Journal of Experimental Psychology: Human Perception and Performance

Off to a Bad Start: Uncertainty About the Number of Targets at the Onset of Multiple Object Tracking

Zheng Ma and Jonathan I. Flombaum

Online First Publication, January 21, 2013. doi: 10.1037/a0031353

Journal of Experimental Psychology:
Human Perception and Performance

0096-1523/13/$12.00 DOI: 10.1037/a0031353

Off to a Bad Start: Uncertainty About the Number of Targets at the Onset of Multiple Object Tracking

Zheng Ma and Jonathan I. Flombaum
Johns Hopkins University

Visual tracking abilities are limited to only a few objects at a time. When do errors arise? We hypothesized that some errors arise prior to tracking; specifically, during the first moments of a trial because of an inability to correctly perceive the number of targets in a display. To test this hypothesis, we modified a basic multiple object tracking (MOT) task in two ways: (1) we distilled the first moments of MOT into a static working memory task, requiring participants to remember and then identify targets among nontargets in displays without motion; (2) we unconstrained the number of responses a participant could make, asking them to terminate each trial when they felt that they had made an adequate number of responses. In Experiment 1, participants made the wrong number of responses in a considerable number of trials, and they tendered the wrong number of responses more frequently with larger loads. Comparisons across different delay durations demonstrated that these results were not caused by temporal decay. Follow-up experiments produced similar results when participants stated the cardinal number of targets perceived in a static trial (Experiment 2), and when they reported whether or not a test display included the same number of targets as a memory display (Experiment 3). Finally, with a typical tracking duration, participants also produced the wrong number of responses frequently (Experiment 4). Thus, some of the difficulty associated with MOT originates from uncertainty about the number of targets at the start of an episode.

Keywords: multiple object tracking, attention, spatial memory, enumeration

Over the last two decades, human visual tracking abilities have attracted considerable research attention. Much of this interest can be credited to the multiple object tracking (MOT) paradigm introduced by Pylyshyn and Storm (1988). In a typical study, participants see a set of featurally identical objects (e.g., circles). A subset of these objects is identified as targets, and then all the objects—targets and nontargets— move through the display. In order to evaluate whether they successfully tracked, participants are asked to identify the original targets at the end of a trial. Among many virtues, the paradigm makes salient the fact that cognitive abilities are limited. People invariably make mistakes when tracking more than three or four objects, and as a result, MOT supplies a fertile opportunity for exploring the causes of cognitive limits. MOT may evidence general principles with respect to how and why cognition is limited by providing us with a clear operational question: why can people track only so many objects at once?

In the current study, we suggest that a previously unidentified factor is that people sometimes fail to perceive the right number of targets at the very start of a trial; that is, they begin uncertain about the number of targets intended for tracking. Indeed, artificial tracking systems for real-world applications not only contend with uncertainty about the positions of objects, but also with uncertainty about the number of objects present in the first place (Smith, Gatica-Perez, & Odobez, 2005). Thus, the general challenges of accurately processing an image, segmenting a display, and selecting the objects in it may pose a specific challenge to tracking mechanisms.

Investigating these challenges is critical for developing an accurate understanding of human tracking mechanisms. For example, one currently influential class of models appeals to "flexible resources" as the limiting factor in tracking (e.g., Alvarez & Franconeri, 2007; Franconeri, Jonathan, & Scimeca, 2010; Horowitz & Cohen, 2010). An operating assumptions in these models —an assumption stated explicitly in formal articulations (Ma & Huang, 2009; Vul, Frank, Alvarez, & Tenenbaum, 2009)—is that participants have a representation of each target in a trial, even with sets as large as seven or eight; they just may not have a precise representation of each item. The experiments reported here explore the possibility that this assumption is not justified.

The Current Study

The current set of experiments finds its most proximate motivation in a recent study that asked participants to enumerate by clicking (Haladjian & Pylyshyn, 2011). Rather than ask participants to key in the number of objects perceived in a display, the researchers had participants click (with a mouse) in the position of each object they remembered. The logic of the experiment was that participants should generate a click for each item they represented and, therefore, that the number of clicks produced should reflect the cardinal number of items enumerated. The data revealed clicks

Zheng Ma and Jonathan I. Flombaum, Department of Psychological and Brain Sciences, Johns Hopkins University.

Correspondence concerning this article should be addressed to Jonathan I. Flombaum, Johns Hopkins University, Department of Psychological and Brain Sciences, Ames Hall, 3400 N. Charles Street, Baltimore, MD 21218. E-mail: flombaum@jhu.edu

nearly equal to items displayed for sets of about six items, but noisy distributions of responses for larger sets. A crucial point is that the response-eliciting display was an empty screen, and all the items presented initially were targets (i.e., there were no nontargets).

In Experiment 1, we sought to ask a similar set of questions for the equivalent of the kind of display that usually commences an MOT trial. Given a subset of targets to select from a larger set, within what range will observers accurately enumerate[1] the targets? To answer this question, we modified a standard MOT task in two ways. First, we distilled the task into a static working memory test, comprising just the initial moments of MOT. Targets turned a unique color, they then became identical in color to the nontargets, and after a delay period, participants clicked on each item that they recalled as a target. Second, we unconstrained the number of responses a participant could supply. All prior work on multiple object tracking has constrained the number of responses participants can make,[2] limiting the ability to determine whether participants represent the wrong number of targets.

The question of interest was whether observers would click on the right number of targets when responding was unconstrained. We found that they did not for loads as small as six in Experiment 1. Follow-up experiments also evidenced observer uncertainty about the number of targets in these displays by requiring an explicit numerical report (Experiment 2) and via a discrimination task (Experiment 3). And finally, Experiment 4 included a tracking period, evidencing observer uncertainty about the number of targets in a typical MOT trial.

Experiment 1: Unconstrained Spatial Working Memory

The most common method for assessing MOT performance is a "mark-all" procedure, in which, at the end of a trial, a participant needs to click with the mouse on all the items she believes are targets. All previous research using the mark-all method has constrained the number of responses a participant can produce to equal the number of targets in a trial (e.g., Scholl, Pylyshyn, & Feldman, 2001). As a result, it could not be revealed had participants overestimated or underestimated the number of items they were supposed to track in these experiments.

To determine whether participants represent the right number of targets at the start of a trial, we distilled it into a static working memory task and we unconstrained the number of clicks a participant could make (see Figure 1). We did not assume that representing the right number of targets would automatically lead to perfect identification of targets at test. But we did expect that if participants perceived the right number of targets at the start of a trial, that they should always produce as many clicks as there were targets. In contrast, any failure to accurately represent the number of targets at the start of a trial should lead participants to produce the wrong number of responses at the end of a trial.

In this first experiment, we also manipulated the duration between target presentation and test, including three different delay intervals during which items remained static on the screen. We did this to meet two goals. First, we wanted to explore the contents of spatial knowledge just before tracking begins in typical MOT. For this reason, we used a delay duration of .5 s, similar to the elapsed time after targets are identified but before motion in typical MOT.

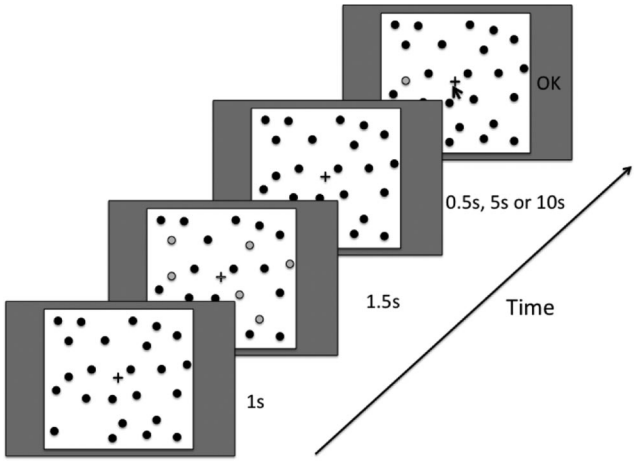

Figure 1. Procedure of Experiment 1. Each trial began with 24 blue discs (shown in black) arranged haphazardly in the display. Next, a subset of 4 to 9 turned yellow (shown in gray), identifying them as targets to be remembered. Finally, all the items turned blue during a retention interval of 0.5, 5, or 10 seconds. At test, participants clicked on all the items they remembered as targets. Participants chose when to terminate a trial by clicking the *OK* button to the right of the display.

Second, we wanted to explore the possibility that longer durations lead to worse numerical performance. This was out of concern that participants could underreport the number of targets because of decay, as opposed to genuine numerical uncertainty. Accordingly, we included durations of 5 s and 10 s, choosing these specific values because they are common as tracking durations in MOT. We predicted that memory duration would not impact the rate of enumeration errors in accord with the hypothesis that enumeration errors reflect uncertainty about the information present in an initial image identifying targets.

Method

Participants. Nine Johns Hopkins University undergraduates participated for course credit. All had normal or corrected-to-normal visual acuity.

Apparatus. Stimuli were presented on a Macintosh iMAC computer with a refresh rate of 60Hz. The viewing distance was approximately 60 cm so that the display subtended 39.43° × 24.76° of visual angle.

Stimuli and procedure. Stimuli were generated with MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). All stimuli were presented in a black square subtending

[1] A point of clarification may be worthwhile (before further explaining the logic of the experiments). We mean *enumerate*, here, in the sense of specifying each individual in a set, as in the sentence, "Let me *enumerate* my reasons for studying MOT." We reserve for the General Discussion a consideration of whether the results relate to work on number perception and estimation.

[2] In our review of the literature we found that any paper that explicitly discussed response collection methods included constraints on the number of clicks a participant could make. Several papers, however, were ambiguous, and so it is possible that some papers did not constrain clicks but failed to mention that fact in their methods.

25.38° × 19.98°. Each trial started with 24 blue disks (diameter 0.94°) along with a white fixation cross (0.47° × 0.47°) in the center (which remained present throughout a trial). The positions of disks were randomly selected, with the restriction that the centers of any two disks be separated by at least 1.65°, and with the restriction that no disk could occlude fixation. After one second, a subset of between 4 and 9 disks turned yellow for 1.5 s, indicating that these were the targets. Finally, all disks turned blue again, remaining so for 0.5 s, 5 s, or 10 s, at which time a participant could begin to make responses.

To indicate to participants that they should identify the targets, a mouse cursor appeared on the fixation point. When a participant clicked on a disk, it turned yellow in order to prevent selecting a given disk more than once and to remove any memory demands with respect to the disks already selected. A selection could be canceled by reclicking a disk. We emphasized to participants that they should guess if uncertain about the identity of particular targets.

To the right of the display there was an *OK* button, which participants clicked once they had made all their responses in a trial. There were 10 trials for each target load (4–9) and duration, leading to a total of 180 trials presented in a random order.

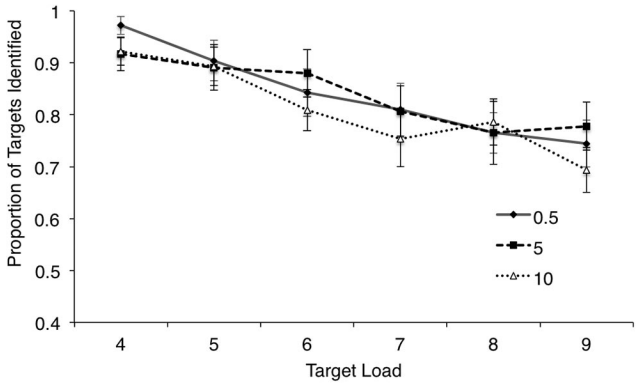

Figure 2. Memory performance in Experiment 1. Error bars show ±1 *S.E.*

Results

Memory accuracy. Before analyzing numerical misperception, we report memory accuracy. We measured accuracy as the proportion of targets correctly identified (see Figure 2). Because these proportions are relative to different target loads (i.e., 10% correct does not mean the same thing for a load of 3 vs. 6), we do not report main effects of target load on memory accuracy (this is true in all subsequent experiments as well). A 3 (duration) × 6 (target load) repeated measures ANOVA showed that duration had a significant main effect on performance, $F(2, 16) = 4.1$, $p < .05$. (Whenever sphericity was violated, we report significance with a Greenhouse-Geisser correction). Accuracy decreased as duration increased, with a significant linear trend analysis, $F(1, 8) = 6.428$, $p < .05$.

Enumeration errors. In order to explore the possibility that participants sometimes represented the wrong number of targets, we analyzed the number of responses made—whether accurate or inaccurate—in each trial. Figure 3 displays the probability of making a given number of responses given a specific target load and memory duration. The darkness of a cell is shaded in proportion to the count for a given number of responses. For example, the cell corresponding to $x = 4$, $y = 4$ with a duration of .5 s is almost perfectly black, indicating that participants selected four items on nearly all four target trials (in point of fact, all but 3 trials). In contrast, the cells corresponding to a target load of eight reveal a widely distributed range of responses. As is clear from the figure, participants did not always make the right number of responses, and the tendency to produce the wrong number of responses increased with increasing target load. As is also clear in Figure 3, when participants enumerated incorrectly, they were far more likely to underestimate than to overestimate (Izard & Dehaene, 2008).

Beyond whether or not participants enumerated correctly, we wanted to obtain a sense of how frequently they did. Accordingly, we computed the proportion of trials in which participants produced the wrong number of responses (see Figure 4). A 3 (time interval) × 6 (set size) repeated measures ANOVA showed a significant main effect of target load, $F(5, 40) = 29.072$, $p < .001$. Error rates increased as target load increased, with a significant linear trend analysis, $F(1, 8) = 62.351$, $p < .001$. A crucial point is that there was no significant main effect of duration, $F(2, 16) = 0.821$, $p > .05$, and there was no significant interaction between duration and target load, $F(10, 80) = 2.098$, $p > .05$.

We also performed a series of one-sample t tests to compare enumeration error rates with zero for each target load and condition. The results suggested that participants started to make enumeration errors at a significant rate at a target load of five across all memory durations. With a Bonferroni correction, rates were significant at loads of eight and nine.

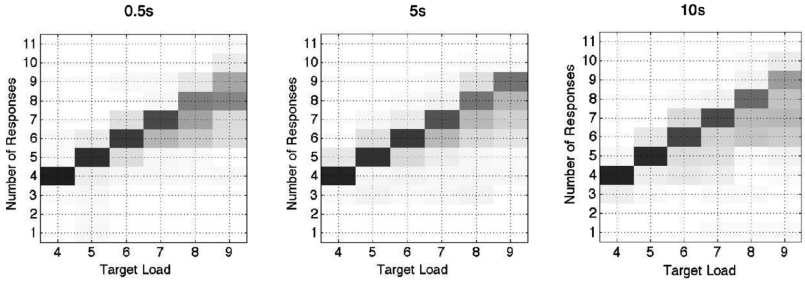

Figure 3. Distributions of the number of responses made given a target load in Experiment 1. Distributions for different retention durations are shown in separate panels.

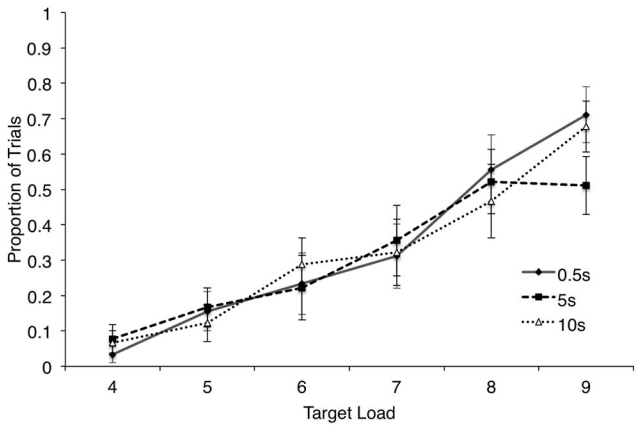

Figure 4. Proportion of trials with an incorrect number of responses as a function of target load and retention duration in Experiment 1. Error bars show ±1 *S.E.*

Discussion

This experiment provided clear evidence of enumeration errors in a simple spatial working memory task, one that was designed to constitute a necessary first step in performing any MOT task. We reserve a detailed discussion of MOT and related issues until the general discussion. But we emphasize here what we view as the most important point: at larger target loads, participants enumerated incorrectly at considerably high rates. Indeed, even with a target load of just five, participants reported the wrong number of targets in more than 10% of trials.

With respect to the possibility of memory decay, we failed to find a significant impact of memory duration on erroneous enumeration rates. If decay accounts for uncertainty about the number of targets, then participants should have been more uncertain at longer delays. Note that the longest delay employed was 20 times larger than the shortest delay.

Still, other kinds of response-related decay and bias may have influenced the results. Perhaps participants in Experiment 1 did know the number of targets in the display, but memory decayed while supplying serial responses, or perhaps participants were averse to guessing when uncertain about the identity of a target. Experiments 2 and 3 were designed to exclude these possibilities.

Experiment 2: Errors in an Explicit Enumeration Task

This experiment sought to extend the findings of the prior experiment in two ways. First, we wanted to determine whether explicit knowledge about the cardinal number of targets at the start of MOT resembles the kind of knowledge implied by unconstrained clicking. Based on a large literature on numerical perception (Brannon, 2006) we expected that when asked to report the number of targets, participants would reveal increasing uncertainty with increasing loads. Such uncertainty would oppose the hypothesis that participants do know the number of targets at the start of a trial, but possess an aversion to guessing about their identities. Similarly, because a cardinal value is reported via a single response, misreporting in this experiment would oppose the hypothesis that knowledge of the number of targets decays as participants execute serial responses.

Second, we sought to investigate explicit numerical estimation in a situation more similar to MOT than is typical in the estimation literature—one with motion and a tracking requirement. Despite a vast literature on visual number estimation, to our knowledge, typical experiments do not involve moving stimuli, or a need to individually select, attend, or track objects. We were concerned that explicit numerical errors may not arise in tasks with motion and a need for tracking, perhaps because tracking demands engage different mechanisms than typical estimation tasks. Thus, we included brief motion in this experiment, and a tracking requirement. Similarly, so that participants would need to activate any tracking-specific mechanisms in all trials, we intermixed click-report and explicit numerical-report trials.

Method

Participants. A new group of 14 Johns Hopkins University undergraduates participated for course credit. All had normal or corrected-to-normal visual acuity. The data from one participant were incomplete and unanalyzed due to operational errors.

Stimuli and procedure. The stimuli and procedure were identical to Experiment 1, with the following exceptions. We added trials that required an explicit report of the number of targets (see Figure 5). Instead of an *OK* button, a sentence appeared: "please type in the number of targets."

Additionally, in place of static disks, we introduced motion in this experiment. After targets were revealed, all of the disks moved randomly through the display for 0.5 s. They all moved linearly with a constant speed of 4.23°/s, a speed that usually affords

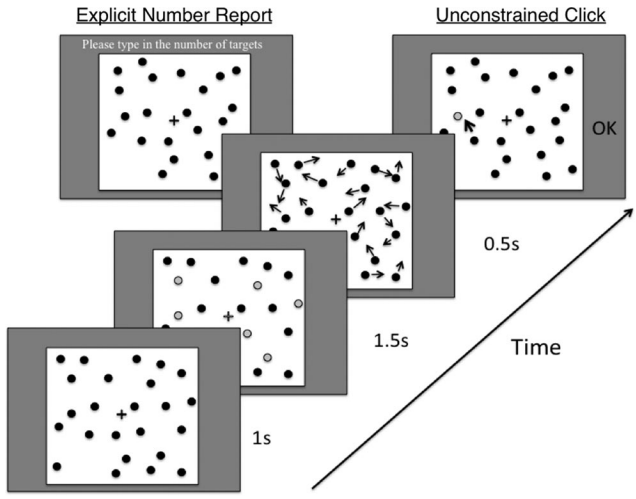

Figure 5. Procedure of Experiment 2. Each trial began with 24 blue discs (shown in black) arranged haphazardly in the display. Next, a subset of 4 to 9 discs turned yellow (shown in gray), identifying them as targets to be remembered. Finally, all the items turned blue again and moved in random directions for a duration of 0.5 seconds. At test, in the "Unconstrained Click" trials, participants clicked on all the items they remembered as targets and terminated a trial by clicking the *OK* button to the right of the display. In the "Explicit Number Report" trials, participants were prompted to type in the number of targets included in that trial.

effective tracking of about four items (Alvarez & Franconeri, 2007). On average, each disk moved 2° in a trial. Disks repelled one another whenever they came within 1.69°, center to center, and they also changed direction to repel from the boundary of the display or the fixation cross. On any frame, each disk had an independent 2% chance of changing direction. New direction vectors could take on any value within 359° relative to the prior vector.

Target loads of four to nine (out of a constant 24) were employed. We included 10 trials for each target load and reporting method, leading to a total of 120 trials. Trials with different target loads and reporting methods were intermixed and randomly distributed.

Results

Tracking accuracy. The average proportion of targets correctly identified across all set sizes in the click report trials was 74.4% (see Figure 6).

Enumeration errors. Distributions reflecting numerical perceptions are depicted in Figure 7. To compare these distributions statistically, we used a Kolmogorov–Smirnov test at each target load. The distributions were not significantly different for loads four, five, six, and seven ($p > .05$ for each comparison). Without a Bonferroni correction, the distributions were significantly different at load eight ($p = .009$). With a correction, they were significantly different only at load nine ($p = .0061$). Thus, the two response methods produced comparable distributions.

In the literature on number perception, one common signature is an increasing standard deviation in numerical reports as a function of the true numerosity, but with little or no change in the coefficient of variation. Figure 8 plots these descriptive statistics for each method. With both response methods, standard deviations increased in proportion to target load, though coefficients of variation remained relatively constant. This is characteristic of a system that becomes more error prone as a function of the magnitude it encounters (e.g., Brannon, 2006; Platt & Johnson, 1971; Whalen, Gallistel, & Gellman, 1999).

To analyze the incidence of incorrect enumeration statistically, we conducted a 2 (response method) × 6 (target load) repeated measures ANOVA on the proportion of trials with incorrect numerical responses (see Figure 9). There was a significant main effect of response method, $F(1, 12) = 8.86$, $p = .012$. Participants made enumeration errors in more trials with the unconstrained clicking method than with the direct report method. There was also a significant main effect of target load, $F(5, 60) = 36.782$, $p < .001$. Rates of incorrect enumeration increased as set size increased, with a significant linear trend analysis, $F(1, 12) = 118.02$, $p < .001$. There was no significant interaction between response method and target load, $F(5, 60) = 2.375$, $p > .05$.

We performed a series of one-sample t tests to compare erroneous enumeration rates with zero for each response method and target load (12 total tests). Enumeration error rates were significantly greater than zero at a target load of five for the click trials (but not the number report trials). With a Bonferroni correction, erroneous enumeration rates were significantly greater than zero for both methods at a target load of six (and above).

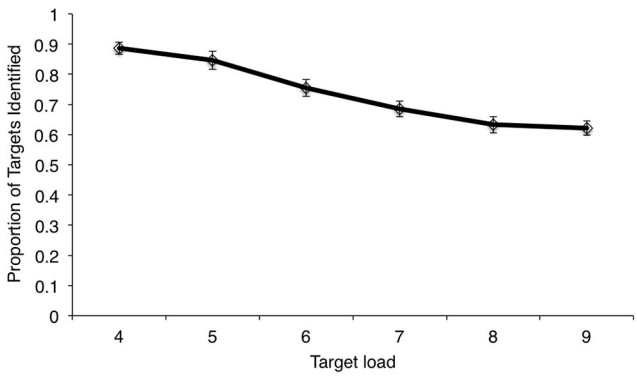

Figure 6. Tracking performance in Experiment 2. Performance was measured as proportion of targets identified. Error bars show ±1 S.E.

Discussion

When asked to explicitly report the number of targets at the start of an MOT trial, participants proved uncertain, reporting the wrong number in a significant number of trials for loads of six and above. With this method, it is unlikely that inaccuracy reflected memory decay or an aversion to guessing.

We did find a difference between the explicit report method used here and the unconstrained clicking method with respect to the rate of trials inaccurately enumerated. But a critical point is that this difference was small relative to the large rate of incorrect enumeration with both methods, and given the fact that the distributions of responses were comparable. Even with explicit reports, participants enumerated incorrectly in over 10% of load six trials and in over 20% of load seven trials.

Experiment 3: Uncertainty About Target Number in a Discrimination Task

In this experiment, we contrived yet another method for exploring the pretracking moments of MOT, a method that is also immune to concerns about memory decay and a potential aversion to guessing. The approach was based on decades of discrimination experiments in the study of human perception. Participants observed a display with a group of identical disks. A subset changed color—as they would in MOT—identifying them as targets to remember. All the disks then became the same color again and remained so for a static retention interval of one second. Finally, three possible events took place at test: (1) all the former targets, and only the former targets, turned a unique color; (2) all the former targets and one additional nontarget turned a unique color; or (3) all the former targets save for a randomly chosen one turned a unique color. Participants' task was to report whether the second set of colored items was the same as the original set or different. They were told that it would be the same in one half of all trials, and that in the remaining half of trials there would be either one added or one missing from the set, with no other deviations possible. We predicted that performance would drop with increasing target loads, reflecting uncertainty about the exact set of items making up the initial target set.

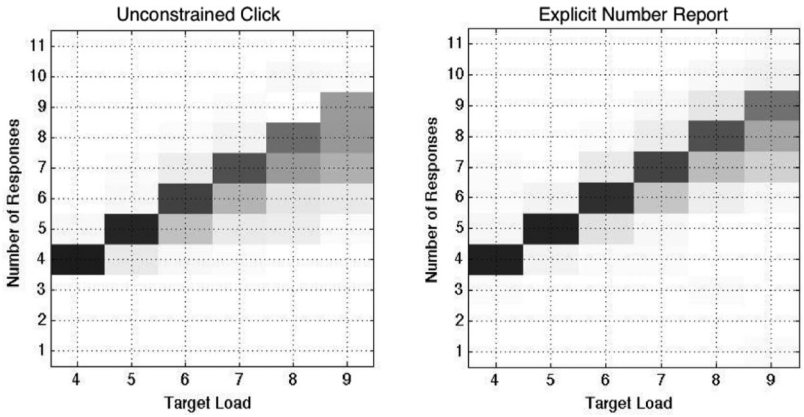

Figure 7. Distributions of the number of responses made given a target load in Experiment 2. Distributions for the two different response methods are shown in separate panels.

Method

Participants. A new group of 11 Johns Hopkins University undergraduates participated for course credit. All had normal or corrected-to-normal visual acuity.

Stimuli and procedure. This experiment was identical to Experiment 1, except as follows. At the start of a trial, a subset of between four and nine disks turned yellow for 1.5 s, indicating that these were the targets. Next, all disks turned blue again, remaining so for one second. Participants were instructed to remember which disks were the targets over this interval. We used a one second rather than a 0.5 s duration to prevent the perception of apparent motion. Finally, another subset of disks turned yellow and the participants were asked to judge whether the second yellow subgroup was exactly the same as the first subgroup. In 50% of trials, the two groups were exactly the same. In 25% of trials all the original targets turned yellow again, but one additional disk was randomly added. In the remaining 25% of trials, one of the original targets was randomly left out of the test group. Participants were made aware of exactly how the "different" trials could differ from the "same" ones, as well as the relative distributions across all the trials. They entered a *same* or *different* response via keypress.

There were 40 trials for each of six target loads between four and nine, resulting in a total of 240 trials. Trials with different target loads and types were intermixed and randomly distributed.

Results

We measured the ability to discriminate sample and test images with d' (see Figure 10). A one-way repeated measure ANOVA showed that there was a significant main effect of target load, $F(5, 50) = 18.47$, $p < .001$. Discrimination abilities decreased as target load increased, with a significant linear trend analysis, $F(1, 10) = 58.40$, $p < .001$.

Discussion

The general purpose of this study was to determine whether participants accurately perceive the set of targets intended for tracking in MOT. In the current experiment, we found that they are prone to misperceptions, as evidenced by a declining ability to discriminate between a complete set of initially presented targets, and a set including one more or one fewer. Declines associated

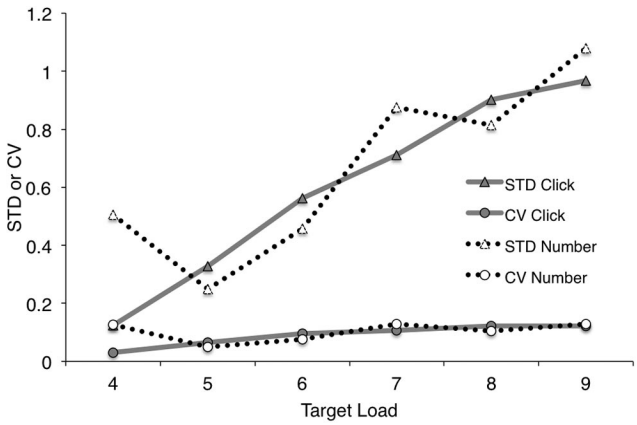

Figure 8. Average standard deviations (STD) and coefficients of variation (CV) as a function of target load and report method in Experiment 2.

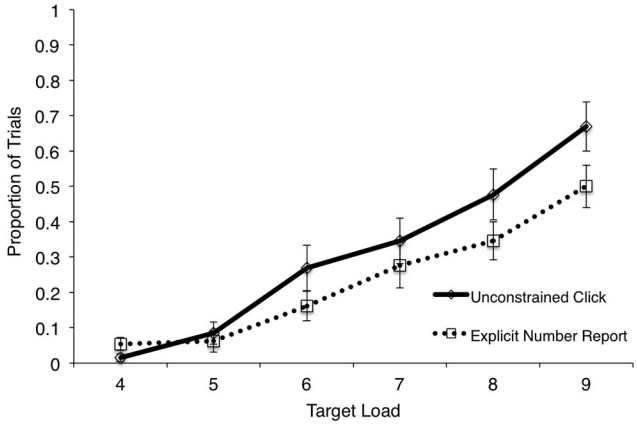

Figure 9. Proportion of trials with an incorrect number of responses as a function of target load and response method in Experiment 2. Error bars show ±1 *S.E.*

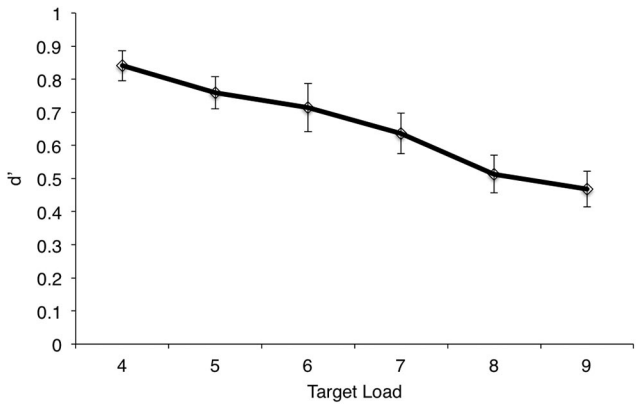

Figure 10. *d'* as a function of target load in Experiment 3. Error bars show ±1 *S.E.*

with increasing number were relatively large. With a load of four targets, mean *d'* was nearly .85. But sensitivity dropped by 15% with a load of just six. It is important to note that this experiment did not require serial responses, and it could not be influenced by an aversion to guessing about target identifies. The difference between *same* and *different* trials only involved the presence or absence of a single item. A best guess should have relied on a participant's full body of knowledge.

Experiment 4: Unconstrained Multiple Object Tracking

This experiment sought to extend our general findings to a more typical MOT task. Participants were asked to select between four and nine targets from a group of 24 total items, and then they tracked those targets as they moved haphazardly through the display for a duration of 5 s. Following the methods of Experiment 1, we unconstrained the number of responses a participant could make, requiring them to terminate trials on their own. To our knowledge, this was the first study with a mark-all method to allow over- and underreporting. We predicted that participants would make enumeration errors, evidencing uncertainty about the number of targets in an MOT trial.

Method

Participants. A new group of 10 Johns Hopkins University undergraduates participated for course credit. All had normal or corrected-to-normal visual acuity.

Stimuli and procedure. The stimuli and procedure were identical to Experiment 2 except as follows. (1) Only unconstrained trials were included. (2) We used a 5 s tracking duration, which has been widely used in previous MOT studies.

There were 10 trials for each target load, producing a total of 60 trials. Trials with different target loads were randomly distributed.

Results

Before reporting analyses, it is worth mentioning that in less than 0.2% of all trials participants made no response at all, clicking *OK* immediately at test. We presume that these trials reflect motor errors. Since all results are the same with and without these trials included, we report results including all of the data collected.

Tracking accuracy. The average percentage of targets correctly identified was 54.95% (see Figure 11). Recall that participants were asked to track loads as large as nine.

Enumeration errors. Distributions reflecting the number of responses made given a particular target load and tracking duration are depicted in Figure 12. A one-way repeated measures ANOVA showed that there was a significant main effect of target load, $F(5, 45) = 30.695$, $p < .001$, on the frequency of trials with the wrong number of responses (see Figure 13). Erroneous enumeration rates increased as set size increased, with a significant linear trend analysis, $F(1, 9) = 85.696$, $p < .001$.

We performed a series of one-sample *t* tests to compare erroneous enumeration rates with zero for each target load (six total tests). The results suggested that enumeration error rates were significantly higher than zero at a target load of five ($p = .017$). With a Bonferroni correction for multiple comparisons, erroneous enumeration rates were significantly different from zero starting at a target load of six. We emphasize that at a load of five, erroneous enumeration took place in about 10% of trials. Perhaps more surprisingly, with durations of just five seconds and a load of just six targets, participants reported the wrong number of items in 40% of trials.

Discussion

This experiment provided direct evidence that enumeration errors can emerge in standard MOT displays and conditions. In fact, enumeration error rates were fairly large in this experiment. Though it is hard to draw conclusions by comparing across experiments, it even appears that at some loads, enumeration error rates were larger in this experiment than in Experiment 1, which did not include tracking. This would be consistent with a recent report suggesting that targets can be lost or given up on entirely during tracking (as opposed to only swapped with nontargets; Drew, Horowitz, & Vogel, 2012). Juxtaposed with Experiments 1–3, the results reported here suggest that some targets may never become acquired successfully, and that additional targets may become lost subsequently during tracking. This latter issue is not the main focus of the current report. But overall, numerical uncertainty

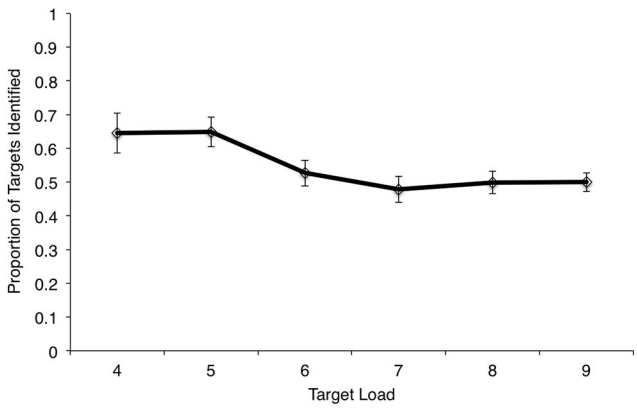

Figure 11. Tracking performance in Experiment 4. Error bars show ±1 *S.E.*

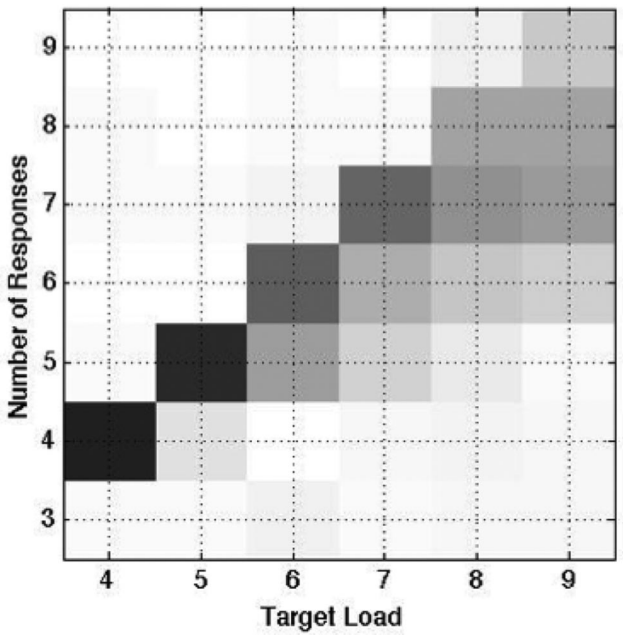

Figure 12. Distributions of the number of responses made given a target load in Experiment 4.

arising at the start of and during an MOT trial seems to be a fact of performance that should be accounted for by current models and theories.

General Discussion

We sought to answer a straightforward but previously unaddressed question: at the start of an MOT trial, do people accurately know the number of targets that are present? The experiments demonstrated that participants do not, at least for displays with more than about four targets. In Experiment 1, we distilled MOT into a task without any tracking demands—a task that amounted to just the presentation and test portions of the typical paradigm. We also unconstrained the number of responses a participant could make at the end of a trial. We reasoned that the number of responses made could be taken as an indication of the number of individuals represented. We found that participants selected the wrong number of targets at significant rates when there were six or more. Additionally, longer memory durations did not impact enumeration in this experiment, excluding memory decay as a cause of our results. Experiment 2 extended the main finding, revealing enumeration errors when participants were asked to explicitly report the number of targets seen. Since these reports involved a single keypress and no need to select individuals, they provided further evidence against decay and an aversion to guessing leading to enumeration errors. Experiment 3 similarly evidenced uncertainty about the number of targets present in a display via a discrimination task, a task that should also be resistant to memory decay effects and a tendency to avoid guessing. Finally, Experiment 4 included a 5-s tracking duration, but unconstrained responses. Participants supplied the wrong number of responses with loads as small as five or six, demonstrating that they may be uncertain about the number of targets in a display when attempting to track multiple visual targets.

Overall, Experiments 1–3 constitute a kind of baseline data not previously present in the literature on multiple object tracking. Given just the target acquisition phase of a trial, how many targets can participants be expected to acquire? We discovered that the answer is not necessarily as many as they were initially asked to select. A complete theory of how MOT is carried out should predict these effects, and it should incorporate errors that arise before tracking begins into predictions about performance. In the remainder of this discussion, we explore potential connections between these results and research on subitizing and number estimation. We then discuss the implications for current models of how humans track multiple moving objects.

Enumeration Errors and Subitizing

The main result of the reported experiments is that given more than five or so, an observer cannot precisely report the number of targets in the kind of image that usually begins an MOT trial. But reports are systematically related to the true numerosity, and error range is proportional to the magnitude. One clear pattern in these results is that the frequency and range of incorrect enumeration is greater with seven or eight targets than with five or six. In the literature on number perception, there exists the hypothesis that a dedicated, parallel, and rapid "subitizing" system affords precise representations of small numbers of objects, leaving a noisy system to handle large numbers (Feigenson, Dehaene, & Spelke, 2004). To what extent may the reported results reflect the influence of the subitizing system?

For several reasons we believe that it may be difficult to relate these results directly to subitizing; that is, to take them as evidence of or as caused by the subitizing system—though they may be consistent with the hypothesis of subitizing. Specifically, the current experiments, by design, only explored numerical knowledge for four or more targets. Subitizing is typically thought to support representations of three or fewer (Trick & Pylyshyn, 1993). Even had we tested loads of three and fewer, it would be difficult to

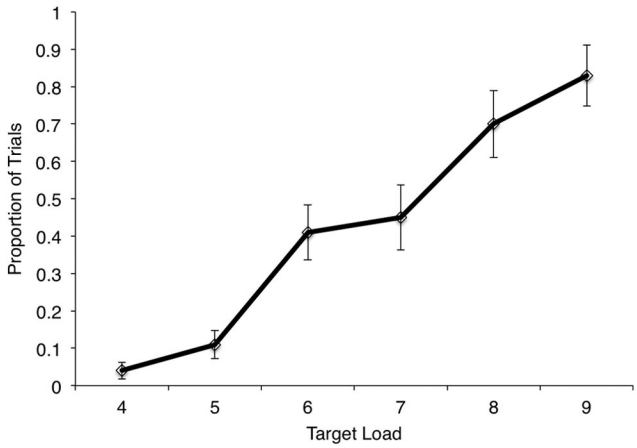

Figure 13. Proportion of trials with an incorrect number of responses as a function of target load in Experiment 4. Error bars show ±1 *S.E.*

conclude with certainty whether patterns of enumeration reflect subitizing. Usually, these inferences are motivated by an "elbow" in enumeration accuracy or speed at the transition from three to four. But since it is known that enumeration inaccuracy scales with magnitude, it is hard to know when an elbow is severe enough to warrant a subitizing interpretation. In fact, the existence of subitizing has been challenged, in the past, on the grounds that a genuine elbow does not exist in typical experiments (Whalen et al., 1999). Similarly, subitizing theories are complicated by recent work suggesting that the subitizing limit may be closer to five or six (Haladjian & Pylyshyn, 2011)—a theory that may be consistent with our data, but would require more focused investigation. And in the current Experiment 3, which used a discrimination procedure, it is worth noting that there was no apparent inflection in d' as a function of load; it appeared to decline linearly throughout the tested range.

Because of these complexities, we remain agnostic as to whether the enumeration errors observed in the current experiments support subitizing theories and whether subitizing supports tracking of three or fewer objects. A recent study observed a one-to-one trade off in MOT abilities and concurrent subitizing with smaller numbers of targets (Chesney & Haladjian, 2011), results that seem compatible with the current experiments. The subitizing system may well support tracking of three or fewer. But the primary concern in the current study was to understand human enumeration capabilities at higher loads in order to evaluate the implications for MOT mechanisms that handle those loads.

Multiple Object Tracking and the Approximate Number System

Imagine a display haphazardly strewn with yellow dots. The display appears briefly—too briefly for you to count each individual dot and slowly increment a total—and then you are asked to report the number of dots that were present. How well would you do? Considerable research has focused on exactly this question, exploring, generally, how humans and other organisms estimate large numbers of stimuli in various classes and modalities (e.g., Cantlon & Brannon, 2007; Jordan & Brannon, 2006; for reviews, Brannon, 2006; Feigenson et al., 2004), as well as exploring displays that bear striking similarity to the starting point of a typical MOT trial (e.g., Barth, Kanwisher, & Spelke, 2003; Barth et al., 2006; Halberda, Sires, & Feigenson, 2006). The primary conclusion is that human adults, infants, and other organisms possess an ability to estimate the numerosity of large numbers of stimuli (>3), but noisily. Given some number of stimuli to enumerate, variance tends to increase in proportion to the correct magnitude such that the coefficient of variation remains constant (Platt & Johnson, 1971), leading researchers to describe the approximate number sense as employing an analog or magnitude-like representational format (Brannon, 2006).

The displays often used to study the approximate number sense bear an uncanny resemblance to the displays that commence most tracking trials (and any of the intervening moments comprised by the duration of a trial). Beyond these cosmetic similarities, there are good reasons for thinking that tracking mechanisms and approximate number mechanisms are related. For example, cues to object segmentation, including uniform connectedness, have been to known to constrain tracking abilities (Scholl et al., 2001), and they have also been shown to constrain approximate number estimation in visual displays (Franconeri, Bemis, & Alvarez, 2009). And both tracking and number approximation are known to recruit regions of the intraparietal sulcus (IPS). The IPS has been the primary area implicated in studies of approximate number representation, both via BOLD responses and in recordings from single units tuned to number (Brannon, 2006; Nieder & Dehaene, 2009; Nieder & Miller, 2004). Similarly, IPS responses have shown sensitivity to the number of items tracked in an MOT trial (Jovicich et al., 2001). Additional regions, including prefrontal areas and other parts of parietal cortex, have also been implicated in studies of both capacities (Culham et al., 1998; Nieder & Miller, 2004; Nieder, Freedman, & Miller, 2002). Thus, tracking and approximate enumeration may share a similar network of neural circuitry. Our results add to the general impression that number perception and object perception rely on at least some shared mechanisms (Burr & Ross, 2008; Durgin, 1995; Ross & Burr, 2010). Future research will hopefully exploit these connections in a variety of ways. From our perspective, one exciting thread may involve the application of early vision models as a "front end" to both enumeration and tracking, an approach that has already met some success in the case of enumeration (Dakin, Tibber, Greewood, Kingdom, & Morgan, 2011).

Implications for Understanding Human Tracking Limits

The fact that observers are uncertain about the properties of the visual world has become a key insight in recent research concerning both visual perception (Purves, Wojtach, & Lotto, 2011) and visual cognition (Bays & Husain, 2008; van den Berg, Shin, Chou, George, & Ma, 2012). But in MOT (and in related work on visual working memory), the application of this insight has typically been limited to the properties of individual objects—such as their positions, colors, shapes, orientations, and so forth—as opposed to the properties of scenes. However, considering observer uncertainty from this perspective should play an important role in the development of more complete models of the mechanisms that support tracking and related processes. From first principles, there is no reason to assume that the visual system would have imperfect knowledge about the positions of items, for instance, but perfect knowledge about their presence. This is salient in related work on artificial tracking systems. We quote from a recent paper describing such a system because the point is made especially clearly:

> Tracking a fixed number of independent, hand-initialized objects is a well studied problem. However, the automatic detection and tracking of a variable number of interacting objects is still difficult, implying three challenging tasks: (1) reliably estimating the number of objects in the scene, (2) keeping the algorithm computationally tractable when multiple objects appear simultaneously, and (3) modeling interactions between varying numbers of objects. (Smith et al., 2005)

Current models of human MOT generally assume something like a "hand-initialized" set of objects as the inputs to tracking,

focusing their efforts, instead, on managing interactions between varying numbers of objects. But there is no reason to assume that the inputs to tracking always comprise a perfectly segmented image. It follows that expectations about poorly segmented displays should constrain theories of MOT.

This applies equally to theories that appeal to "fixed resources" and ones that appeal to "flexible resources." According to fixed resource theories, tracking is limited by a limited set of discrete representations (e.g., Cavanagh & Alvarez, 2005; Drew & Vogel, 2008; Pylyshyn, 2001). In at least one respect, our results fit conveniently with such theories. After all, if participants have only a limited set of representations, then of course they should be uncertain about the presence of targets above a certain total number. As would be predicted, our participants underestimated frequently, and often by a large margin.

But our data also place constraints on how fixed resource theories can account for performance when observers are tasked with tracking more than three objects. Enumeration in the reported experiments was systematically related to the true number of targets, and there was variability from trial to trial, even at times with smaller target loads. A straw-man version of a fixed resource theory might predict purely random enumeration behavior at larger target loads, and only underestimation (by a wide margin). We doubt that this would be the version endorsed by many, but nonetheless, a more viable model would need to supply a specific account for both how an observer comes to perceive, select, and then track more than three targets, and also for their patterns of responses in the experiments reported here. For example, fixed resource theories may appeal to some form of grouping (e.g., Yantis, 1992) to explain, in general, why observers appear able to track more than three objects better than would be predicted if they tracked only three and guessed on the rest. Our data now place a second set of constraints on the kinds of predictions that such an account would need to make; it would also need to explain the approximate numerical knowledge that observers seem to possess about larger target sets and the precise kind of knowledge that they seem to lack. A fixed resource model could potentially do so by appealing to grouping mechanisms with approximate number mechanisms riding on top, a proposal that seems plausible. But a specific articulation of how these mechanisms work and interact would be necessary to evaluate such a model quantitatively and in comparison to others.

In contrast to fixed resource theories, flexible resource theories hypothesize that representations of tracked objects consume a finite but continuous pool of resources. When one tracks more objects, one allocates fewer resources to each, and as a consequence, each object is represented less precisely (Alvarez & Franconeri, 2007; Franconeri et al., 2010; Horowitz & Cohen, 2010; Ma & Huang, 2009; Vul et al., 2009).

As currently articulated, flexible resource theories assume that observers represent each and every item in a display. This can be seen most clearly in two formal implementations (Ma & Huang, 2009; Vul et al., 2009). To their credit, these models are transparent, making the relevant assumptions obvious. It may seem inconceivable that observers can represent hundreds or even tens of items simultaneously—and to be fair, the authors of the relevant models may not have intended for such commitments. So we emphasize that the issue is not representing tens of objects, but six to nine, where model parameters have been fit to human results under the assumption that all targets are tracked. This assumption needs to be relaxed in order to obtain more realistic parameter estimates, and the implications of the models could be rather different as a result.

In both model versions (Ma & Huang, 2009; Vul et al., 2009), tracking abilities are described as Bayesian inferences that relate one set of observations to new observations. Put simply, an observer stores the current position of each target at some moment in time, and at the next moment the observer infers which of the currently observed objects is the best match for each of the remembered targets. A crucial point is that a representation of each target originates during the first moments in a trial. Here, one can think of an observer as receiving a sample from each target when they change color. But uncertainty about the positions of the samples leads to inferred probability distributions for the positions of the targets. Thus, observers emerge with a representation of each target, but uncertain about their positions.

This perspective assumes that an observer already knows how many targets there are in a display, and that she knows which noisy sample came from which object. It seems more likely, however, that whatever noisy samples observers receive, they use them as the basis for determining both the number of objects present and their positions. Thus, they should be jointly uncertain about at least these features. As we have shown, participants often start a trial with an ignored handicap—uncertainty about the presence of targets.

To be sure, this fact could be incorporated into models that may still depend on flexible resources at some point in processing. It is critical, though, that the extent to which limited resources take the responsibility for task errors could change considerably. According to current models, responsibility is assigned via a cascade of causes and effects: limited resources reduce the precision of spatial knowledge, leading to errors via confusions between targets and nontargets (Bae & Flombaum, 2012; Franconeri et al., 2010; Vul et al., 2009). The inference of limited resources emerges from the observation of worse performance for many targets than would be expected given performance with few (e.g., Alvarez & Franconeri, 2007), combined with better model fits when spatial precision is assumed to decline with increasing loads (as opposed to remaining constant; Ma & Huang, 2009; Vul et al., 2009). We have shown that some errors arise before tracking ever begins because of an incomplete set of representations. This means that the difference between expected performance at large loads (given performance with small loads) and observed performance is likely smaller than it has appeared in previous reports.

Practically, there are a number of ways to include expectations about numerical uncertainty in formal models. For example, models could use distributions like the ones we collected in Experiments 1 and 2 to determine, probabilistically, how many targets a model should start with given an initial set. Further, one could model explicit inferences about the presence of targets given noisy samples. And finally, models of image segmentation could be integrated with models of tracking, accounting for the process of tracking at an earlier starting point in the visual pathway than is typical. Only via some combination of these approaches will we be able to evaluate the role of limited resources in limiting tracking abilities.

Conclusion: Tracking From Uncertain Visual Inputs

One of the major advances in theories of visual tracking has involved the identification of uncertainty in visual processing that leads to errors; for instance, uncertainty about the location of any given target at some moment in time. At the core of these insights is the reminder that tracking mechanisms may operate over object representations, but that object representations are extracted from images. As a result, tracking can be limited by the inescapable computational challenges associated with image processing. In the current study, we have shown that image processing results in uncertainty about more than just the features of individual objects. It can result in uncertainty about the very presence of objects.

References

Alvarez, G., & Franconeri, S. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*(14), 1–10. doi:10.1167/7.13.14

Bae, G. Y., & Flombaum, J. I. (2012). Close encounters of the distracting kind. Explaining the cause of visual tracking error. *Attention, Perception, & Psychophysics, 74,* 703–715. doi:10.3758/s13414-011-0260-1

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition, 86,* 201–221. doi:10.1016/S0010-0277(02)00178-6

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition, 98,* 199–222. doi:10.1016/j.cognition.2004.09.011

Bays, P. M., & Husain, M. (2008). Dynamics shifts of limited working memory resources in human vision. *Science, 321,* 851–854. doi:10.1126/science.1158023

Brainard, D. H. (1997). The psychophysics toolbox, *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Brannon, E. M. (2006). The representation of numerical magnitude. *Current Opinion in Neurobiology, 16,* 222–229. doi:10.1016/j.conb.2006.03.002

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology, 18,* 425–428. doi:10.1016/j.cub.2008.02.052

Cantlon, J. F., & Brannon, E. M. (2007). Basic math in monkeys and college students. *Plos Biology, 5*(12), e328. doi:10.1371/journal.pbio.0050328

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences, 9,* 349–354. doi:10.1016/j.tics.2005.05.009

Chesney, D. L., & Haladjian, H. H. (2011). Evidence for a shared mechanism used in multiple-object tracking and subitizing. *Attention, Perception & Psychophysics, 73,* 2457–2480. doi:10.3758/s13414-011-0204-9

Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. H. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology, 80,* 2657–2670.

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *PNAS: Proceedings of the National Academy of Sciences of the United States of America, 108,* 19552–19557. doi:10.1073/pnas.1113195108

Drew, T., Horowitz, T. S., & Vogel, E. K. (2012). Swapping of dropping? Electrophysiological measures of difficulty during multiple object tracking. *Cognition*. Advance online publication. doi:10.1016/j.cognition.2012.10.003

Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *The Journal of Neuroscience, 28,* 4183–4191. doi:10.1523/JNEUROSCI.0556-08.2008

Durgin, F. H. (1995). Texture density adaptation and the perceived numerosity and distribution of texture. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 149–169. doi:10.1037/0096-1523.21.1.149

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8,* 307–314. doi:10.1016/j.tics.2004.05.002

Franconeri, S. L., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition, 113,* 1–13. doi:10.1016/j.cognition.2009.07.002

Franconeri, S. L., Jonathan, S., & Scimeca, J. (2010). Tracking multiple objects is limited only by object spacing, not speed, time, or capacity. *Psychological Science, 21,* 920–925. doi:10.1177/0956797610373935

Haladjian, H. H., & Pylyshyn, Z. W. (2011). Enumerating by pointing to locations: A new method for measuring the numerosity of visual object representations. *Attention, Perception, & Psychophysics, 73,* 303–308. doi:10.3758/s13414-010-0030-5

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science, 17,* 572–576. doi:10.1111/j.1467-9280.2006.01746.x

Horowitz, T. S., & Cohen, M. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception, & Psychophysics, 72,* 1765–1775. doi:10.3758/APP.72.7.1765

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106,* 1221–1247. doi:10.1016/j.cognition.2007.06.004

Jordan, K. E., & Brannon, E. M. (2006). Weber's Law influences numerical representations in rhesus macaques (Macaca mulatta). *Animal Cognition, 9,* 159–172. doi:10.1007/s10071-006-0017-8

Jovicich, J., Peters, R., Koch, C., Braun, J., Chang, L., & Ernst, T. (2001). Brain areas specific for attentional load in a motion-tracking task. *Journal of Cognitive Neuroscience, 13,* 1048–1058. doi:10.1162/089892901753294347

Ma, W. J., & Huang, W. (2009). No capacity limit in attentional tracking: Evidence for probabilistic inference under a resource constraint. *Journal of Vision, 9*(11), 1–30. doi:10.1167/9.11.3

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience, 32,* 185–208. doi:10.1146/annurev.neuro.051508.135550

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate visual cortex. *Science, 297,* 1708–1711. doi:10.1126/science.1072493

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *PNAS: Proceedings of the National Academy of Sciences of the United States of America, 101,* 7457–7462. doi:10.1073/pnas.0402239101

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision, 10,* 437–442. doi:10.1163/156856897X00366

Platt, J. R., & Johnson, D. M. (1971). Localization of position within a homogeneous behavior chain: Effects of error contingencies. *Learning and Motivation, 2,* 386–414. doi:10.1016/0023-9690(71)90020-8

Purves, D., Wojtach, W. T., & Lotto, R. B. (2011). Understanding vision in wholly empirical terms. *PNAS: Proceedings of the National Academy of Sciences of the United States of America, 108,* 15588–15595. doi:10.1073/pnas.1012178108

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80,* 127–158. doi:10.1016/S0010-0277(00)00156-6

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197. doi:10.1163/156856888X00122

Ross, J., & Burr, D. C. (2010). Vision senses number directly. *Journal of Vision, 10*(2), 1–8. doi:10.1167/10.2.10

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple-object tracking. *Cognition, 80,* 159–177. doi:10.1016/S0010-0277(00)00157-8

Smith, K., Gatica-Perez, D., & Odobez, J. M. (2005, June). *Using particles to track varying numbers of objects*. Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA.

Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 331–351. doi:10.1037/0096-1523.19.2.331

van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *PNAS: Proceedings of the National Academy of Science, 109,* 8780–8785. doi:10.1073/pnas.1117465109

Vul, E., Frank, M. C., Alvarez, G. A., & Tenenbaum, J. B. (2009). Explaining human multiple objects tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Advances in Neural Information Processing Systems, 22,* 1955–1963.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10,* 130–137. doi:10.1111/1467-9280.00120

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology, 24,* 295–340. doi:10.1016/0010-0285(92)90010-Y

Received July 31, 2012
Revision received November 5, 2012
Accepted November 21, 2012 ■