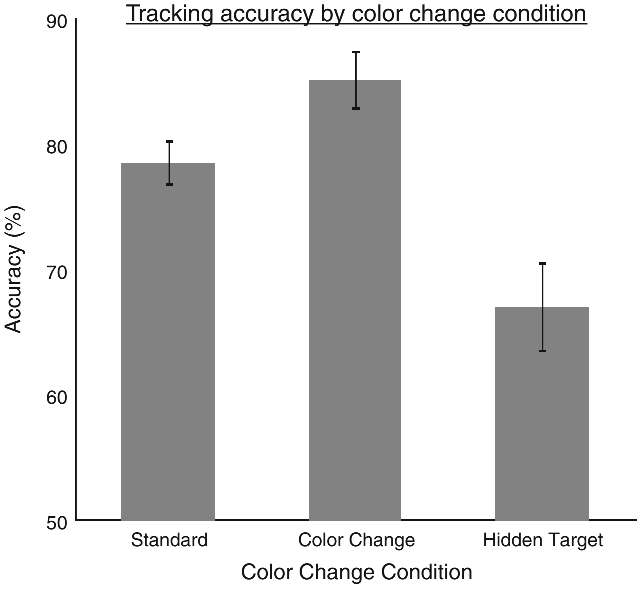

Fig. 4 Mean tracking accuracy as a function of condition in Experiment 2. Error bars indicate ±1 standard error of mean accuracy. Standard trials and color change trials were identical to those in Experiment 1, while hidden target trials were those in which only three targets were revealed at the start of a trial. Color changes took place on both color change trials and hidden target trials

It might be expected that performance on the hidden target trials would be worse than performance on standard trials, since performance on standard trials depended on how often participants lost one item (or more), while performance on hidden target trials depended on how often they found the missing item. In other words, on hidden target trials, we set up our participants to "catch up" to performance on standard trials, and thus, this experiment was perhaps not an ideal test of how easily they can use color changes to recover targets. In Experiment 3, we report a further test of the possibility that participants engage in target recovery. But first, here, we note that performance on hidden target trials was surprisingly low, suggesting that attempting target recovery actually makes it harder to track currently tracked items. Indeed, average performance on hidden target trials was below 75% correct, which would be at or below chance (depending on expectations about successful guessing) had they generally been able to track just three targets. Thus, participants seemed unable to successfully track as many targets as they were initially supplied while also attempting to use color changes to find a target. In Experiment 1, had they used target recovery to find lost targets, they would have needed to both use those cues and continue to track any targets that had not been lost.

Experiment 3: Further evidence against target recovery

As we noted in the discussion above, Experiment 2 was perhaps not an ideal test of the possibility of target recovery, since it asked that participants use color changes to catch up to performance in a standard four-target trial from a starting point with only three targets. Accordingly, in the present experiment, we considered how much color changes can increase the number of items one is able to track given the same starting point. We compared two conditions.[1] On standard trials, there were no color changes, and participants were simply asked to track three targets, all of which were identified at the start of the trial. Second, we included hidden target trials that were identical to those used in Experiment 2. Participants were shown three targets at the start of each trial and then were told to use color changes to find one additional hidden target. In this way, the standard trials supplied a baseline: How well can a participant track three targets? And we could then ask how likely a participant was to find a hidden target while maintaining the same level of tracking performance for the three original targets. On the basis of the results of Experiment 2, we predicted that participants would not be able to track any more items on the hidden target trials than on the standard trials of this experiment, a result that would demonstrate the ineffectiveness of color changes as cues for lost target recovery.

Method

Participants A new group of 8 Johns Hopkins University undergraduates participated in exchange for course credit. Each had normal or corrected-to-normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Stimuli, apparatus, and procedure This experiment was identical to Experiment 2, with two exceptions. (1) Standard trials included only three targets (and the requirement to identify only three items at test). (2) Color change trials were not included. However, hidden target trials were included, which were identical to those in Experiment 2. Each participant completed one block of standard trials and one block of hidden target trials (order counterbalanced across participants). Each block included 20 trials, amounting to a total of 40 trials in this experiment.

Results and discussion

Performance on standard trials was reasonably good, 84% on average across the 8 participants. As was expected, on the basis of the results of Experiment 2, performance on hidden target trials was relatively poor—62%, on average. We note, however, that performance in these two trial types should not be compared directly, since they included different numbers of targets. Instead, we computed the probability of correctly identifying three or more targets across all trials for each trial type. There was no significant difference between the two trial types, $t(7) = 0.01$, $p > .05$. Participants identified three or four targets on 60.6% of hidden target trials, and three targets on 61.2% of standard trials. Figure 5 displays the likelihood of identifying zero to four targets in each condition of this experiment. Participants reported all four targets correctly on only 6.8 % of all hidden target trials. Guessing correctly on the hidden target would involve chance performance of 7.6% (1/13). These results demonstrate directly the difficulty of using nontarget color changes to infer the identity of a target *while also tracking other objects,* which would be necessary if color changes were used for target recovery in Experiment 1. Therefore, these results suggest that color change benefits are not due to target recovery.

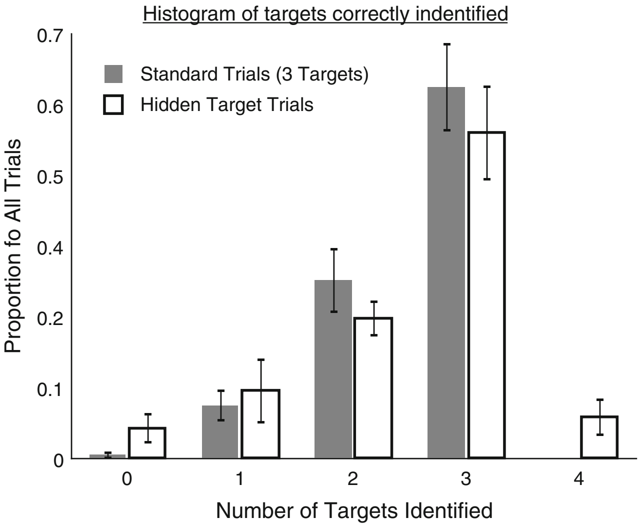

Fig. 5 Proportions of trials on which zero to four targets were correctly identified as a function of condition in Experiment 3. Error bars indicate ±1 standard error of the mean

---

[1] We thank Todd Horowitz and an anonymous reviewer for suggesting Experiment 3.

Experiment 4: At what distance do objects become confusable?

This experiment explored the impact of our color change manipulation as a function of inter-item distance. To this end, we systematically varied the distance at which color changes took place. We reasoned that the effect of color change by distance should depend on the underlying resolution of spatial attention.

Method

Participants A separate group of 11 Johns Hopkins University undergraduates participated in exchange for course credit. Each had normal or corrected-to-normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Stimuli, apparatus, and procedure This experiment was identical to Experiment 1, with two exceptions. (1) Items always moved at a speed of 5.43°/s (previously called "medium"). (2) Only one sixth of all trials were "standard." All others included color changes that took place when targets and nontargets were within 2.5°, 3.0°, 3.5°, 4.0°, or 4.5° of one another. There were an equal number of trials using each of these distances. Each participant completed a total of 60 trials.

Results and discussion

Figure 6 displays mean tracking accuracy as a function of color change distance. A one-way ANOVA revealed a significant main effect of change distance, $F(5, 50) = 4.87$, $p = .001$, $\eta^2 = .32$. A linear contrast was significant, $F(1, 10) = 19.18$, $p < .01$, accounting for 78% of the variance. But inspection of the graph suggested a major improvement in performance when changes took place at a distance of 3.5° or more. A post hoc analysis confirmed this intuition. A Scheffé correction compared standard, 2.5°, and 3.0° conditions with 3.5°, 4.0°, and 4.5° conditions, revealing a significant difference, $F(1, 10) = 23.14$, $p < .0001$. This contrast accounted for 95% of the variance, far more than the linear contrast. These results imply that the color change benefit was effective only when the distance between a target and a nontarget was about 3.5°. Surprisingly, there was a distance outside which the color change manipulation was not helpful, and when color changes were provided below a critical threshold, they had no impact as well.

To replicate this counterintuitive effect, we repeated Experiment 4 with a new group of 15 observers and object speeds that were 30% faster (Experiment 4b). Figure 7 displays mean tracking accuracy as a function of color change distance in the replication experiment. We found the same basic pattern in our data—that is, a clear inflection point; only, in this experiment, the critical distance appeared to be 4.0°. A linear contrast using Scheffé correction was not significant, $F(1, 14) = 10.11 < F_{\text{crit}} = 11.72$, but a contrast between a group including standard, 2.5°, 3.0°, and 3.5° trials, as compared with a group including 4.0° and 4.5° trials, was significant, $F(1, 14) = 12.73 > F_{\text{crit}} = 11.72$, explaining 77% of the variance.

Taken together, these results suggest that there is a critical distance that depends on the resolution of spatial attention at which close encounters become hazardous. Of course, we expect that attentional resolution becomes poorer at greater

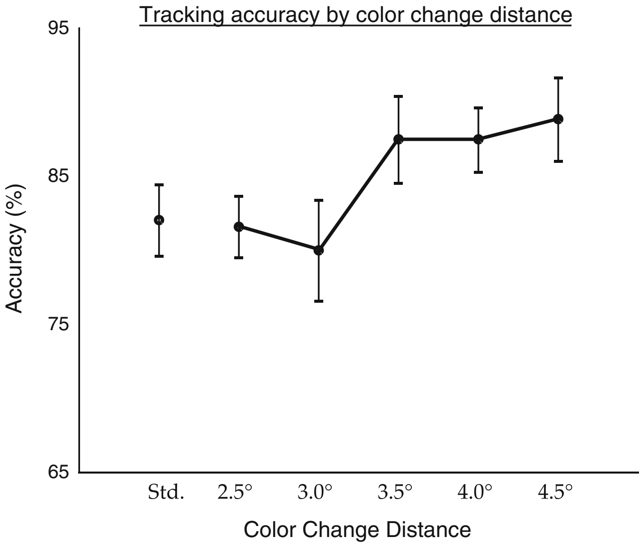

Fig. 6 Mean tracking accuracy as a function of color change distance in Experiment 4. Error bars indicate ±1 standard error of the mean. "Std." refers to standard trials on which no color changes took place. Experiment 4 employed "medium" speeds from Experiment 1

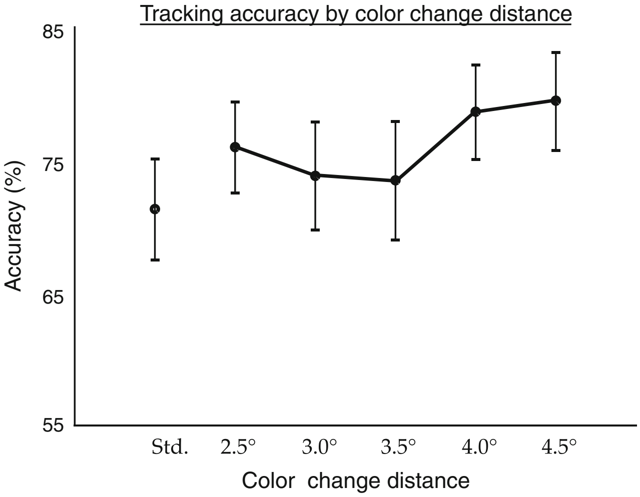

Fig. 7 Mean tracking accuracy as a function of color change distance in Experiment 4b. Error bars indicate ±1 standard error of the mean. "Std." refers to standard trials on which no color changes took place. Experiment 4b employed "fast" speeds from Experiment 1

distances from eccentricity and obeying Bouma's rule (Bouma, 1970; Pelli, Palomares, & Majaj, 2004; Whitney & Levi, 2011). Indeed, this is an explicit feature of a recent tracking model that depends on spatial precision (Vul et al., 2009). Thus, we emphasize here not the specific numbers we discovered, 3.5°–4.0°, but the step function that seemed to describe the influence of color changes on performance. We suspect that the specific numbers reflect an average of critical distances at different eccentricities. However, the step functions reveal the dynamics of object confusions. Specifically, confusions happen at only those moments when targets and nontargets are confusable, and providing disambiguating information when they are not confusable (when they are far enough apart) does not help. Moreover, attempting to disambiguate objects only when they are very near one another—presumably, after they have already become confusable—does not help either.

Experiment 5: Are there color change benefits at large distances?

In Experiment 4, we tested the impact of color changes to nontargets at various distances, but in such a way that if a color change was induced at one distance, it remained present at all shorter distances. As a result, those data may not warrant the inference that no benefits can be gained when color changes are induced at greater distances. In Experiment 4, it appeared that there was no difference in performance for changes that were initiated at 3.5°, as compared with 4.5°, suggesting that there was no added benefit to providing color information well before a close encounter. But if changes initiated at 3.5° brought performance up to ceiling, we simply may not have been able to observe the additional gains accrued by initiating changes at longer distances. Accordingly, in the present experiment, we initiated changes to nontargets at a distance of 5.5°, but those changes were undone when a nontarget moved within 4.5° of a target.

Method

Participants A new group of 10 Johns Hopkins University undergraduates participated for course credit. Each had normal or corrected-to-normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Apparatus, stimuli, and procedure This experiment was identical to Experiment 1, except that on color change trials, color changes to nontargets were induced whenever a nontarget was within 4°–5.5° of a target.

Results and discussion

Figure 8 displays mean tracking accuracy as a function of speed and color change condition. A two-way ANOVA revealed no main effect of color change condition on tracking accuracy, $F < 1$. There was a main effect of speed, $F(2, 18) = 17.588, p < .001, \eta^2 = .22$, however, and an interaction between speed and color change condition, $F(2, 18) = 3.844$, $p = .04$, $\eta^2 = .08$. The interaction was caused by better performance on standard trials, as compared with color change trials, at medium speed, but a simple main effect using Scheffé correction was not significant, $F(1, 18) = 6.96 < F_{\text{crit}} = 22.06$. These results confirm the impression left by Experiment 4, that there are no gains to changing nontarget colors when those nontargets are outside a confusable distance of a target.

Experiment 6: Does performance usually depend on the number of close encounters?

The experiments conducted so far have suggested that tracking is limited by correspondence problems that emerge when objects approach within a critical distance of one another. Changing nontarget features during these critical moments demonstrated that correspondence mistakes can be prevented. In two experiments, performance improved only when color changes were induced at a fixed distance from a target, and surprisingly, in Experiment 5, no benefit accrued for changes that took place at further distances.

In the present experiment, we sought to demonstrate that close encounters at a distance of about 4° predict tracking

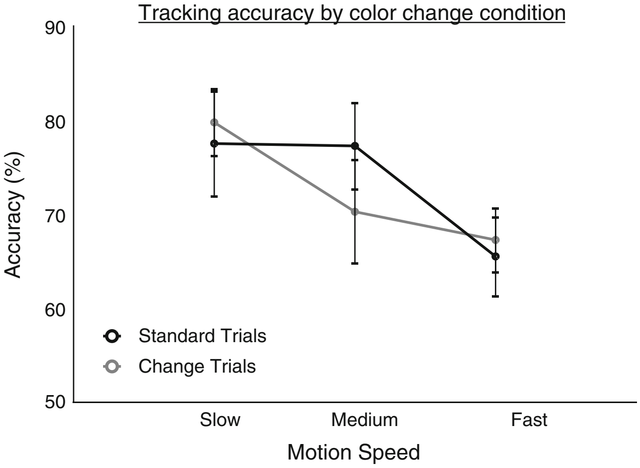

Fig. 8 Mean tracking accuracy as a function of speed and color change condition in Experiment 5. Error bars indicate ±1 standard error of the mean. Color changes in this experiment took place only when a nontarget was within 4°–5.5° of a target

performance in a standard MOT task, one without any color changes. As participants tracked, we counted the number of close encounters that took place between targets and nontargets at a distance of 4°. We expected that, trial by trial, the number of close encounters that took place would predict tracking performance.

Method

Participants A separate group of 11 Johns Hopkins University undergraduates participated in exchange for course credit. Each had normal or corrected-to-normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Apparatus, stimuli, and procedure This experiment was identical to Experiment 4, except as follows. (1) There were never any color changes; every item in the display was indistinguishable from every other item throughout a trial. (2) In order to produce a wide range of number of close encounters, we tested participants on trials with four or five targets among 10 or 12 nontargets. Half of all trials included four targets, while half included five. Counterbalanced with this manipulation, half of all trials included 10 nontargets, while half included 12.

Data analysis Because we included trials with four and trials with five targets, we separated the data in this experiment by target load. (Percent correct cannot be compared and combined for different target loads.) Additionally, we *collapsed data across participants* and binned trials into nine bins on the basis of the number of close encounters, for each target load, so that each bin included at least 10% of all trials (i.e., 35 trials). This was done because there were relatively small numbers of trials with many and few close encounters. The outcome was that we had nine total bins for each target load, each with a range of number of close encounters. The independent variable in the correlational analyses, below, and on the *x*-axes in Fig. 10, was the average number of close encounters in each bin.

Results and discussion

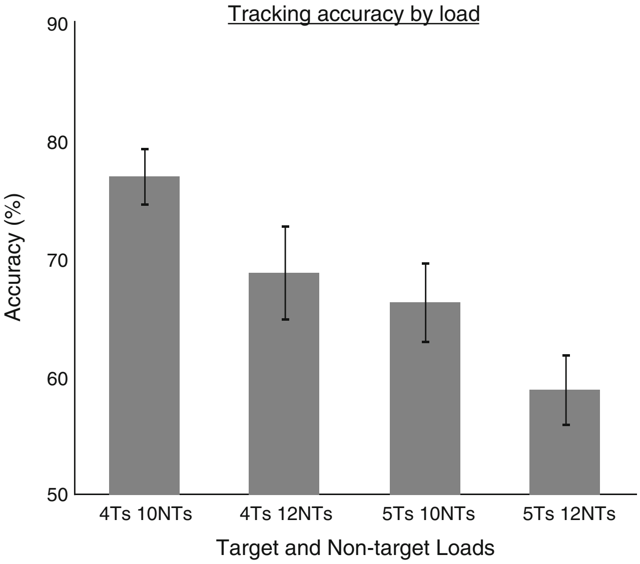

Fig. 9 Mean tracking accuracy as a function of target (T) and nontarget (NT) loads in Experiment 6. Error bars indicate ±1 standard error of the mean

Figure 9 displays mean tracking accuracy as a function of number of targets and nontargets. Figure 10 displays the relationship between percentage of targets identified (tracking performance) and the average number of close encounters in each bin by tracking load. We fit power laws to these results because inspecting the data suggested these fits and because, once lost, a target cannot be lost again. These relationships were significant for both target load four, $t(7) = 3.77$, $p < .01$, $R^2 = .671$, and target load five, $t(7) = 4.24$, $p < .01$, $R^2 = .721$, suggesting that power law declines in performance as a function of the frequency of close encounters. Linear fits were also significant, $t(7) = -3.00$, $p = .020$, $R^2 = .562$, for a target load of four, and for a target load of five, $t(7) = -3.68$, $p < .01$, $R^2 = .660$. These results furnish proof of concept that the step function inflection of about 4° revealed in Experiment 4 can be used to understand tracking performance in a standard version of an MOT task.

General discussion

In the present study, we sought to supply evidence that tracking errors in MOT are caused by close spatial encounters in which imprecise spatial knowledge causes targets and nontargets to become confusable. Experiment 1 supplied this evidence by improving tracking performance through a color change manipulation that reduced the confusability of targets and nontargets whenever they were engaged in a close encounter. Experiments 2 and 3 demonstrated that participants could not use nontarget changes to infer the identity of a hidden target while also tracking additional items, excluding a target recovery account of the results of Experiment 1. Experiment 4 varied the distance at which nontargets changed colors and suggested that items become confusable when they are about 4° from one another, at least when four targets are tracked. Moreover, this experiment suggested that color change benefits do not accrue when changes are supplied at very short distances, perhaps because errors accrue quickly within the margin between the distance at which objects become confusable and smaller values. It is worth noting here that this experiment also

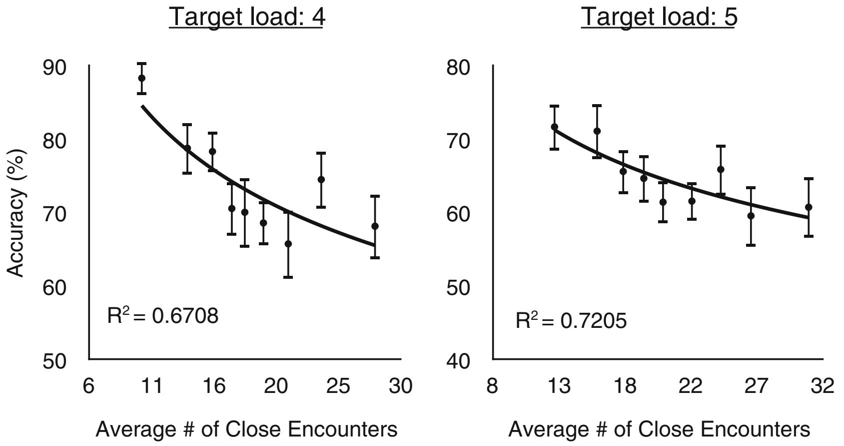

Fig. 10 Percentage of targets correctly identified as a function of the number of average close encounters (in nine bins) in Experiment 6. Close encounters were counted when a nontarget approached within 4° of a target. Bins were created to each include at least 10% of the data, and the average number of close encounters in each bin is the independent variable (see the Method section in Experiment 6). Error bars indicate ± 1 standard error of the mean. Curves reflect fitted power law functions

supplied a kind of evidence against a target recovery account of the main color change effect. Specifically, if participants use nontarget color changes to find a previously lost target, the ability to do so should only be enhanced at closer distances, where color changes are providing more precise information about the identity of a lost target. Experiment 5 revealed no advantage for changes induced only at large distances (and not persisting at shorter distances). In other words, this experiment suggested that items that are not confusable in the first place–because they are far enough apart from one another—cannot be rendered any less confusable. Finally, Experiment 6 demonstrated that the distances at which color changes had their impacts in Experiments 1–5 could be used to predict tracking performance on trials when no color changes were supplied.

Overall, these experiments demonstrated that tracking errors can be reduced or prevented, that the distances at which targets and nontargets become confusable can be measured, and that the measured distances obtained relate to performance limits in standard MOT tasks. Future studies should explore, in even greater detail, why specific distances are the ones at which objects become confusable and the relationship between these distances and the time or effort that may be needed to resolve potential confusions favorably. Additionally, future studies can employ our color change manipulation to investigate the distances at which objects become confusable as a function of, for example, speed, eccentricity, target load, and object size, among other factors.

What constrains tracking performance?

Since the seminal experiments of Pylyshyn and Storm (1988), perhaps the central question about object tracking has been the following: “Why is it limited in the first place?” Pylyshyn took a theoretically motivated approach, arguing early on that it is limited because we possess only a limited number of discrete pointer-like representations, FINSTs. More recently, numerous studies have addressed this question by cataloging display features that influence tracking performance, including object speed, trial duration, target load, number of nontargets, and the spatial separation between items (e.g., Franconeri et al., 2010; Franconeri, Lin, Pylyshyn, Fisher, & Enns, 2008; Oksama & Hyönä, 2004; Sears & Pylyshyn, 2000; Shim, Alvarez, & Jiang, 2008). Looking through this list, it becomes apparent that these display features may share a more proximate reason for limiting tracking performance; namely, they influence the likelihood that targets and nontargets will come in close enough proximity to one another to become confusable. Several studies have supplied evidence that confusions sometimes take place, resulting in the tracking of nontargets (O'Hearn et al., 2005; Sears & Pylyshyn, 2000), and recently Franconeri and colleagues (2010) have suggested that inter-item spacing accounts for all tracking limits, even providing striking evidence that other factors can be reduced to interactions with spacing. The present results fit well with this point of view, supplying evidence that confusions during close encounters constitute the most proximate causes of tracking errors by demonstrating that such confusions can be prevented.

Previous work has left ambiguous why close spatial encounters lead to tracking errors. Franconeri and colleagues (2010), for instance, argued that target and nontarget representations include inhibitory surrounds that engage in destructive interference when items approach in close proximity, including when targets approach within close proximity of one another. Although our study was not designed to test such an account, it does not accord well with the present results. Specifically, it is not obvious why changing a nontarget's color would affect the center–surround inhibitory structure of target and nontarget representations and their resulting interactions. Franconeri and colleagues also argued that targets interfere with one another when they approach in close proximity. But our manipulation likely had no impact on how such encounters were negotiated, although it improved performance nonetheless. It may be that our results can be reconciled with an account of tracking limits that relies on inhibitory interactions, and it may also be that some of the errors in our experiments were caused by target–target interactions. But these possibilities do not seem necessary to understand the results presented.

Moreover, prior work has demonstrated clearly that attentional resolution is limited (Intriligator & Cavanagh, 2001; Tsal & Bareket, 2005) and that the precision of spatial working memory is limited (Bays & Husain, 2008). Put simply, human spatial knowledge is, and really must be, imprecise. As a consequence, nearby items should often be confusable (Moore, Lanagan-Leitzel, & Fine, 2008), particularly the more similar they are in appearance. If confusions between targets and nontargets are to be expected because of imprecise spatial knowledge, it seems unnecessary to hypothesize additional mechanisms to account for errors emerging during close spatial encounters. At a minimum, we would want to quantify the number of errors that can be accounted for on the basis of confusions, alone, in order to know how much variance remains to be accounted for by other mechanisms. Doing this will be challenging given that failures due to close spatial encounters likely emerge probabilistically. Indeed, a recent model of MOT accounts for many tracking phenomena simply by formalizing the task as probabilistic inferences about target identities given uncertain spatial knowledge (Vul et al., 2009). Our results are consistent with this model, identifying the moments during which inferences about target identities are most likely to fail.

The nature of multiple-object-tracking commodity limits

While recent theories regarding tracking limits differ considerably from earlier approaches, all theories to date seem to share an assumption that, at some point, tracking is a resource-limited process. But newer theories have conceived of the underlying resource differently, as a flexible, continuous commodity, rather than a fixed and discrete one. According to such flexible-resource accounts, a limited commodity becomes consumed by targets, and it provides less representational resolution the more times it is divided (Horowitz & Cohen, 2010). Accordingly, when more targets are tracked, each is tracked with less spatial precision, and perhaps features such as speed, crowding, and the number of nontargets also consume limited resources, resulting in their own contribution to declining spatial precision in target representations (Alvarez & Franconeri, 2007; Bettencourt & Somers, 2009; Holcombe & Chen, in press).

Crucially, while all prevailing theories of multiple-object tracking assume commodity limits that constrain performance, fixed and flexible theories make very different predictions about when and how errors emerge. Fixed-resource theories, like Pylyshyn's FINSTs (1989, 2001), but including other views as well (e.g., Drew & Vogel, 2008; Luck & Vogel, 1997) predict that mostly random errors should arise, probably at the start of a trial, because more targets are presented to track than can be accommodated by a limited number of discrete representations. In contrast, flexible-resource theories predict that errors should arise at various points during a trial, because degraded representational precision leads to confusions among tracked objects (e.g., Alvarez & Franconeri, 2007; Horowitz & Cohen, 2010; Vul et al., 2009). A vigorous debate concerning fixed and flexible resources and the specific kinds of errors they predict has appeared over recent years in the visual working memory literature (Alvarez & Cavanagh, 2004; Awh, Barton, & Vogel, 2007; Bays & Husain, 2008; Fukuda, Awh, & Vogel, 2010; Zhang & Luck, 2008).

The present results are problematic for fixed-resource views. Our experiments demonstrate that errors arise at various points in a trial and that they are caused by specific kinds of interactions between objects, not simply an inability to represent some objects, causing them to be left out at the start of a trial. We acknowledge, however, that it may be that both kinds of errors emerge and that multiple kinds of limits exist—limits on both the total number of representable items and the precision of items. What this study demonstrates, then, is the possibility of directly studying the kinds of errors that produce tracking performance and their microgenesis. The recent move toward explicit and quantitative models of how tracking takes place should be followed by a move toward observing and preventing, when possible, the specific kinds of errors that the models predict.

A computational-level analysis of multiple object tracking

One of the enduring mysteries in human visual cognition concerns the fact that it often appears limited. In his influential work on approaches to studying vision, Marr (1982) provided at least one explanation for this mystery: that visual cognition may be limited because, at the computational level, it faces genuine and insurmountable epistemic limits. The visual system cannot know about the world with certainty. The present results point to a pair of computational limits that make human tracking abilities error prone. First, we are incapable of knowing the positions of objects with certainty. Second, we are incapable of solving correspondence problems with certainty. As a result, humans will eventually commit errors when tracking multiple visual objects.

Author Note We thank Justin Halberda and members of the Attention and Cognition lab at Johns Hopkins University for helpful comments and conversations.

References

Allen, R., McGeorge, P., Pearson, G., & Milne, A. (2006). Multiple-target tracking: A role for working memory. *Quarterly Journal of Experimental Psychology, 59*, 1101–1116.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*, 106–111.

Alvarez, G., & Franconeri, S. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*(13, Art. 14), 1–10.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items, regardless of complexity. *Psychological Science, 18*, 622–628.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*, 851–854.

Bettencourt, K., & Somers, D. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision, 9*(7, Art. 9), 1–11.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*, 177–178.

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

Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *Journal of Neuroscience, 28*, 4183–4191.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 32–39.

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., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected at once? *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1003–1012.

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. Y., Pylyshyn, Z. W., Fisher, B., & Enns, J. T. (2008). Evidence against a speed limit in multiple object tracking. *Psychonomic Bulletin & Review, 15*, 802–808.

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology, 20*, 177–182.

Holcombe A. O., & Chen, W. Y. (in press). Tracking a single fast-moving object exhausts attentional resources. *Cognition*. doi:10.1016/j.cognition.2011.10.003

Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception, & Psychophysics, 72*, 1765–1775.

Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics, 69*, 172–184.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*, 171–216.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281.

Makovski, T., & Jiang, Y. (2009a). Feature binding in attentive tracking of distinct objects. *Visual Cognition, 17*, 180–194.

Makovski, T., & Jiang, Y. (2009b). The role of visual working memory in attentive tracking of unique objects. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1687–1697.

Marr, D. (1982). *Vision*. New York: Freeman.

Moore, C. M., Lanagan-Leitzel, L. K., & Fine, E. M. (2008). Distinguishing between the precision of attentional localization and attentional resolution. *Perception & Psychophysics, 70*, 573–582.

O'Hearn, K., Landau, B., & Hoffman, J. (2005). Multiple object tracking in people with Williams syndrome and in normally developing children. *Psychological Science, 16*, 905–912.

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

Pashler, H. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*(12, Art. 12), 1136–1169.

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

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*, 127–158.

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. (2009). What have we learned about attention from multiple object tracking (and vice versa)? In D. Dedrick & L. Trick (Eds.), *Computation, cognition, and Pylyshyn* (pp. 49–78). Cambridge, MA: MIT Press.

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

Shim, W. M., Alveraz, G. A., & Jiang, Y. V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychonomic Bulletin & Review, 15*, 390–397.

St.Clair, R. L., Huff, M., & Seiffert, A. E. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision, 10*(4, Art. 18), 1–13.

Tsal, Y., & Bareket, T. (2005). Localization judgments under various levels of attention. *Psychonomic Bulletin & Review, 12*, 559–566.

Vul, E., Frank, M. C., Tenenbaum, J. B., & Alvarez, G. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems 22* (pp. 1955–1963). La Jolla, CA: Neural Information Processing Systems Foundation.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15*, 160–168.

Wolfe, J. M., Place, S. S., & Horowitz, T. S. (2007). Multiple object juggling: Changing what is tracked during extended multiple object tracking. *Psychonomic Bulletin & Review, 14*, 344–349.

Zelinsky, G. J., & Todor, A. (2010). The role of "rescue saccades" in tracking objects through occlusions. *Journal of Vision, 10*(14, Art. 29), 1–13.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*, 233–235.