High-capacity, transient retention of direction-of-motion information for multiple moving objects

Christopher Shooner

Department of Electrical and Computer Engineering, University of Houston, Houston, TX, USA

Srimant P. Tripathy

Bradford School of Optometry and Vision Science, University of Bradford, Richmond Road, Bradford, UK

Harold E. Bedell

Center for Neuro-Engineering and Cognitive Science, College of Optometry, University of Houston, Houston, TX, USA

Haluk Öğmen

Department of Electrical and Computer Engineering, Center for Neuro-Engineering and Cognitive Science, University of Houston, Houston, TX, USA

The multiple-object tracking paradigm (MOT) has been used extensively for studying dynamic visual attention, but the basic mechanisms which subserve this capability are as yet unknown. Among the unresolved issues surrounding MOT are the relative importance of motion (as opposed to positional) information and the role of various memory mechanisms. We sought to quantify the capacity and dynamics for retention of direction-of-motion information when viewing a multiple-object motion stimulus similar to those used in MOT. Observers viewed three to nine objects in random linear motion and then reported motion direction after motion ended. Using a partial-report paradigm and varying the parameters of set size and time of retention, we found evidence for two complementary memory systems, one transient with high capacity and a second sustained system with low capacity. For the transient high-capacity memory, retention capacity was equally high whether object motion lasted several seconds or a fraction of a second. Also, a graded deterioration in performance with increased set size lends support to a flexible-capacity theory of MOT.

Introduction

The "problem of phenomenal identity," as defined by the Gestalt psychologist Joseph Ternus, refers to the ability of the human visual system to establish and to maintain the identities of objects despite the fact that the attributes that define these objects can undergo drastic changes (Ternus, 1926, 1938). For example, as different perspective views of a moving object are imaged on the retina over time, the position, shape, size, color, and texture of the moving object can change. Nevertheless, our visual system is capable of preserving the identity of the object (e.g., *my neighbor's dog*) despite these drastic featural changes. This presents a dual problem for the visual system: On the one hand, how are identities of objects established and maintained despite changes in the defining features of the objects ("feature invariance")? On the other hand, how are specific features in the retinotopic image attributed to the correct objects ("feature specificity")?

High-level object recognition and categorization are relatively invariant to most specific features, such as position, viewing angle, shading, etc. In contrast, phenomenally, perception is feature specific, in that at any time instant, we do not perceive an abstract object devoid of its features but rather we perceive an object *with* its specific features. Over the years, several experimental paradigms and theoretical constructs have been developed to address different parts of this general problem. For example, in his seminal work, Ternus used an experimental paradigm, known now as "Ternus" or "Ternus-Pikler displays" to probe conditions that lead to the maintenance or exchange of object identities (Ternus, 1926, 1938). These displays are still widely used to investigate both the problem of feature invariance (rev. Petersik & Rice, 2006) and feature specificity (Boi, Öğmen, Krummenacher, Otto, & Herzog, 2009; Öğmen, Otto, & Herzog, 2006; Öğmen & Herzog, 2010; Otto, Öğmen, & Herzog, 2008). Kahneman, Treisman, and Gibbs (1992) used an "object-specific priming" paradigm and proposed that phenomenal identity of objects is

created by opening "object files" and are maintained by indexing these object files by their instantaneous location. According to their account, feature specificity is obtained by inserting features into object files.

Pylyshyn and Storm ([1988](#)) devised an experimental paradigm, known as multiple-object tracking (MOT), to investigate how identities of objects are maintained in the absence of all features but position and motion. Observers were presented with a multiple-object motion stimulus containing ten *identical* objects in random motion. They were instructed at the beginning of each trial to selectively track a specific subset of the ten objects, and, after a period of Brownian-type motion, tested on their ability to distinguish which objects belonged to this target set. Subjects could typically track a subset of up to five targets with 85% accuracy. Experiments of this type can be said to use a report criterion of object identity or class (i.e., target or not). It is clear, however, that maintenance of this information requires the use of other stimulus features, viz., object position or motion. This is apparent from the fact that object-class information can only be obtained *directly* from the stimulus at the beginning of the trial, when targets are assigned, and any stimulus information used to maintain this assignment during object motion (when all objects are identical) must come from other stimulus features. If we therefore consider object class a second-order stimulus feature, we are left with the questions of what first-order features contribute to its maintenance, and what mechanisms realize this processing. This problem is broad, and the present work relates indirectly to several aspects of the issue; namely, the unknown significance of direction-of-motion information and the unknown mechanisms by which memory is used for multiple-object tracking.

The question of motion processing in MOT has been addressed by several researchers. Keane and Pylyshyn ([2006](#)) tested the hypothesis that motion trajectories are extrapolated to allow prediction of the object's future location. To this end, they modified the MOT paradigm to include a brief gap, midway through object motion, during which objects disappeared. Their results show that tracking performance is least hurt by this interruption if the objects reappear at their last seen locations; the alternative condition, in which objects reappeared at a location that would be predicted by their previous trajectories, showed significantly lower performance levels. The authors therefore rejected the "prediction" hypothesis, concluding instead that proximity is the most important factor determining object continuity over a gap. This result was replicated by Fencsik, Klieger, and Horowitz ([2007](#)), with an important addition and a new conclusion. They compared the "prediction" case, in which moving objects disappeared and reappeared at a forward-extrapolated location, to a new condition, in which an identical "jump" was made from the site of disappearance to the site of reappearance, but with no object motion occurring before disappearance. In this condition, objects simply remained

in their initial locations for several seconds before the "gap," at which time they disappeared and reappeared at a new location. Tracking performance was evaluated as before by the observer directly reporting which objects were targets. Performance was significantly higher in the "motion-preview" condition than in the "static-preview" condition, but only when observers were tracking one or two objects; the trend was reversed when four objects were tracked. The authors concluded that motion information is used in multiple-object tracking but is likely not as useful as positional information (proximity), which is apparently the dominant factor.

These studies were aimed at determining the relative importance of two first-order stimulus features, position and direction of motion, but again used as their response criterion the second-order feature of object class. Any task of this sort, requiring maintenance of object identity, necessarily involves feature invariance and therefore may not be the best way to study feature specificity directly. Addressing the broad question of phenomenal identity in natural viewing conditions requires that these complementary problems be studied in parallel. Regarding the involvement of motion and position cues, Iordanescu, Grabowecky, and Suzuki ([2009](#)) showed recently that direction of motion can affect the report of object position. In their experiment, observers viewed a multi-colored MOT stimulus with three targets (of different color) and seven distracters with similar color variety. At the end of each trial, all objects disappeared, an audible cue named one of three colors, and the observer reported the location of the target having that color using a computer mouse. Here the maintenance of object class was implicitly required, as accurate performance involved reporting on the correct object of the randomly named color. The report criterion, however, was object position. Direct analysis of this first-order property allowed several interesting conclusions, but of special interest here is the finding that the distribution of localization errors was skewed toward locations *ahead* of the object's actual final position. This dependence on object trajectory supports the claim of Fencsik et al. ([2007](#)) that both position and motion contribute to the localization of target objects.

The issue of extraction of direction-of-motion information, when tracking multiple trajectories, was systematically studied by a series of experiments looking at the thresholds for detecting a change in the direction of motion of a target object (a moving dot) undergoing bilinear motion, in the presence of other similar distracter objects in linear motion (Narasimhan, Tripathy, & Barrett, [2009](#); Tripathy & Barrett, [2004](#)). Set-size effects on deviation thresholds were found to be dramatic, with thresholds increasing almost 20-fold for a set size of four. These set-size effects were unlike those seen in Pylyshyn and Storm ([1988](#)), where performance dropped little over the same range of set sizes when object class was to be reported. However, when the target deviation was substantially supra-threshold, the effective number of trajectories

tracked was comparable to the number of objects tracked in the Pylyshyn paradigm (Tripathy, Narasimhan, & Barrett, 2007).

The definition of the term "tracking" may vary across these studies. For example, some researchers, implicitly or explicitly, define the term "tracking" as the ability to maintain object-class distinction during motion. In a typical MOT study, this would correspond to distinguishing the target items from the distractors. Deviation detection experiments also require detailed direction-of-motion information. The term multiple trajectory tracking (MTT), as opposed to MOT, can be used to highlight this important difference. Because the experiments reported here did not use distractors (see the Methods for each experiment, below), the term tracking, as defined above, does not strictly apply to our study. Table 1 summarizes the differences between the present study and the broad classes of MOT and MTT studies. However, although the present experiments were not designed to directly test the significance of motion information for tracking, they extend substantially our understanding of the amount of motion information extracted by the visual system while viewing stimuli that consist of multiple moving objects. We therefore will consider our results within the context of previous MOT and MTT studies below.

The original mechanistic explanation of MOT, provided by Pylyshyn and Storm (1988), involved a low-level, encapsulated visual system that is responsible for tracking individual objects by means of internal pointers, or indices. This system is preattentive and parallel, with a capacity determined by internal architecture, namely by how many indices are available for assignment to individual objects. Details are not given to explain exactly how a pointer continues to point to its assigned object as that object moves among others, but the assertion is that this maintenance is performed automatically at a level lower than those involving higher functions such as endogenous attention or short-term memory. Other models of MOT, especially those that propose a serial mechanism for processing objects sequentially, rely explicitly on visual short-term memory (Oksama & Hyona, 2008). Additionally, there is evidence that the transient streaks which a moving object leaves in *sensory* memory may be used in tracking an object's trajectory (Narasimhan et al., 2009; Tripathy, Öğmen, & Narasimhan, in press).

A first step in establishing whether sensory memory is used for the general problem of phenomenal identity is to determine what type of information is stored and how long this information remains available. Thus, the goals of our work were to establish whether direction of motion is stored in visual memory during multiple-object motion tracking and, if so, to determine the characteristics of the underlying memory mechanisms. Visual memory for different stimulus attributes such as orientation, speed, and contrast has been investigated by several researchers (for a review, see Magnussen & Greenlee, 1999). It has been shown that a single direction of motion can be stored and recalled reliably in visual memory for extended periods of time both in humans (Blake, Cepeda, & Hiris, 1997) and monkeys (Pasternak & Zaksas, 2003). Blake et al. (1997) tested memory for multiple directions of motion by presenting different directions sequentially and in parallel. In the former case, they found a strong decay in performance as the number of serially presented directions increased, with a trend showing a strong primacy and a relatively weak recency effect. In the experiment more closely related to MOT, multiple directions were displayed simultaneously and the results indicated a strong degradation of performance with an increase in the number of directions but no significant temporal decay in memory after post-stimulus time intervals as long as 30 s. Blake et al. did not test explicitly whether there was a partial-report advantage, a hallmark of iconic memory (Sperling, 1960). Earlier work by Demkiw and Michaels (1976) showed a clear partial-report advantage for direction of motion. Demkiw and Michaels tested the decay of memory using three values of post-stimulus delay, 0 ms, 500 ms, and 1 s, and found only a modest decay in performance. Treisman, Russell, and Green (1975) found a partial-report advantage only in a subset of their observers. For these observers, they tested whether memory for motion direction decayed by using three cue timings: when the stimulus started to move, when stimulus motion terminated, and 1 s after motion termination. They found a significant decay of performance with a 1-s delay compared to zero delay. Taken together, these findings support the existence of visual memory for direction of motion; however, no coherent picture emerges from these studies concerning the characteristics of this memory. Here we investigated the capacity and the

	Trajectories	Target/distractor distinction	Report criterion
Current study	Linear with constant speed	No	Direction of motion
MOT studies (e.g., Pylyshyn & Storm, 1988)	Complex pseudo-random trajectories often with variable speed.	Yes	Object class (target or distractor)
MTT studies (e.g., Tripathy & Barrett, 2004)	Bilinear with constant speed.	No	Trajectory deviation direction (up or down)

Table 1. Comparison of various motion tracking paradigms.

dynamics of this memory by using some of the same tests that Sperling (1960) introduced to demonstrate what was later termed "iconic memory" and adaptations of these tests used by Narasimhan et al. (2009) to study sensory memory for direction of motion when tracking multiple moving objects.

Experiment 1: Capacity of visual memory for direction of motion assessed by set size

Methods

Stimuli were created using the Visual Stimulus Generator (VSG5) video card (Cambridge Research Systems) and presented on a 20-inch NANAO FlexScan color monitor with a resolution of 656 × 492 pixels. Observers viewed the screen from a chin rest positioned at a distance of 1 meter, such that each pixel subtended approximately 1.7 minutes of visual angle and the angular size of the entire monitor screen was approximately 23 × 17.5 deg. Stimuli were presented at a frame rate of 100 Hz.

At the beginning of each trial, a variable number of objects (N = 3 through 9) appeared on the viewing screen at randomly chosen but non-overlapping locations and remained stationary until 1 s after appearance, at which time they began to move along linear trajectories, each in an independently chosen direction. Objects were circular disks with diameters of 1 deg visual angle and a luminance of 7 cd/m^2 on a 65-cd/m^2 background (dark objects on a light field). In addition to multiple-object conditions, a baseline condition containing only a single moving object was included for comparison. Object speed was fixed at 5 deg/s for all trials. Objects did not interact with each other, i.e., they moved across each other without any change in their velocity, but "bounced" off the edges of the viewing screen by reversing either the horizontal or vertical component of their velocity.

Object motion continued for 5 s. When object motion ended, one randomly chosen object was immediately marked with a small red dot at its center to prompt the observer's response. The observer then reported the direction of motion of the selected object using a computer mouse. In the case of an object that bounced off an edge and thus changed direction of motion, the observer was asked to report the final direction of motion. To report, the observer moved the mouse cursor, causing a direction indicator to appear. This indicator was a line segment of constant length extending from the center of the object towards the position of the cursor. The direction of the indicator could be controlled with the mouse cursor with 1 deg resolution. Report was carried out by adjusting the indicator to reflect the direction of object motion, and then

clicking the mouse button. This method was chosen to avoid limiting the observer's response speed.[1] After the observer clicked the mouse, an additional indicator appeared to indicate the true direction of motion. Note that no objects were assigned as targets, but all objects were equally likely to be selected for report. There was no noticeable motion aftereffect with our stimuli.

Trials with different set sizes were randomly interleaved within an experimental block. One hundred trials per set size summed to 700 total trials, which were broken into five separate sessions of approximately 20 minutes each. Short training sessions were performed before the experiment to ensure stability of observer performance. In no case was a significant learning effect observed. Measurement of single-object performance was added as a baseline test. To obtain these data, two sessions of 40 trials each were conducted separately from the primary sessions.

Four observers, including one author (CRS), participated in this and all subsequent experiments. With the exception of the author, all subjects were naïve to the specific purposes of the experiments. After completion of the main experiment, three of the four observers (including the author) repeated the entire experiment with the duration of motion reduced from 5 s to 200 ms. This test was motivated by the initial results of Experiment 1 and its significance will be discussed later.

Results

Each trial yielded a scalar value representing the error angle, or the difference in deg between actual and reported directions of motion. This angle was measured, with 1-deg resolution, relative to the true direction of motion (assigned = 0 deg) with clockwise errors represented by positive values (up to 180 deg) and counterclockwise errors by negative values (to −180 deg). To assess performance, the (unsigned) magnitude of the error angle was averaged across all trials in a condition. The resulting mean performance is plotted in Figure 1 for all four observers tested in the 5-s stimulus duration condition and for the three observers tested in the 200-ms stimulus duration condition.

For two of the three observers, there was no noticeable difference between the two conditions, whereas one observer showed a clear *increase* in performance at the shortened duration.

Analysis

A two-way repeated-measures across-subjects ANOVA with Greenhouse–Geisser correction for non-sphericity shows that the effect of set size is significant ($F[6,12]$ = 42.15, p = 0.02) and neither duration nor the interaction between set size and duration are significant (for duration: ($F[1,2]$ = 2.91, p = 0.23); for the interaction term: $F[6,12]$ =
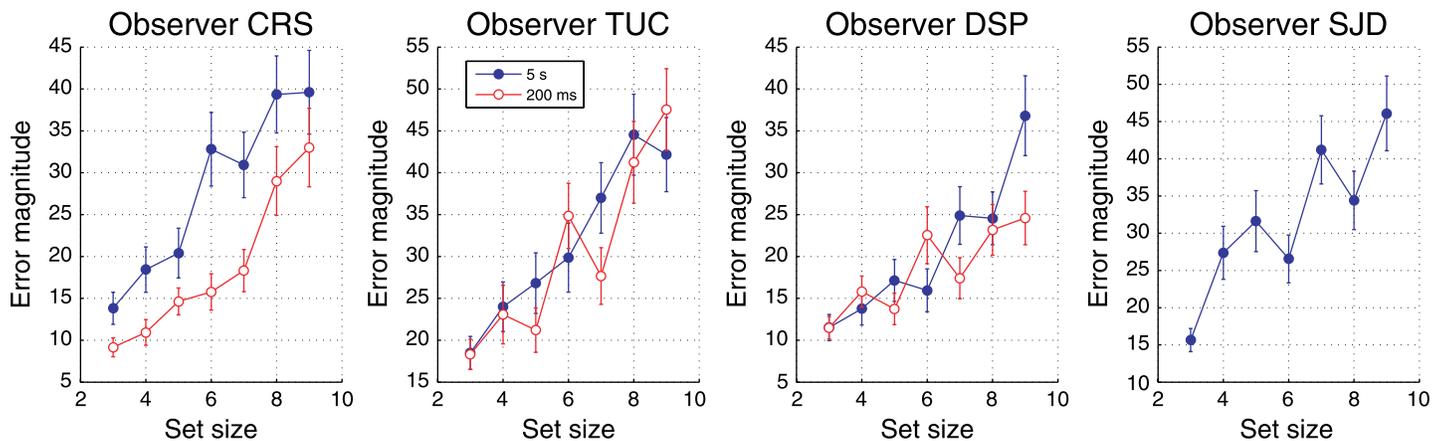
Figure 1. Magnitude of error angle versus set size for four observers. For three observers, data are shown for both 5 s and 200 ms motion duration. Bars represent standard error of the mean.

0.80, $p = 0.51$). In order to assess observer-specific effects, the two duration conditions were compared using a two-way within-subject ANOVA which compared mean error with respect to set size and stimulus duration, separately for each of the first three observers shown in Figure 1. The result was, as expected, that set size was a significant main effect for all observers, but stimulus duration was significant only for observer CRS. The interaction between set size and stimulus duration was not significant for observer CRS or observer TUC but was significant for observer DSP. Intuitively, this cross term describes the effect of stimulus duration on the observed set-size *trend*. For the first two observers, the shape of the set-size effect was preserved in the shortened duration condition, but for observer DSP, this shape varied somewhat, as seen in Figure 1, where this observer's errors increased more steeply with set size in the longer duration condition.

The measured error magnitude was converted into a normalized metric of performance by scaling and inversion such that an error of 180 deg was assigned a performance value of 0.0, and a perfect report (0 deg error) was assigned a value of 1.0. We refer to this metric as normalized performance and use this as our primary measure of performance across conditions. The raw error angle, a, was converted to normalized performance (P_n) using the equation $P_n = 1 - |a|/180$. A uniform distribution of errors ranging from -180 to 180 deg would result in a mean P_n of 0.5. This value therefore represents the average P_n that we would expect if an observer were to guess randomly (i.e., chance performance).

For each of the four observers, normalized performance was averaged over all trials of a given set size and a duration of motion equal to 5 s. The effect of set size on performance is shown in Figure 2a. Initial inspection reveals a regular decrease in performance, which is apparent even for the three-object condition as compared with the single-object baseline. Additionally, the maximum set size ($N = 9$) yielded a mean performance above

0.75 in most cases. This is equivalent to a mean error angle of 45 deg magnitude.

Effective number of objects tracked

Effective number of objects tracked (abbreviated ENOT) is a concept used by Scholl, Pylyshyn, and Feldman (2001) as well as Tripathy et al. (2007) to describe MOT or MTT performance in terms of the number of objects that the observer would have to track perfectly in order to achieve the observed percentage of correct responses. This method assumes that a perfectly tracked object is always reported correctly, and report on all other objects is random. Because our normalized performance is analogous to percent correct, with perfect and chance performance defined as 1.0 and 0.5, respectively, our performance data can also be converted to effective number of objects tracked. The functional definition of ENOT used here is

$$\text{ENOT} = N(2P_n - 1), \tag{1}$$

where P_n represents normalized performance. The normalized performance data from Figure 2a, transformed to ENOT, are shown in Figure 2b. Note that the one-to-one line represents perfect performance, and the deterioration in performance with set size appears to plateau to a value near 5 objects. To test this observation, we fit each of the four curves in Figure 2b with a second-order polynomial (not shown). For all observers, the quadratic terms were statistically significant ($p < 0.05$ for all observers). The resulting parabolas for the four observers reach maxima of 5.0, 4.6, 6.3, and 4.7 objects.

In the context of Pylyshyn's MOT paradigm, Hulleman (2005) suggested that the equation for ENOT above should be adjusted to take into account that:

    i. observers might have tracked v of the distractors along with m of the targets and used the information
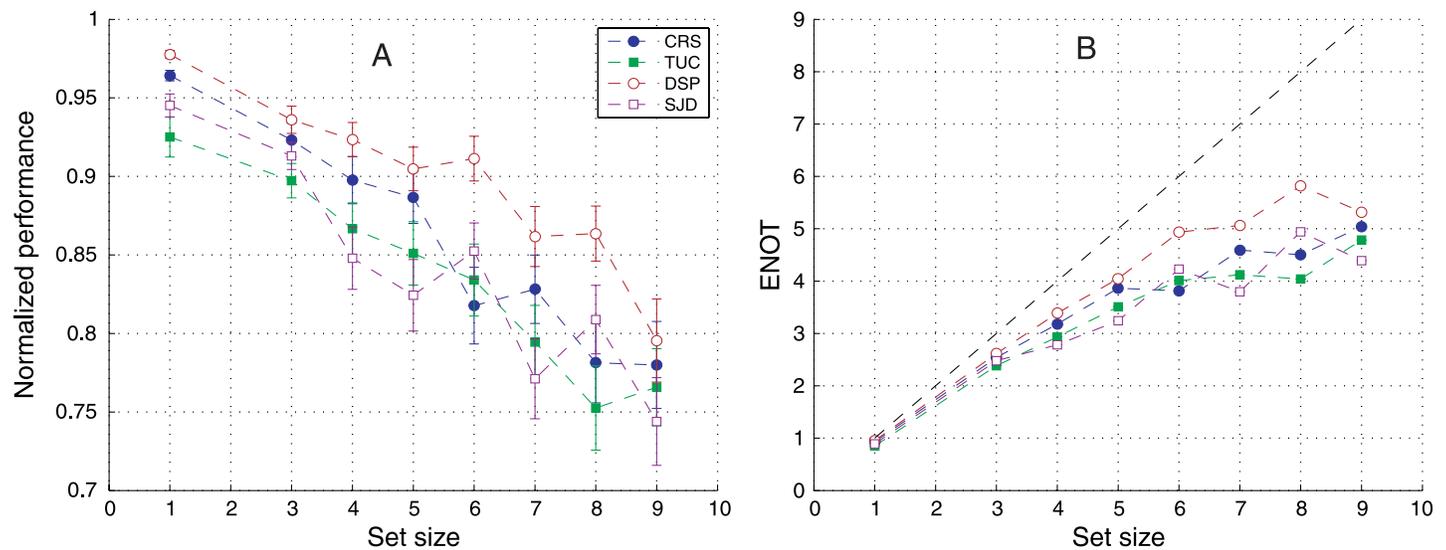
Figure 2. (a) Effect of set size on normalized performance in a direction-of-motion reporting task, for a 5-s duration of motion. (b) Normalized performance data were transformed to effective number of object tracked (ENOT).

regarding the tracked distractors to improve their chances of guessing the identity (target vs. distractor) of any untracked probe; and

ii. observers might know how many targets and distractors were present among the untracked objects and that could influence their chances of guessing the identity of any untracked probe.

In our experiment, the equations for ENOT were not adjusted because there were no distractors of the kind present in traditional MOT stimuli (all objects in our stimuli were potential targets) and awareness of the number of untracked targets would not have influenced the accuracy for reporting the direction of motion of any untracked probe.

In addition to mean performance, it is possible to view the reporting error as a random variable and analyze its distribution. This statistical characterization of performance is provided in the Appendix A.

Discussion

The nature of the observed set-size effect is gradual and consistent, suggesting no special significance of any one set size in the range tested. A fixed-architecture model of MOT, such as that originally proposed by Pylyshyn and Storm (1988), would claim that the tracking system has a certain capacity and that set size should have no effect on performance until this capacity is exceeded, after which performance should drop off with increased set size (proportional to $1/N$). Instead of a trend of this type, we observed an apparently linear decrease in performance starting with a set size of one (Figure 1). Linear regression

of the performance data shown in Figure 1 was consistent across observers and represents a constant increase in error of approximately 4 deg per object. These findings are in agreement with both Alvarez and Franconeri (2007) and Howard and Holcombe (2008), who also show graded set-size effects in two different tracking tasks. A linear increase in the deviation threshold, starting with a set size of one, was observed in Tripathy and Barrett (2004) but the drop in performance was dramatically steeper than that observed here.

Although these results disagree with the idea of a fixed tracking capacity, it is interesting to note that our analysis of effective number of objects tracked (ENOT) hints at an asymptotic value near 4 or 5 objects in the limit of very large set sizes. This is reminiscent of Alvarez and Cavanagh (2004), whose analysis of the trade-off between memory capacity and object complexity yielded a theoretical upper bound on capacity of 4.7 objects (found in the limit of very low object complexity). In addition, Tripathy et al. (2007) found that ENOT increased with the angle of deviation in their task, reaching around 4 or 5 objects for a target deviation of 76°. The nature of their stimulus did not permit testing with larger deviations.

Performance on our task was tested, for three observers, in two drastically different motion-duration conditions. The primary experiment used a duration of 5 s, similar to typical MOT stimuli, but the second condition involved motion that lasted only 200 ms. While there remains some likelihood of a motion aftereffect (MAE) for stimulus duration of 5 s (although the observers did not report any MAE), the brief stimulus duration eliminates any such possibility. Moreover, in this fraction of a second, there is not enough time for eye movement or endogenous directing of attention, yet performance levels for two observers appeared unaffected by the shortened duration,

and in the case of a third observer, performance actually improved. The implication is that the retention of motion information, reflected in observer performance, is not reliant on selective attention being deployed during object motion. Attention is focused sharply, however, immediately after motion termination, when a single target object is cued. Together, these facts raise the question of just how important spatial attention is for the encoding of object information in MOT and what mechanisms could lead to the high retention of object motion information in the absence of attention-enhanced object encoding. One possible answer is that sensory memory is capable of storing object motion information and that this capability may, in some cases, be utilized for report. An extensive discussion of sensory memory and MOT can be found in Narasimhan et al. (2009).

The progressive drop we found with an increase in the number of objects and equivalently the number of directions of motion are in agreement also with the results of Blake et al. (1997). In this sense, the memory involved in storing the direction of motion, while of high capacity, does not follow the definition of a "pure iconic memory" which would predict only a negligible decay in performance. Before we discuss potential reasons for this difference, we will address in the next section the temporal dynamics of the memory. As discussed in the Introduction, Blake et al. did not find any significant decay in memory for delays lasting several seconds. Similarly, Demkiw and Michaels (1976) demonstrated a partial-report advantage but found only a modest temporal decay with a 1-s delay period. In contrast, Treisman et al. (1975) found a significant decay in 1 s, although they did not have any intermediate samples to characterize the time course of the decay. Because rapid decay for delays less than 1 s and the partial report advantage are hallmarks of iconic memory, in the following two sections we will examine these two issues by using variants of the multiple-object tracking stimulus and task introduced in the first experiment.

Experiment 2: Temporal characteristics of memory

Methods

Methods were identical to Experiment 1, with the following adjustments. The number of objects in the stimulus was fixed at nine. Duration of motion was fixed at 200 ms, and after termination of motion, a variable-duration delay preceded the cueing of a single object for report. During this interval, the stimulus was held constant, displaying the final frame of motion, all objects visible and unchanged. The seven cue delays used were 0, 50, 100, 250, 500, 1000, and 3000 ms. After the delay, a randomly selected object was cued by the appearance of a red dot at its center, and report was carried out as in Experiment 1. The seven cue-delay conditions were randomly interleaved, 100 trials each, for a total of 700 trials per observer, which were separated into five sessions of approximately 20 minutes each.

Results

Figure 3 shows mean normalized performance as a function of cue delay for each observer. A sharp deterioration from 0 to 250 ms is visible for all observers, but performance levels off by approximately 1 s to a final value that varies among observers. Note that this asymptotic value ranges from approximately 0.6 to 0.7 (normalized units).

Analysis

A simple descriptive model of exponential decay was used to analyze the decay in performance with cue delay. This model was generalized to allow an arbitrary maximum at zero delay and an arbitrary final value. The functional form is given by

$$P_n(t) = A + B \ e^{-t/\tau}. \tag{2}$$

This model was fit to each observer's data individually, and the resulting model outputs are plotted as the smooth curves in Figure 3, along with the behavioral data. The time constant τ varied among observers from 87 ms (TUC) to 433 ms (DSP), while the asymptotic performance value, A, varied from 0.62 (TUC) to 0.7 (CRS) in normalized units.

Discussion

Clearly cue delay has a considerable effect on performance, and this effect is strongest for values less than 1 s. This rapid decay suggests a strong similarity to iconic memory. The trend appears to level off at larger delays, and in fact t-tests show that for no observer was the performance significantly different in the 1-s-delay and 3-s-delay conditions ($p > 0.3$ for observer DSP, $p > 0.9$ for others). Fitting an exponential-decay model to the data allowed a more formal description of these trends and quantified differences among observers. The relatively stable performance for delays larger than 1 s suggests that a subset of the contents of rapidly decaying iconic-type memory has been stored in a longer-lasting memory.

Towards our goal of a characterization of memory capacity for object motion information, the exponential
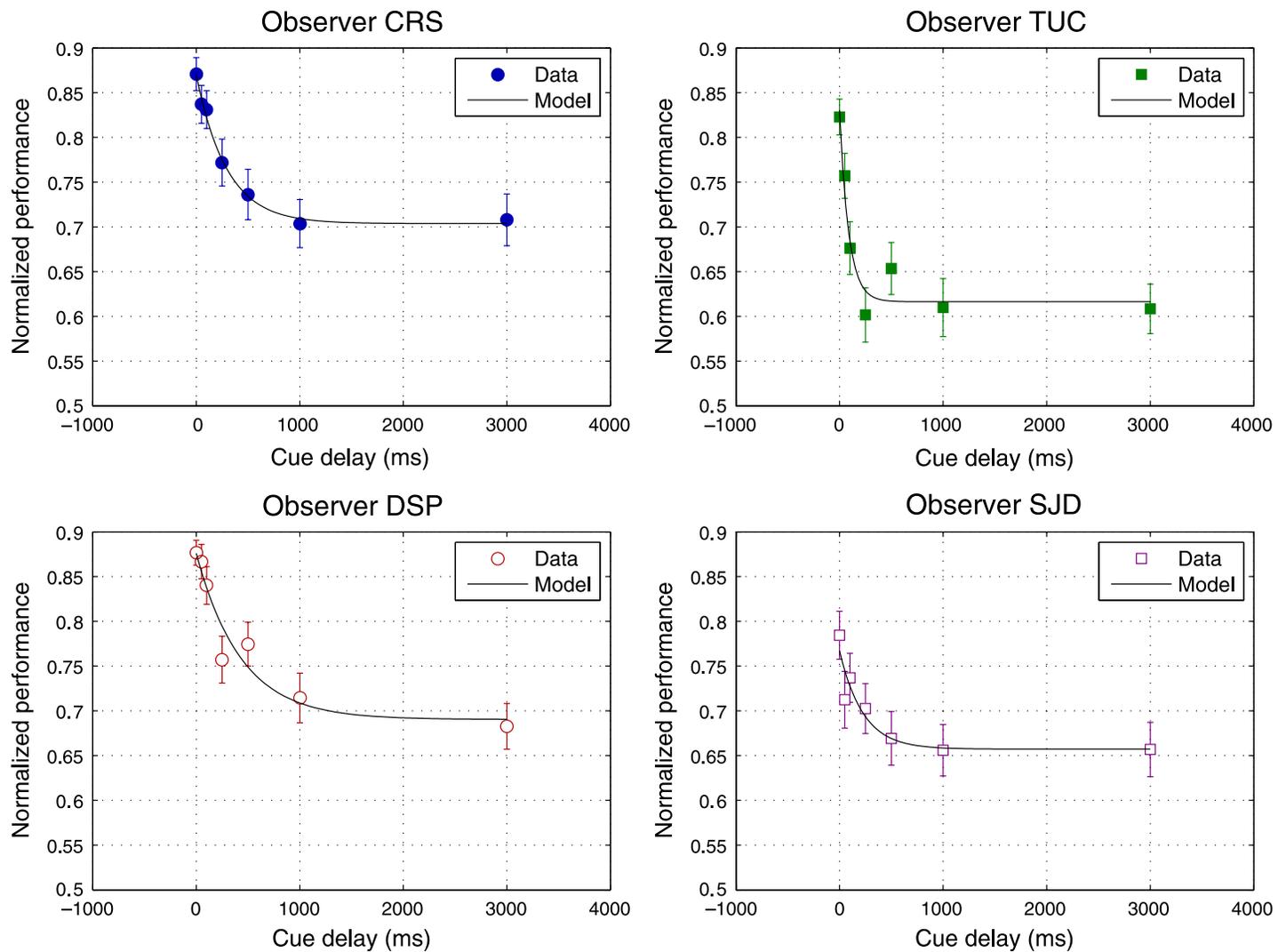
Figure 3. Mean performance as a function of cue delay was modeled as an exponential decay.

model is a satisfactory description of how recall performance varies along the time dimension. More mechanistic considerations, however, allow an interesting interpretation of the results. As in the case of the original work on iconic memory (Sperling, 1960), the sharp deterioration in performance in the first second after stimulus termination lends itself naturally to an explanation involving transient sensory memory mechanisms. Similar steep drops in performance, seen in a deviation detection task when a delay of up to 400 ms was introduced halfway through the motion trajectory, were also attributed to sensory memory (Narasimhan et al., 2009).

What processes can be grouped under the title "sensory memory" is not clear, but in the present discussion we use this term generally to encompass all low-level, stimulus-dependent, visual mechanisms in which a brief stimulus may cause a transient sensory response which outlasts the stimulus itself. Due to the transient nature of sensory memory, it is necessary for information to be transferred to higher level mechanisms capable of storing the

information long enough to allow processing and report. The nature of these higher mechanisms is not the topic of interest here, but the observed cue-delay effect shows clearly that whatever mechanisms are responsible for retaining stimulus information for longer than 1 s lack the capacity to match those mechanisms that hold the information for the first second post-stimulus.

Note that the zero-delay condition in this experiment is identical to the 9-object condition in Experiment 1, and we expected similar performance in the two cases. In fact, observers performed noticeably better in Experiment 2 on this same stimulus. Possible causes of this include a practice effect and a variation in strategy: subjective reports from observers suggest that different tactics may be used in tracking large numbers of objects compared to those used for smaller set sizes, and the constant set size used in Experiment 2 may have supported a more stable strategy. Experiment 1 can be viewed as consisting only of the zero-delay condition. Given the clear deterioration of performance with larger delays, it is safe to assume that

the generally high performance levels seen in Experiment 1 are contingent on this immediate-cueing aspect of the stimulus. Variation in strategy may also account for the difference between our results and those of Blake et al. (1997) who did not find a decrease in performance as a function of cue delay and stated that "this hallmark characteristic is not present in our partial report results. We believe, therefore, that the sequence of events in our partial report task place the task outside of the realm of iconic memory" (Blake et al., 1997, p. 360). Their partial report experiment consisted of a 3 × 3 design (number of distinct directions of motion (3, 5, and 7) and cue delay (0, 10, and 30 s)). The nine stimulus conditions were randomly interleaved from trial to trial within a block. Because our experiments had 9 different directions, we can compare our results to theirs for their 7-direction condition. Their data show average motion-direction errors of approximately 50, 65, and 60 deg for cue delays of 0, 10, and 30 s, respectively. In terms of normalized performance, these error values correspond to 0.72, 0.64, 0.67, respectively. Inspection of our Figure 3 indicates that the normalized-performance values found in their experiment are in the asymptotic range of our observers' normalized performance. As mentioned earlier, this asymptotic performance remains stable between the 1- and 3-s intervals used in our study and is likely to extend to larger cue-delays as observed by Blake et al. As such, it appears that the experimental design used by Blake et al. did not capture the initial transient part of the performance curve but rather sampled only its asymptotic (stable) regime. This could be due to their mixed design that included widely separated cue delays: By distributing their attention span over the 0- to 30-s post-stimulus interval, observers may not have been effective in reading out the contents of high-capacity memory which decays within the first 300 ms of the post-stimulus interval.

In sum, our results on the temporal dynamics of memory for motion direction are in agreement with those of Treisman et al. (1975). In addition, the denser and more extended sampling of delay values allowed us to demonstrate the existence of *dual* memory systems, one a high-capacity but transient mechanism (iconic-memory type) and the second a relatively lower capacity but more sustained mechanism. This dual memory concept may reconcile the apparently contradictory findings by Blake et al. (1997) and Treisman et al. (1975).

Experiment 3: Partial versus full report

The goal of the third experiment was to apply the logic used by Sperling (1960) in order to test for a partial report advantage in our memory task.

Methods

The experimental methods used in the previous two experiments were again adjusted slightly. The number of objects was fixed at nine, and the cue was presented simultaneously with the termination of motion (no cue delay was used). Duration of motion was again fixed at 200 ms. In single-report trials, one randomly selected object was cued, as in Experiments 1 and 2; in full-report trials, all objects were cued simultaneously, indicating that the observer must report the direction of motion for each object individually. In the case of full report, the observer was free to report on all nine objects in any order. As objects were reported, they disappeared, allowing the observer to distinguish those still to report. After reporting all nine objects, the observer was shown all objects along with direction indicators showing the true and reported directions of motion.

Preliminary tests suggested that the red dot used as a cue in previous experiments was problematic for this test. In the case of full report, the appearance of this cue on all objects at once produced a highly distracting transient signal. The cue was adjusted so that, instead of a red dot appearing at the center of the object, the entire object changed color to a dark red. The luminance of this cue was chosen to approximately match the original luminance of the object, thereby minimizing the transient luminance signal. One hundred trials of each condition (single and full report) were run by each of the four observers.

Results

Assessment of full-report performance was performed separately for each observer by first averaging each report-order condition across all full-report trials. That is, the performance on the first object reported was averaged across trials, as was the second-reported object, and so on. Overall full-report performance was then taken as the average of these report-order averages. Single-report performance was simply averaged across all single-report trials. The basic comparison of full and single report is shown in Figure 4, which also shows average normalized performance as a function of report order.

A clear partial-report advantage is apparent for all four observers. Interestingly, Figure 4 also shows that the mean performance for the first object reported in full-report trials is not significantly different from the single-report mean, except in the case of observer TUC, whose first-report performance was significantly higher ($p < 0.01$).

Discussion

In the full-report condition, when observers are free to report the objects in any order, we assume that they report
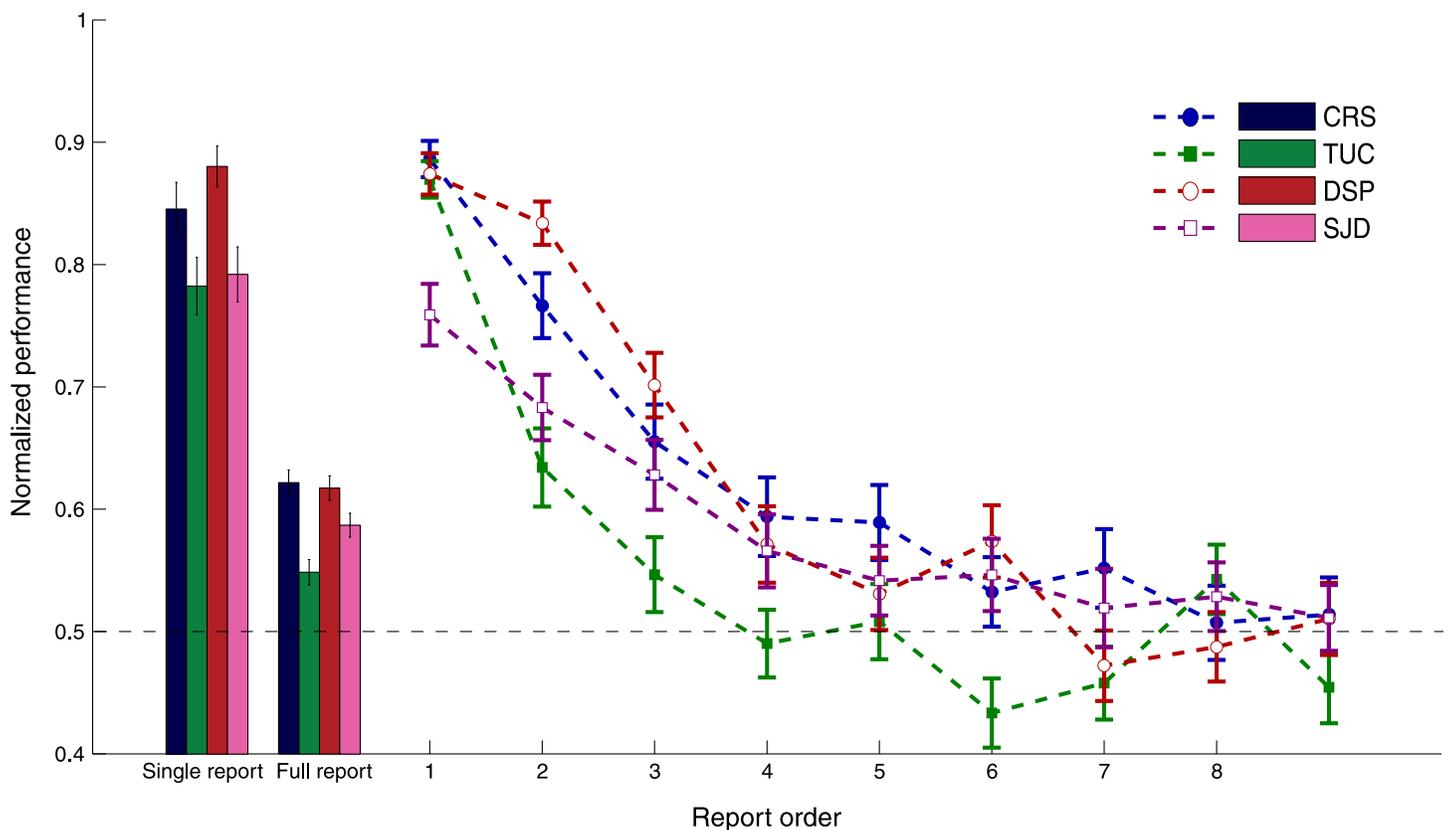
Figure 4. Mean normalized performance for single-report trials compared to full-report performance (average as well as full-report performance with respect to report order).

first on the objects for which they have the best information. This assumption is corroborated by observer's subjective descriptions of their experience as well as the report-order data shown above. It bears repeating here that in the *single*-report condition, the report requirement is randomly chosen from nine possible objects. The similarity between single-report and first-report performance therefore suggests that, for most observers, the quality of information available for the object which the observer deems "best" is actually available for *any* object, but immediate cueing of that and only that object is necessary in order for that information to be successfully utilized for report.

This observation has interesting implications for the discussion of sensory memory in this task. Obviously, information for a given object is available with high fidelity immediately after motion ends; individually cueing this object leads to the high performance levels we observe in the single-report mean. This same information, however, is apparently unavailable to the observer if, as in the full-report condition, several other objects are reported first. One cause of this may be that sensory memory plays an important role in the single-report task, but the short duration of sensory memory does not allow accurate processing of a large number of objects before the memory trace has faded. Another possible contribution

is from interference caused by the action of reporting; it is possible that object information held in memory deteriorates more quickly when attention is directed to the processing and report of other objects. It would be interesting to investigate whether the nature of the report criterion, for example categorical (e.g., letter identification) versus analog (direction of motion), can influence this putative interference.

Narasimhan et al. (2009) investigated the issue of partial-report advantage in their deviation detection task by either having the deviating target dot change color halfway through the dot trajectory or by having the non-deviating distracter dots disappear at the same instant that the target dot started to deviate. In both cases, the deviating trajectory was uniquely identified and the observer was required to report the direction of deviation of the identified trajectory only. Having the target trajectory uniquely identified halfway through the trial resulted in only a small reduction in deviation thresholds, i.e., a small partial-report advantage. The authors speculated that a deviation could only be detected reliably after a substantial part of the post-deviation trajectory was presented, but while waiting for the post-deviation trajectory to be presented, the pre-deviation trajectory would have decayed substantially. This decay may have compromised the observed partial report advantage. In the current

experiment, all task-relevant information is available at the time of cueing, allowing the retrieval process to begin immediately, before sensory memory has decayed. This likely explains the more convincing partial-report advantage as compared with previous studies.

The partial-report advantage reported here is also consistent with the findings of Demkiw and Michaels (1976) and Treisman et al. (1975). However, Treisman et al. found a weaker partial-report advantage, which was present only in a subset of their observers. Their observers were required to report a subset of six objects undergoing apparent motion around circular paths in six spatially distinct areas of the viewing screen. The effect measured in the current study may have been of a greater magnitude because of the use of linear trajectories and analog reportings instead of circular paths and binary reports (observers reported clockwise or counterclockwise motion rather than direction). The overall weaker magnitude of the effect generated with the stimulus design used by Treisman et al. may explain why only a subset of their observers produced a statistically significant difference between partial and full report conditions.

General discussion

The primary goal of this work was a functional description of memory capacity and dynamics, and the structure of the experiments therefore paralleled previous work on iconic memory (Sperling, 1960). The specific memory task under consideration involved reporting direction of motion information from a multiple-object-motion stimulus, modeled after typical stimuli used in multiple-object tracking studies (cf. Cavanagh & Alvarez, 2005; Tripathy & Barrett, 2004). The results of these experiments can therefore be viewed as a corroboration and expansion of previous work on iconic memory as well as novel evidence for the role of such memory mechanisms in MTT and possibly in MOT.

Regarding the question of iconic memory, the clear partial-report advantage seen in Experiment 3 and the rapid decay of available information seen in Experiment 2 strongly suggest that high-capacity, transient memory mechanisms contribute in some form to the high performance levels achieved in the task of reporting on a single object trajectory. The reasoning behind this was outlined in the previous section; the single-report method provides a measurement of the quality of direction information available for *any* randomly selected object immediately after termination of motion and can therefore be equated with the amount of stimulus information retained in the system at time zero. Performing the same sampling of memory 1 s later, however, reveals a greatly reduced quality of information. The properties of high capacity and short duration for information storage fit best with the

low-level, stimulus-dependent, and transient nature of sensory memory.

To put the time course of these motion-related memory mechanisms into context, let us mention the study by Shioiri and Cavanagh (1992), who found persistence of motion-defined figures to be in the range of 100–150 ms. Although their stimuli were very different from ours, there is similarity in the time course of the two phenomena. At the neurophysiological level, Supèr, Spekreijse, and Lamme (2001) reported a continuing modulation of V1 neurons in monkeys to a motion-defined stimulus for approximately 250 ms. Studies that assessed the temporal integration of motion information, which should be related to the duration of persistence, also provide estimates in the range of 80–130 ms (McKee & Welch, 1985; Simpson, 1994; Snowden & Braddick, 1991).

Aspects in which our results differ from those of Sperling (1960) should be mentioned. In Experiment 2, performance deteriorated with cue delay but reached a "steady-state" value after approximately 1 s and remained at that level thereafter. Assuming this decreasing trend does indeed end after 1 s, we can take the performance at larger delays (e.g., 3 s) to represent the average amount of stimulus information that is successfully transferred from sensory memory into more stable, enduring mechanisms such as short-term memory. In fact, Sperling's results agree with this explanation exactly, as performance in his cue-delay experiment reached a minimum value equal to his estimate of immediate memory span for each observer. Our equivalent to immediate memory span is the measurement of average performance in the full-report condition of Experiment 3. Experiment 2 results, however, consistently show a "steady-state" performance significantly greater than the observer's full-report average (paired *t*-test across observers, $p < 0.001$). The explanation for this which we find most feasible is that the action of reporting interferes with memory mechanisms such that information regarding objects which are reported later is corrupted. Interference could be due in part to the reporting method chosen. As the observer reported on objects, those objects disappeared, possibly causing a distracting "off" transient. Experiment 3 shows that, in the full-report case, performance drops to near chance by the fifth to the seventh object reported. Because observers begin to respond immediately in the full-report condition, it is possible that interference occurs either in the "read-out" from sensory memory or in the maintenance of information in short-term memory thereafter.

The aspect of this work that contributes most significantly to the understanding of iconic memory relates to the nature of the information being reported. Whereas Sperling's experiments required recall of a brief, static stimulus, our stimulus is dynamic, and the report criterion (direction of motion) is, in fact, a property not of the stimulus at a single instant but of the change in the stimulus over time. This fact disagrees with the simplest idea of a "visual image," but only insomuch as a mental

representation of a physical stimulus resembles a photograph. Based on the known existence of low-level neural mechanisms sensitive to motion, it is easy to imagine that the temporal dynamics of these processes could lead to the retention of motion information post-stimulus in a manner similar to the more commonly studied persistence of static stimuli. Now, it is logical to argue, in opposition to this idea, that motion processing, by nature, must be transient and faster than processing of sustained, static stimulus content, and therefore "should not" display a significantly sustained response. A similar ecological consideration was, after all, one of the main arguments against Sperling's original model of the visual image (Haber, 1983), but now, as then, the experimental evidence shows that the information is there and available for report. If and how this is useful to perception remains to be seen.

An interesting point of discussion regarding a "dynamic icon" relates to the underlying mechanisms used to store motion information. It has been suggested that direction of motion can, in some cases, be inferred from the response of orientation-selective neurons (Geisler, 1999), the cells which we typically associate with form detection, and not motion. If this is true, and the cellular mechanisms encoding form and movement are indeed shared, then the possibility exists that the dynamic icon and the classical "static" icon are in fact the same mechanism, simply viewed from two different angles. The present results may shed some light on this question: Geisler (1999) interpreted his findings to indicate that oriented cells contribute to motion detection only for speeds greater than one object width per 100 ms. We find evidence for an icon-like retention of motion memory although our motion stimuli used a fixed speed of 0.5 object widths per 100 ms. An important consideration, however, is that the critical speed found by Geisler was a nonlinear function of object size (see Figure 2 of Geisler, 1999), and the shape of this trend suggests that objects of the size used here (1 deg) may affect oriented cells when moving at velocities less than one dot width per 100 ms. Regardless of specific mechanisms, if it is indeed true that motion information is stored not based on the relative activities of directionally selective motion detectors but rather on spatio-temporal trajectory information, then the drop of performance as a function of set size (Figure 1) may be explained based on increased spatio-temporal proximity or overlap between the stored trajectories as the number of disks is increased. Proximity or overlap may lead to decreased performance due to factors such as crowding and masking. From this perspective, stimulus duration presents a trade-off: Although it can improve performance by allowing more time to process and store stimuli, it can lead also to deterioration in performance due to increased proximity and overlap between trajectories. As mentioned above, one observer performed better at shorter stimulus duration but there was no significant effect of stimulus duration for the other two observers. Further studies are needed to test these hypotheses.

With respect to MOT, the present findings can be interpreted several ways. As a preface, it should be pointed out that the multiple-object motion stimuli used in this study differ somewhat from those used in other MOT studies: First, the duration of motion used in Experiments 2 and 3 was only 200 ms, whereas typical tracking tasks require the observer to follow object motion for several seconds at least. Experiment 1 used both the 200-ms duration and a 5-s duration on three observers and found no significant difference between conditions for two of the three. This provided our primary motivation to use the shortened duration exclusively in the latter two experiments. Another aspect which should be considered is that the task employed here did not require the selective tracking of some objects and not others. This represents an important difference between our and standard MOT studies. While it can be said that our stimuli were effectively MOT stimuli in which all objects were targets, there is likely a significant difference between tracking all objects and selective tracking only a subset. Whatever attentional mechanisms perform the selection of certain objects over others were undoubtedly operating in a different mode, if at all. Finally, it could be argued that our task could be accomplished by reporting the direction of motion perceived at the location where the cue appears. If so, then observers would not need to track the objects during motion and our results may not apply directly to MOT and MTT paradigms that require active tracking of objects. The underlying assumption behind this view is that motion information is associated with spatial locations and remains so even after the offset of the motion stimulus. An alternative view is that motion information needs to be associated (bound) to specific objects. This process of binding can be viewed as the act of tracking different objects during motion. If this latter alternative is true then, given the similarities between our stimuli and those typically used in MOT research, it is reasonable to conclude that the capability for the brief retention of large amounts of direction-of-motion information, as observed here, is also present in a typical tracking task. The implication for models of MOT is similar to the implication of Sperling's icon to models of visual information processing; instead of requiring a mechanism capable of processing the stimulus immediately, the theoretician is free to propose mechanisms which, in the absence of the stimulus itself, still act upon stimulus information as long as the trace of that stimulus is held in sensory memory. It has already been suggested by Narasimhan et al. (2009) that traces of object motion in sensory memory may be an important factor in the task of detecting deviations in linear trajectories, but it is also interesting to consider the possible role of such memory in other tracking tasks. For example, the finding of Keane and Pylyshyn (2006) that moving objects that disappear briefly and reappear at their extrapolated positions are tracked less well than if the objects reappear at their last seen locations can be explained readily by the gap in the

sensory trace in the former condition. There is a standing debate whether dynamic multiple-object processing is parallel or serial (for discussions from both sides, see Cavanagh & Alvarez, 2005; Narasimhan et al., 2009; Oksama & Hyona, 2008; Tripathy et al., in press). Because a serial processing model necessarily relies heavily on memory, the finding of an iconic-like memory for multiple-object motion may provide new material for future mechanistic explanations.

In this study, the observed effect of set size (Experiment 1) is gradual and consistent across the entire range tested. There is no evidence for the special significance of any one set size (e.g., a "capacity" of four or five), and the deterioration in performance is clear even in the three-object condition. Insofar as our results apply to tracking tasks, this result can be taken as evidence against the so-called "fixed architecture" theory of MOT that originated with Pylyshyn and Storm (1988) and agrees with the findings of Howard and Holcombe (2008) and Alvarez and Franconeri (2007), who also show graded decreases in tracking performance with set size. An alternative theory is that tracking has a flexible capacity, in which the exact number of objects that the observer can handle simultaneously depends on the complexity of the task or, equivalently, the amount of processing that must be allocated to each object (Alvarez & Franconeri, 2007; Tripathy et al., 2007), or on the stimulus uncertainty that increases with the number of items (Ma & Huang, 2009). The nature of such a flexible capacity could be determined either by constraints of attentional distribution or timing or by the available memory capacity at various levels of processing. The present findings offer some new evidence on the matter, as we measured the same graded set-size effect even when motion lasted only a fraction of a second. Because it is assumed that the endogenous direction of attention is too slow to have a significant effect on this time scale and, as expounded above, it is thought that sensory memory plays a large role in determining performance on this task; this result provides evidence that memory restrictions influence the "flexible" capacity proposed in both of the above studies.

Appendix A

Statistical characterization of performance

Means of errors

In addition to mean performance, it is possible to view the reporting error as a random variable and analyze its distribution (e.g., Ma & Huang, 2009; Zhang & Luck, 2008). One hundred trials were available for each set size, for each observer, and each of these sets of 100 was treated separately as a random sample from an unknown

distribution. The first step in this line of analysis was to evaluate the simple statistics of each sample. Figure A1 shows the mean *signed* error angle for each condition, along with error bars representing the standard error of the mean. This analysis is useful because it reveals that observer responses are not always centered around zero error, as would be expected; instead, certain conditions show a bias toward clockwise errors and some a bias toward counterclockwise errors. Conditions in which these deviations from zero are statistically significant[2] are marked by an asterisk in Figure A1. The reason for this bias is unknown, but subsequent analyses attempt to allow for bias by never assuming a zero-mean distribution.

Distributions of errors

The next step in analysis was to calculate an empirical distribution of errors for each condition. Histograms were calculated using 18 bins, each 20 deg in width, and then converted into empirical probability density functions by scaling to unit area. Figure A2 shows these distributions for the two extreme set sizes, one and nine, for one observer. Visual inspection suggests that the shape of the distribution changes with set size, leading to a smaller number of near-correct responses and an increased number of large errors in the larger set-size condition. Another observation is that the change in distribution shape does not resemble the widening of a Gaussian distribution, as the width of the main lobe appears relatively unchanged while the height of the "tails" of the curve increases substantially. This observation was analyzed more formally using the procedure described below. Although only the extreme values are shown, the apparent change in distribution shape with increased set size was observed as a general trend.

Error analysis using circular statistics

Because the random variable under consideration is circular (-180 deg and 180 deg represent the same angle), an accurate analysis of the distribution should use a PDF defined on the circle. The distribution considered here is the wrapped Gaussian; this distribution accounts for circularity and is equivalent to an infinite summation of normal distributions, corresponding to a single Gaussian "wrapping" around the circle, overlapping itself an infinite number of times. A nonlinear optimization routine was used to find the best-fitting wrapped Gaussian for each distribution. In practice, a finite number of overlaps must be assumed, but this proved inconsequential, as increasing the number of Gaussians used in the model had negligible effect. Figure A3 shows the same data as Figure A2, along with the best-fitting wrapped Gaussian for each set size. Only three Gaussians were included in the summation, but due to the small variance of the model distributions, the
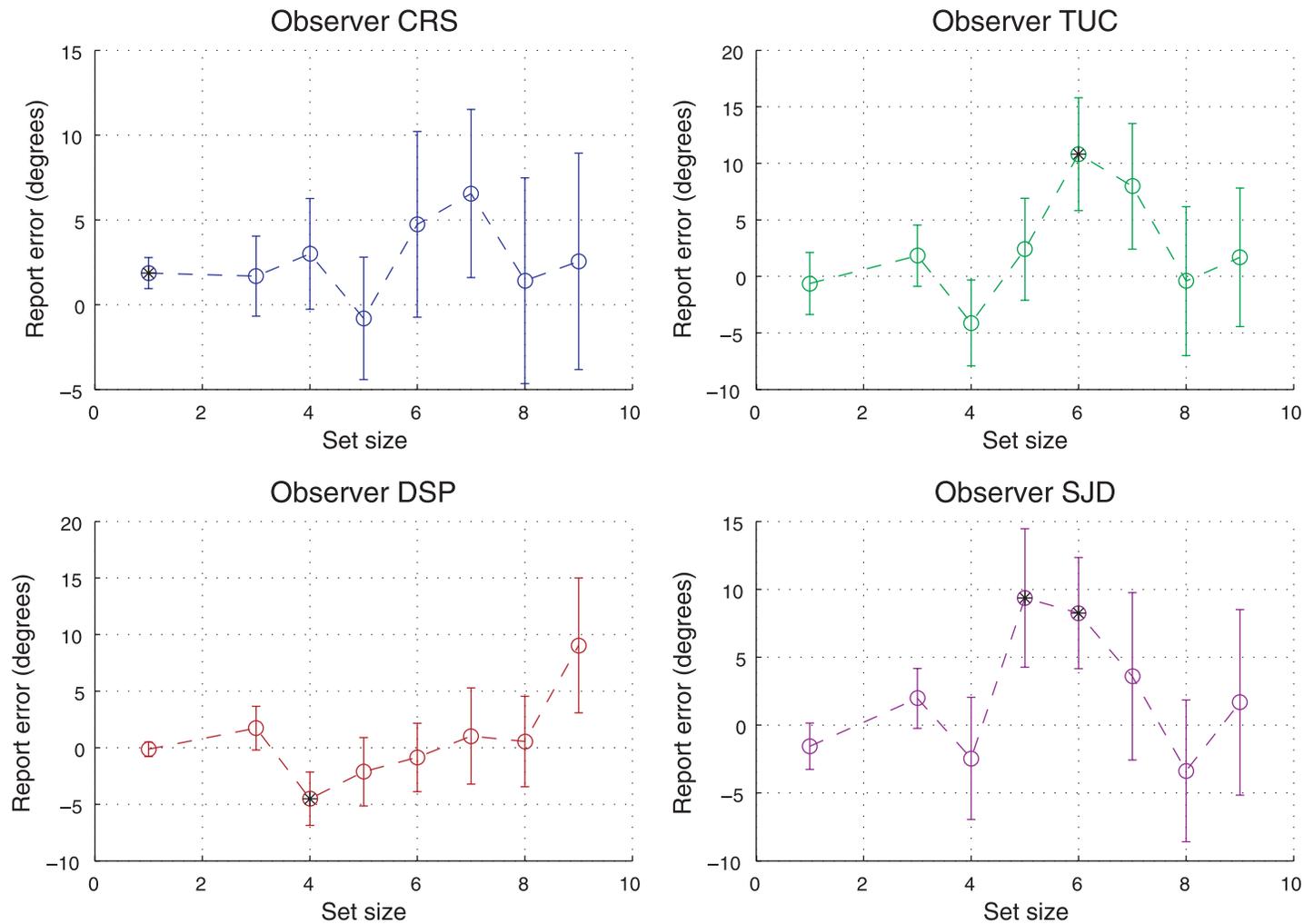
Figure A1. Mean report error for each set size. Positive values represent clockwise error angles. Asterisks mark means that are significantly different from zero ($p < 0.05$).

contribution from the "wrap around" effect was negligible. For this reason, increasing or decreasing the number of Gaussians used in the model did not affect the result. We therefore conclude that the distributions could equivalently be modeled as single Gaussians, and the circular nature of the data was ignored.

Normality of distribution of errors

Reverting to a simple Gaussian model allowed a more formal evaluation of normality by use of the Anderson–Darling test. This statistical test evaluates the null hypothesis that the distribution is Gaussian and results in a statistic that can be interpreted in terms of standard Gaussian rejection regions. Specifically, a test statistic greater than 1.65 allows rejection of the null hypothesis at a confidence level of 0.05. Each distribution was tested separately, and the resulting test statistics are plotted in Figure A4, where the dashed line represents the 0.05

rejection threshold just mentioned. These results again suggest that, generally, errors are normally distributed only at small set sizes. At larger set sizes, the shape of the distribution changes and can no longer be modeled with a Gaussian PDF. In general agreement with the analysis shown in Figure A4, statistical resampling indicated that each of the distributions except one (subject CRS for a set size of one) is significantly leptokurtic, indicating greater height in the tails than a normal distribution.[3]

Combined model for distribution of errors

An alternative model results from considering the performance expected from a limited capacity hypothetical observer (LCHO). In its strictest formulation, this model predicts that an observer reports perfectly on tracked objects and randomly on non-tracked objects. If this expectation is relaxed somewhat to account for the variability observed when tracking a single object, the
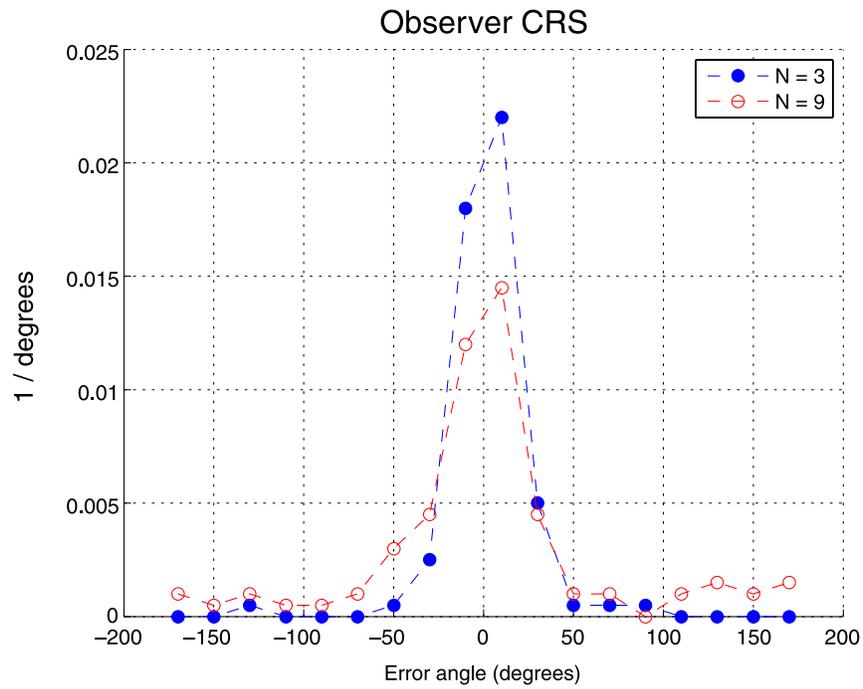
Figure A2. Empirical probability density functions computed from the distributions of one observer's reporting errors for two set sizes.
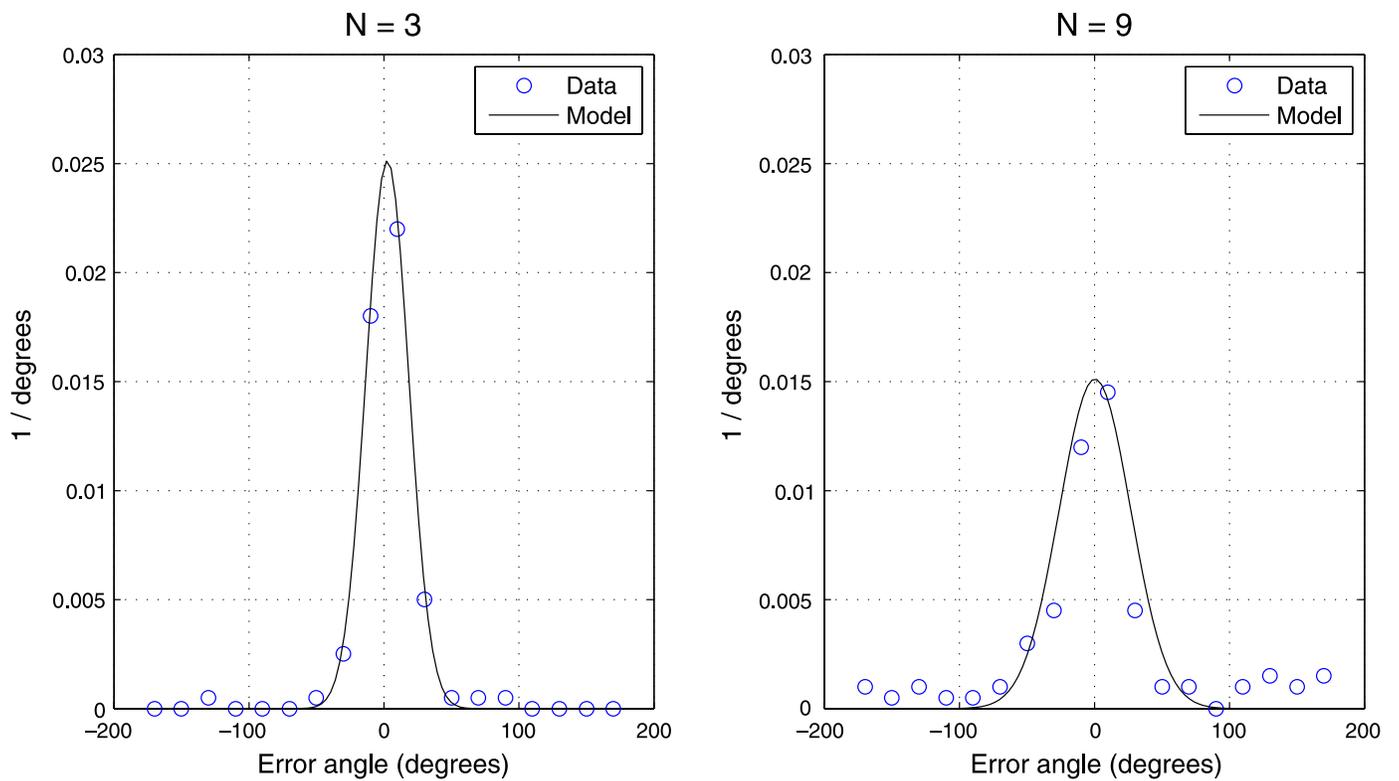

Figure A3. Wrapped Gaussian models of probability distributions shown in Figure A2. At the larger set size, the Gaussian model underestimates the "tails" of the empirical distribution.
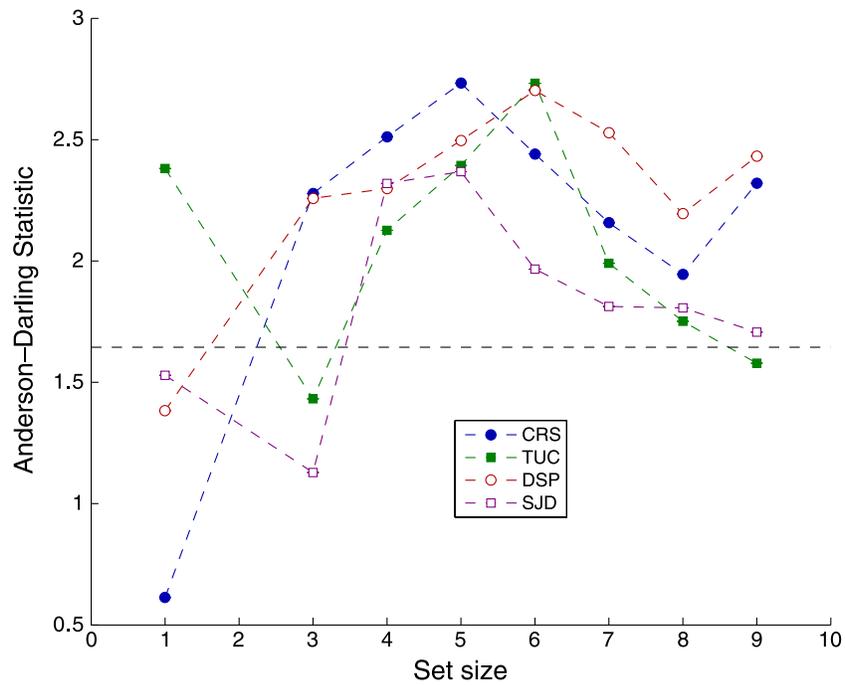
Figure A4. Anderson–Darling test statistics used to evaluate deviation from normality. Values above the threshold (dashed line) allow rejection of the hypothesis of normality at a 0.05 confidence level.

model can be restated in terms of response distributions of the type analyzed above. Specifically, the prediction is that errors in reporting tracked objects are normally distributed around zero, and errors in reporting non-tracked objects are uniformly distributed across all 360 deg. The measured distribution of errors, then, would be a weighted combination of these two distributions, where the weight is exactly the probability that a randomly selected object is one that is tracked. As mentioned above, it was desirable to allow for a non-zero distribution mean, and the model
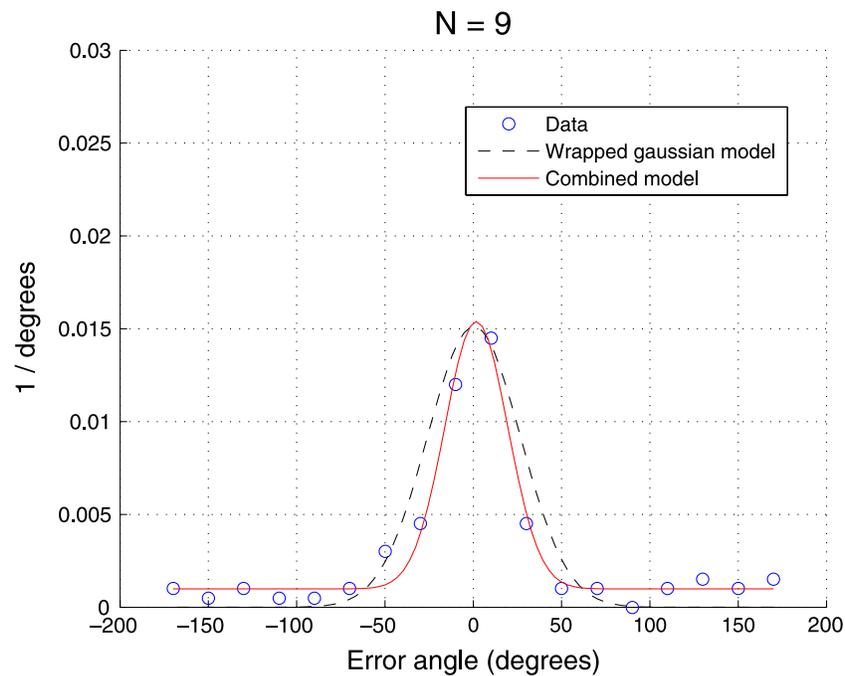

Figure A5. Comparison of two distribution models along with data for one observer for one set size. The combined model has both a Gaussian and uniform component and better accounts for large "tails" seen in the empirical PDF.

was adjusted accordingly. The functional form of this prediction is

$$p_\theta(\theta) = (w)N(\mu, \sigma) + (1 - w)U(-180, 180), \quad (A1)$$

where $N(\mu, \sigma)$ represents a normal distribution with mean μ and standard deviation σ, $U(-180, 180)$ represents a uniform distribution on the interval -180 to 180 deg, and $p_\theta(\theta)$ is the resulting probability density function. As a model, this function has three free parameters: μ and σ, the mean and standard deviation of the normal component, and w, the weighting factor combining the two components.

This combined model was also optimally fit to the behavioral data. The resulting model output appeared to better capture the change in distribution shape with set size. Qualitatively, it can be seen from visual inspection that the large "tails" of the empirical PDF are better accounted for by the combined model than by the Gaussian model considered previously. To allow comparison, the outputs of both models are plotted in Figure A5 for set size of 9, along with the example data previously shown in Figures A2 and A3.

For the purpose of quantitatively comparing these two models, we employed the following methods. First, model residuals were calculated for each distribution, and then averaged over all set size conditions. The result, shown in Figure A6, is an average residual curve for each observer,

showing the nature of the errors between the models and the empirical PDFs. Concerning the wrapped Gaussian model, the consistently negative residuals at extreme values confirm our earlier observation that this model underestimates the tails of the empirical distributions. Another observation is the sharp peaks at angles slightly to either side of the center of the distribution. This corresponds to an overestimation by the model of the width of the distribution. This can also be seen qualitatively in Figure A5, above. Overall, residuals for the combined model are of smaller magnitude and exhibit no clear trend across the range of angles.

As an additional metric of model performance, the coefficient of determination (R^2) was also calculated separately for each distribution. Figure A7 shows this measurement for each observer and each set size. As already noted, the correlation of the wrapped Gaussian model with the data deteriorates with increasing set size, and the combined model appears to produce a better fit overall. A new observation, however, is that the combined model also shows a deterioration with increased set size. This trend in the model performance implies that it has not fully captured the nature of the change in distribution shape with set size. As set-size effect was the subject of this experiment, this can be considered a weakness in the model, but the generally high performance of the model justifies its use in further analysis below.
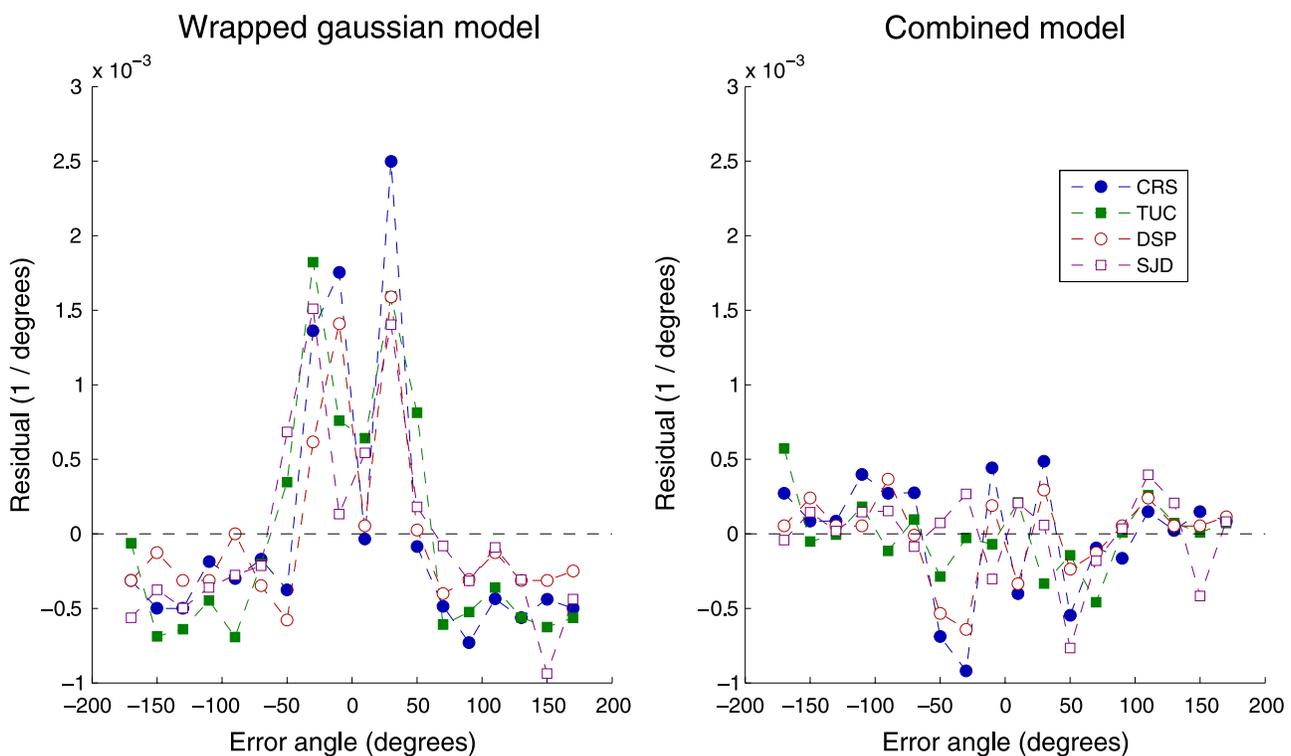

Figure A6. Model residuals, calculated as (model output) − (data), for both the wrapped Gaussian and combined distribution models. The wrapped Gaussian underestimates the height of the distribution tails while overestimating the width of the main lobe.
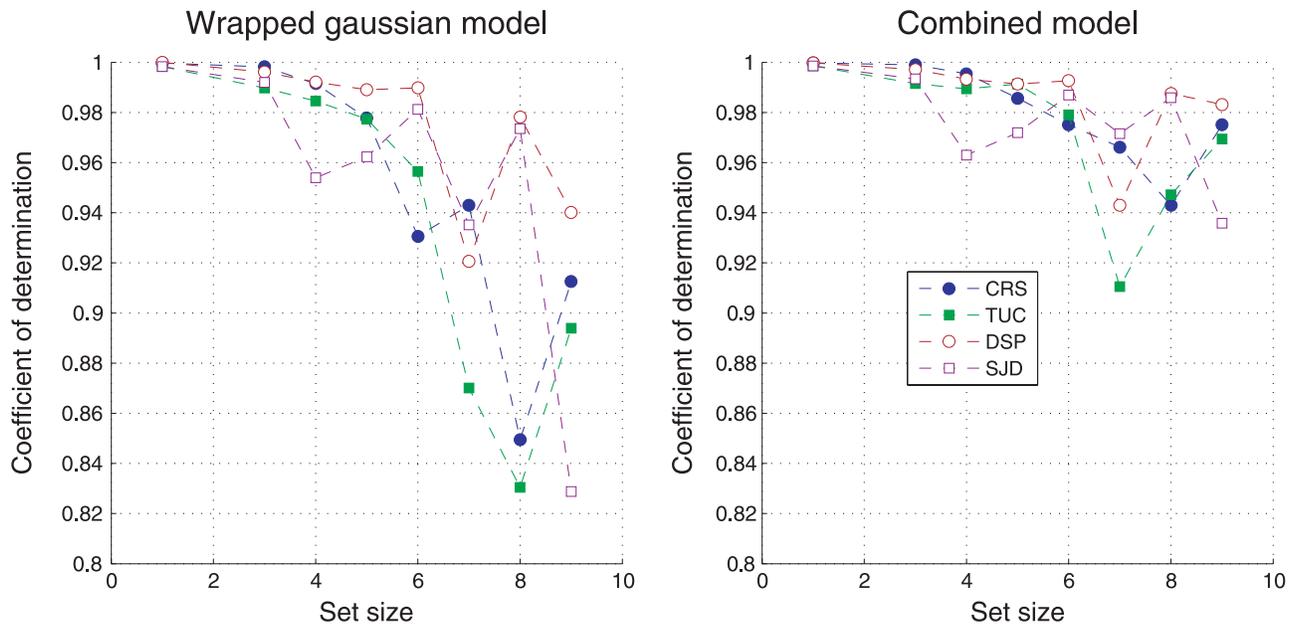
Figure A7. Coefficient of determination (R^2) for both the wrapped Gaussian and combined distribution models. Note that the combined model shows overall better performance but both exhibit a noticeable set-size effect.

Coefficients of the combined model

As a final treatment of these data, the coefficients of the combined model were analyzed with respect to set size. Because both the width of the Gaussian component and the weight combining the two components were left as free parameters to be optimized algorithmically, the resulting best-fit parameters may provide some insight into the data. Figure A8 shows both these parameters as functions of set size for each observer.

These model parameters describe the change in shape of the model PDF with set size, but again, the non-uniform
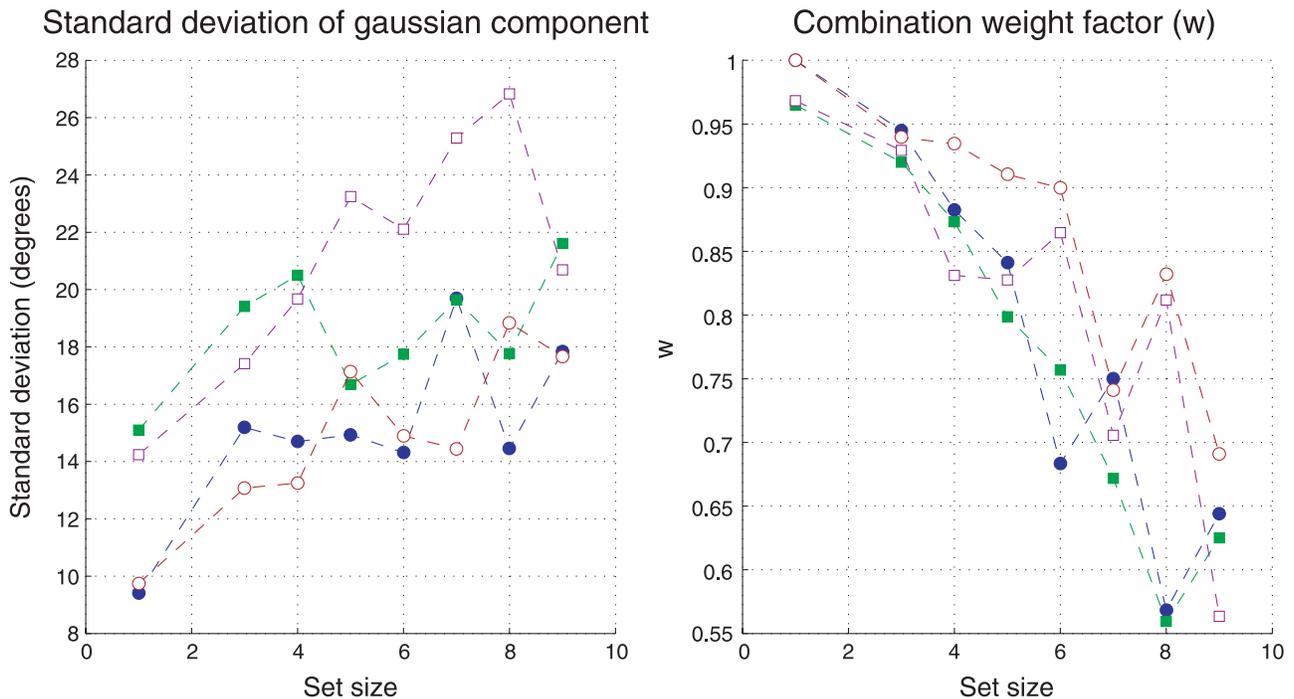

Figure A8. Parameters of the combined distribution model were optimized separately for each set size. The best-fitting parameters are shown for each observer.

performance of the model, as mentioned above, should be considered when making any inferences regarding the behavior of the actual underlying distribution. Because the limited-capacity hypothetical observer concept provided the motivation for the combined-distribution model structure, it is worth noting that the LCHO model, as described previously, provides a specific prediction for how these model parameters should relate to set size. Specifically, the model assumes a constant quality of information for tracked objects, regardless of set size, and would therefore predict a constant shape for the normal component of the model (in disagreement with the increase in the variance parameter, shown in Figure A8). Regarding the weighting factor, which represents the probability that a randomly chosen object was tracked, the LCHO model predicts the relationship $w = \min[1, C/N]$, where C is the fixed capacity of the observer. Although noisy, the trend seen in the empirically derived weighting factor (Figure A8) does not appear to follow this $1/N$ shape and would likely be better fit with a simple linear model.

Acknowledgments

Commercial relationships: none.
Corresponding authors: Haluk Öğmen.
Emails: ogmen@uh.edu.
Address: N308 Eng. Bldg. 1, University of Houston, Houston, Texas, USA.

Footnotes

[1]For similar methods, see Blake et al. (1997); Place and Horowitz (2006).

[2]Note that this statistical test did not adjust for multiple comparisons. This is because each condition was analyzed independently, and a general test for report bias was not performed. It must be pointed out that in making a total of 32 comparisons, we have increased the total probability of a type I error. The individual biases reported herein should not be used to draw any general inferences.

[3]Similar to the analysis of the mean signed errors that was described above, the definition of 95% confidence intervals for sample kurtosis included no adjustment for the fact that we tested multiple distributions.

References

Alvarez, G., & 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):4, 1–10, http://www.journalofvision.org/content/7/13/14, doi:10.1167/7.13.4. [PubMed] [Article]

Blake, R., Cepeda, N. J., & Hiris, E. (1997). Memory for visual motion. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 353–369. [PubMed]

Boi, M., Öğmen, H., Krummenacher, J., Otto, T. U., & Herzog, M. H. (2009). A (fascinating) litmus test for human retino- vs. non-retinotopic processing. *Journal of Vision, 9*(13):5, 1–11, http://www.journalofvision.org/content/9/13/5, doi:10.1167/9.13.5. [PubMed] [Article]

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

Demkiw, P., & Michaels, C. F. (1976). Motion information in iconic memory. *Acta Psychologica, 40,* 257–264. [PubMed]

Fencsik, D., Klieger, S., & Horowitz, T. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception & Psychophysics, 69,* 567–577. [PubMed] [Article]

Geisler, W. (1999). Motion streaks provide a spatial code for motion direction. *Nature, 400,* 65–69. [PubMed]

Haber, R. (1983). The impending demise of the icon: A critique of the concept of iconic storage in visual information processing. *Behavioral and Brain Sciences, 6,* 1–54.

Howard, C., & Holcombe, A. (2008). Tracking the changing features of multiple objects: Progressively poorer perceptual precision and progressively greater perceptual lag. *Vision Research, 48,* 1164–1180. [PubMed] [Article]

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

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision, 9*(4):1, 1–12, http://www.journalofvision.org/content/9/4/1, doi:10.1167/9.4.1. [PubMed] [Article]

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219.

Keane, B., & Pylyshyn, Z. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cognitive Psychology, 52,* 346–368. [PubMed] [Article]

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):3, 1–30, http://www.journalofvision.org/content/9/11/3, doi:10.1167/9.11.3. [PubMed] [Article]

Magnussen, S., & Greenlee, M.W. (1999). The psychophysics of perceptual memory. *Psychology Research, 62,* 81–92. [PubMed]

McKee, S. P., & Welch, L. (1985). Sequential recruitment in the discrimination of velocity. *Journal of the Optical Society of America A, Optics and Image Science, 2,* 243–251. [PubMed]

Narasimhan, S., Tripathy, S., & Barrett, B. (2009). Loss of positional information when tracking multiple moving dots: The role of visual memory. *Vision Research, 49,* 10–27. [PubMed] [Article]

Öğmen, H., & Herzog, M. H. (2010). The geometry of visual perception: Retinotopic and non-retinotopic representations in the human visual system. *Proceedings of the IEEE, 98,* 479–492.

Öğmen, H., Otto, T., & Herzog, M. H. (2006). Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Research, 46,* 3234–3242. [PubMed] [Article]

Oksama, L., & Hyona, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology, 56,* 237–283. [PubMed] [Article]

Otto, T. U., Öğmen, H., & Herzog, M. H. (2008). Assessing the microstructure of motion correspondences with non-retinotopic feature attribution. *Journal of Vision, 8*(7):16, 1–15, http://www.journalofvision.org/content/8/7/16, doi:10.1167/8.7.16. [PubMed] [Article]

Pasternak, T., & Zaksas, D. (2003). Stimulus specificity and temporal dynamics of working memory for visual motion. *Journal of Neurophysiology, 90,* 2757–2762. [PubMed] [Article]

Petersik, J. T., & Rice, C. M. (2006). The evolution of explanations of a perceptual phenomenon: A case history using the Ternus effect. *Perception, 35,* 807–821. [PubMed]

Place, S. S., & Horowitz, T. S. (2006). Which way did it go? Measuring trajectory information in multiple object tracking [Abstract]. *Journal of Vision, 6*(6):767, 767a, http://www.journalofvision.org/content/6/6/767, doi:10.1167/6.6.767.

Pylyshyn, Z., & Storm, R. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197. [PubMed]

Scholl, B., Pylyshyn, Z., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition, 80,* 159–177. [PubMed] [Article]

Shioiri, S., & Cavanagh, P. (1992). Visual persistence of figures defined by relative motion. *Vision Research, 32,* 943–951. [PubMed]

Simpson, W. A. (1994). Temporal summation of visual motion. *Vision Research, 34,* 2547–2559. [PubMed]

Snowden, R. J., & Braddick, O. J. (1991). The temporal integration and resolution of velocity signals. *Vision Research, 31,* 907–914. [PubMed]

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74,* 1–29.

Supèr, H., Spekreijse, H., & Lamme, V. A. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science, 293,* 120–124. [PubMed]

Ternus, J. (1926). Experimentelle untersuchungen über phänomenale Identität (Experimental investigations of phenomenal identity). *Psychologische Forschung, 7,* 81–136.

Ternus, J. (1938). The problem of phenomenal identity. In W. D. Ellis (Ed.), *A source book of Gestalt psychology.* London: Routledge and Kegan Paul.

Treisman, A., Russell, R., & Green, J. (1975). Brief visual storage of shape and movement. In *Attention and performance* (pp. 699–721). London: Academic Press.

Tripathy, S., & Barrett, B. (2004). Severe loss of positional information when detecting deviations in multiple trajectories. *Journal of Vision, 4*(12):4, 1020–1043, http://www.journalofvision.org/content/4/12/4, doi:10.1167/4.12.4. [PubMed] [Article]

Tripathy, S., Narasimhan, S., & Barrett, B. (2007). On the effective number of tracked trajectories in normal human vision. *Journal of Vision, 8*(4):8, 1–18, http://www.journalofvision.org/content/8/4/8, doi:10.1167/8.4.8. [PubMed] [Article]

Tripathy, S. P., Ogmen, H., & Narasimhan, S. (in press). Multiple-object tracking: A serial attentional process? In C. Mole, D. Smithies, & W. Wu (Eds.), *Attention: Philosophical and psychological essays.* Oxford University Press.

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