Why do people appear not to extrapolate trajectories during multiple object tracking? A computational investigation

Sheng-hua Zhong
Department of Computing, Hong Kong Polytechnic University, Hong Kong
Department of Psychological and Brain Sciences, The Johns Hopkins University, Baltimore, MD, USA

Zheng Ma
Department of Psychological and Brain Sciences, The Johns Hopkins University, Baltimore, MD, USA

Colin Wilson
Department of Cognitive Science, The Johns Hopkins University, Baltimore, MD, USA

Yan Liu
Department of Computing, Hong Kong Polytechnic University, Hong Kong

Jonathan I. Flombaum
Department of Psychological and Brain Sciences, The Johns Hopkins University, Baltimore, MD, USA

Intuitively, extrapolating object trajectories should make visual tracking more accurate. This has proven to be true in many contexts that involve tracking a single item. But surprisingly, when tracking multiple identical items in what is known as "multiple object tracking," observers often appear to ignore direction of motion, relying instead on basic spatial memory. We investigated potential reasons for this behavior through probabilistic models that were endowed with perceptual limitations in the range of typical human observers, including noisy spatial perception. When we compared a model that weights its extrapolations relative to other sources of information about object position, and one that does not extrapolate at all, we found no reliable difference in performance, belying the intuition that extrapolation always benefits tracking. In follow-up experiments we found this to be true for a variety of models that weight observations and predictions in different ways; in some cases we even observed worse performance for models that use extrapolations compared to a model that does not at all. Ultimately, the best performing models either did not extrapolate, or extrapolated very conservatively, relying heavily on observations. These results illustrate the difficulty and attendant hazards of using noisy inputs to extrapolate the trajectories of multiple objects simultaneously in situations with targets and featurally confusable nontargets.

Introduction

Multiple object tracking (MOT; Pylyshyn & Storm, 1988) is among the most popular and productive paradigms for investigating the underlying nature of visual cognition. In a typical experiment, a set of featurally identical objects moves about a display independently, and the task is to track a subset of the objects that were initially identified as targets (Figure 1). This task demands sustained effort; it cannot be accomplished via eye movements that shadow the motion of all targets; and basic display factors such as speed, duration, and the numbers of targets and nontargets afford direct and intuitive manipulations of task difficulty. The MOT paradigm has proven remarkably useful for identifying general properties of visual processing, such as the utility of inhibition alongside selective attention (Pylyshyn, 2006), the reference frames over which visual cognition operates (Liu et al., 2005), and the underlying units of selective attention (Scholl, Pylyshyn, & Feldman, 2001).

Recent advances have added a further dimension to the study of MOT by characterizing the computational problems at the core of the task. Specifically, visual tracking of multiple objects can be formalized as a

Citation: Zhong, S.-h., Ma, Z., Wilson, C., Liu, Y., & Flombaum, J. I. (2014). Why do people appear not to extrapolate trajectories during multiple object tracking? A computational investigation. *Journal of Vision*, *14*(12):12, 1–30, http://www.journalofvision.org/content/14/12/12, doi:10.1167/14.12.12.

doi: 10.1167/14.12.12 Received July 17, 2013; published October 13, 2014 ISSN 1534-7362

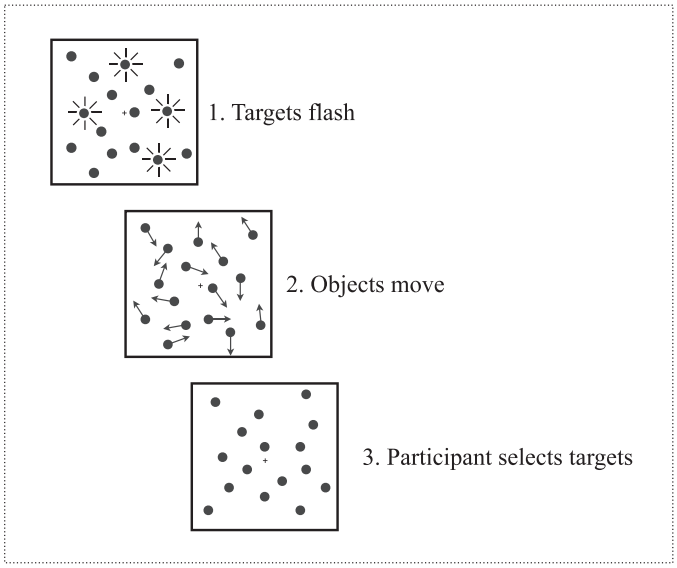

Figure 1. Sequence of events in a typical multiple object tracking (MOT) experiment. (1) A group of featurally identical objects appear, and a subset flash, identifying them as targets for tracking. (2) All objects move independently through the display for a duration typically lasting between 6 and 20 s. (3) The objects stop moving, and the participant must identify the targets (with a mouse or via keypress).

correspondence problem that requires identifying noisy measurements of objects in the current moment with noisy representations inferred a moment ago, a problem frequently described in the context of motion perception more generally (e.g., Dawson, 1991). Tracking errors can arise from incorrect correspondence inferences, in particular when a nontarget is mistakenly identified with a target (Bae & Flombaum, 2012; Franconeri, Intriligator & Cavanagh, 2001; Franconeri, Jonathan, & Scimeca, 2010). Two reports have explored computational models that adopt a probabilistic approach to representations of object position and correspondence (Ma & Huang, 2009; Vul, Frank, Alvarez, & Tenenbaum, 2009). These models employ Bayesian inference to infer moment-by-moment positions from noisy measurements, and to infer object correspondences between current positions and those in memory. This inchoate line of work provides a framework for understanding the representations and computations involved in tracking at the algorithmic level.

In the current study, we push this line of research forward by utilizing a similar framework, one that is common in computational applications for visual tracking, the Kalman filter. We use the model to investigate a basic but perplexing aspect of human performance. Specifically, we investigate the role and algorithmic mechanisms of velocity extrapolation to support tracking. As we discuss below, extrapolation has been explored behaviorally for some time in the MOT literature, but with surprising and conflicting results. Computational models can be useful here because they force one to specify the algorithms that would support extrapolation, and they supply an opportunity to evaluate their accuracy and utility with respect to the task. Here, they will allow us to ask varieties of questions that can be difficult to address exclusively through behavioral experiments, in particular, whether observers should extrapolate—would it actually help tracking performance?—and if they do, how apparent would it be?

Do observers extrapolate when doing MOT?

"Do observers extrapolate?" has been the question addressed by several groups of investigators. This work has been driven by the intuition that extrapolation should assist in the process of moment-by-moment correspondence inference. Put simply, if an observer can project with reasonable certainty where each object is headed, she should be able to use those projections in the interpretation of noisy measurements of future positions. But surprisingly, a fair amount of evidence has suggested that observers may not extrapolate (Fencsik, Klieger, & Horowitz, 2007; Franconeri, Pylyshyn, & Scholl, 2012; Howard, Masom, & Holcombe, 2011; Keane & Pylyshyn, 2006).

The first study to directly examine extrapolation—and to obtain negative evidence—was conducted by Keane and Pylyshyn (2006). They introduced a global interruption in the middle of typical MOT trials. After a few seconds of tracking, the whole display became blank for 307 ms. When the stimuli reappeared, participants were asked to identify the targets. To motivate extrapolation, the authors used trajectories with nearly perfect inertia, trajectories in which objects almost always maintain their speed and direction (in a way that is not typical of the broader MOT literature, as discussed further below). Nonetheless, observers were better able to identify targets when they remained stationary during the global interruption compared to when the objects reappeared at positions that were consistent with their trajectories prior to the interruption. Fencsik et al. (2007) replicated this finding, strengthened it with novel controls, and concluded that if participants do successfully use any motion-related information, they only do so when tracking one or two objects.

Similarly, Franconeri et al. (2012) investigated extrapolation through occlusion, as opposed to global interruption. It has been demonstrated in a number of studies that observers can track objects through moments of occlusion with seemingly no cost to performance (Flombaum, Scholl, & Pylyshyn, 2008; Scholl & Pylyshyn, 1999). This study explored how

participants identify targets on one side of an occluder with targets on the other side. The authors compared performance for objects that appeared at expected disocclusion locations with objects that emerged following hidden trajectory perturbations. Across several experiments, a clear but counterintuitive pattern emerged. Observers identified disoccluding objects with occluded ones based only on spatial proximity, as though they expected objects to appear as close as possible to where they disappeared regardless of their specific motion trajectories prior to occlusion.

In addition to investigations utilizing interruptions, a few studies have shown that performance in standard MOT trials is relatively insensitive to differences between predictable and unpredictable trajectories. Vul et al. (2009) demonstrated that performance is not affected by changes in object inertia—that is, how likely an object is to maintain its current bearing. Similarly, Howe and Holcombe (2012) discovered that participants do not perform better when tracking four objects that move along straight paths, as opposed to objects that alter their direction every so often; though they did find a performance difference in trials with only two targets. In that study, objects always maintained the same speed, a feature that was not present in Vul et al.'s experiment (2009). Thus experiments manipulating object inertia suggest that observers only utilize extrapolation successfully when tracking fewer than four objects, and only when object speeds and bearings do not change. Finally, Howard et al. (2011) supplied evidence to suggest that representations of location during MOT tend to lag behind objects' true positions, perhaps indicating a conservative extrapolation strategy.

The picture that emerges from these results is that under many circumstances, especially those including more than two targets, human observers either do not extrapolate or do not benefit from their forward predictions of object positions. But alongside these findings, related experiments have shown that observers could extrapolate, and perhaps do so in some contexts. In particular, several studies have shown that human observers possess the raw materials necessary for extrapolation. They can report (albeit with noise) the direction of a target's motion in an MOT trial (Horowitz & Cohen, 2010; Iordanescu, Graboweky, & Suzuki, 2009; Shooner, Tripathy, Bedell, & Öğmen, 2010). One study even demonstrated that a moving pattern within an object can impair tracking if it moves against the object's trajectory, suggesting an automatic encoding of bearing (St. Clair, Huff, & Seiffert, 2010) and a cost caused by signals that make this encoding difficult.

Moreover, two studies supply evidence that observers extrapolate when tracking only one or two objects in the MOT paradigm. Using their global interruption paradigm, Fencsik et al. (2007) found improved target identification when objects underwent hidden motion during an interruption that was preceded by motion, compared to a case in which static objects preceded the interruption. They could and did utilize trajectory information in a comparative sense in this particular circumstance. But even this effect only obtained with a tracking load of two or fewer. Similarly, Howe and Holcombe (2012) found improved tracking performance for two targets in trials with reliable speeds and bearings, compared to trials where those parameters could randomly change. But like Fencsik et al., they observed no such advantage for trials with more than two targets.

In addition, a broader literature has demonstrated conclusively that observers effectively utilize trajectory predictions when tracking moving targets outside the context of the MOT paradigm, for example when tracking a ball in a variety of sports (e.g., Bennett, Baures, Hecht, & Benguigui, 2010; Diaz, Cooper, Rothkopf, & Hayhoe, 2013; Spering, Schütz, Braun, & Gegenfurtner, 2011). An important caveat is that this literature almost always involves tracking a single object in situations without featurally identical nontargets. Nonetheless, it demonstrates that the relevant trajectory information can be acquired by the human visual system, and that it can be used to generate effective predictions.

Taken together then, the existing research on the subject of extrapolation in object tracking and MOT is puzzling. Why do observers who can encode velocity, can be biased by it, and can even extrapolate when tracking only one or two objects, not seem to extrapolate in a setting that involves the tracking of larger numbers of targets among featurally identical nontargets? Our goal is to address this question by interrogating the underlying assumptions that have motivated previous studies. In particular, we focus on the assumption that extrapolating would benefit performance in MOT.

To make the assumption more concrete, consider the global interruption experiments. Objects in the relevant conditions always appeared exactly where they should have given their trajectories prior to interruption. But if participants were extrapolating, is it safe to assume that they would extrapolate to exactly the right positions? Perhaps observers did make predictions, but the predictions were relatively inaccurate, such that the post-interruption positions did not conform to observers' expectations in practice (i.e., did not conform any better than the positions at which the objects were most recently perceived). More importantly, in the context of MOT specifically, for predictions to have an impact on *task performance*, predictions would need to help observers discriminate targets from nontargets. Predictions would have to have discriminatory power

precise enough to have a marginal impact on the instances when targets and distractors tend to become confused.

Assuming noisy knowledge of bearing—especially noisy knowledge that appears to decline with tracking load (Horowitz & Cohen, 2010)—how accurate could human extrapolations have been? From a rational perspective, an observer should only use extrapolations in proportion to their precision and utility. The possibility that observers should extrapolate only under very limited circumstances, or extrapolate somewhat conservatively when tracking multiple objects, might help to account for the mixed results in the literature. Accordingly, we sought to investigate how accurately we should expect human observers to extrapolate—and whether it would benefit them—given noisy inputs.

The current study

Whether extrapolation benefits tracking performance clearly depends on the fidelity with which observers can predict object motion given the information that they receive from a display (together with their implicit assumptions about trajectories). In other words, it depends on whether observers really can extrapolate very well in the context of MOT. Given known limitations associated with human perception and memory, particularly in the spatiotemporal domain and with large tracking loads (Anstis, 1974; Gegenfurtner, Xing, Scott, & Hawken, 2003; He, Cavanagh, & Intriligator, 1996; Intriligator & Cavanagh, 2001; Latour, 1967; Lichtenstein, 1961; Sperling & Weichselgartner, 1995; White & Harter, 1969), highly accurate extrapolation may not be possible. Note, by highly accurate, here, we mean accurate enough to impact MOT performance—to aid in target non-target discrimination.

To investigate the effectiveness and utility of extrapolating, we compare tracking performance across models that are endowed with human-like perception and memory limits, but that differ with respect to the weight that extrapolations bear in the inferential processes involved in tracking. Specifically, there are known limits on how quickly human observers can sample visual inputs (Howard, Masom, & Holcombe, 2011; Landau & Fries, 2012; White & Harter, 1969), and it is also known that they represent object positions and velocities imprecisely (Adelson & Bergen, 1986; Bays & Husain, 2008; Bouma, 1970; Burr & Thompson, 2011; Gegenfurtner, Xing, Scott, & Hawken, 2003; Intriligator & Cavanagh, 2001; Stocker & Simoncelli, 2006). These noisy inputs could lead to relatively imprecise and possibly misleading extrapolations. Moreover, human observers may possess resource constraints in the form of limited memory or attention, potentially impairing them further with greater tracking loads (Bays, Catalao, & Husain, 2009; Mazyar, van den Berg, & Ma, 2012; Vul, Frank, Alvarez, & Tenenbaum, 2009). Such limits could make extrapolation computations less effective, and potentially even detrimental if they consume limited resources. In the case of a resource-limited observer, whether to extrapolate or not should depend on whether doing so produces a marginal performance advantage. Without a performance advantage, dispensing with extrapolation could produce resource savings that can be allocated to other processes.

The computational experiments described below utilize the Kalman filter as a tool for investigating the effectiveness and utility of extrapolation given noisy inputs. The Kalman filter is a basic framework for recursively estimating the values of unknown variables from noisy measurements. It is applied frequently in computer vision applications for tracking (BarShalom, & Fortmann, 1988; Boykov & Huttenlocher, 2000), and it supplies the main framework for two prior modeling efforts in MOT (Ma & Huang, 2009; Vul, Frank, Alvarez, & Tenenbaum, 2009). Using the Kalman filter as a computational foundation, we begin by implementing three models of multiple object tracking, one that does not utilize extrapolated predictions, and two models that weight extrapolations in different ways. We endow all of the models in the current study with perceptual limits in the range of human observers. Specifically, we test each model under three different levels of spatial uncertainty about object position, and with three visual sampling rates spanning the fast and slow ends of previously measured human capabilities.

One tangential contribution of the current research involves the implementation of this sampling limit. The two previous MOT modeling studies mentioned assumed that inputs to tracking are sampled at the framerate of the relevant experimental displays and underlying code. We return to this issue in the General discussion, noting here that temporal uncertainty in the form of a limited temporal sampling rate should hamstring extrapolation algorithms, just as spatial uncertainty should. The current work makes several additional contributions, pushing forward what appears to be a productive future for modeling human object tracking with Kalman filter-like algorithms. We discuss these contributions in detail in the General discussion.

In addition to comparing the performance of three different kinds of models, we test the models across a wide range of trajectory types and tracking loads commonly utilized in studies of human MOT ability. Importantly, we also test the models on trajectories in which objects tend to maintain their speeds and bearings—trajectories that are less common in the

experimental literature, but which should supply the greatest opportunity for extrapolation to confer an advantage. Studies interested in extrapolation have often utilized these kinds of trajectories; although studies focused on other issues typically utilize trajectories with frequent and unpredictable speed and bearing changes. We test with both kinds of trajectories because it is not safe to assume that a human observer in an MOT experiment would know whether the trajectories she faces are predictable or not. If an observer wanted to make predictions only when faced with very dependable trajectories, she would first need to infer that a given display includes those trajectories. Thus the consequences of extrapolating in displays with unpredictable motion are relevant to understanding how and when observers extrapolate in more reliable settings. Large costs to extrapolating at the wrong time may make it advantageous to be conservative all the time, that is, for an observer who does not know in advance whether it is the right or wrong time to extrapolate.

The organization of what remains is as follows. Model details are described in the General methods. We then present data from human participants who were tested on a subset of trials that the models were also tested on. This is to demonstrate that our models generally perform within the range of human observers. We then turn to two initial computational experiments. Computational Experiment 1 compares a Kalman filter with a model that does not utilize extrapolations at all. The Kalman filter makes adaptive extrapolations based on previous experience accumulated over the course of a trial. Computational Experiment 2 compares a model that utilizes extrapolated predictions in a relatively more rigid way with a model that does not extrapolate at all. After discussing the results of these two experiments, we report several replications of Experiment 1 using variations of the models, which were designed to examine slightly different approaches to making predictions. Finally, in Computational Experiment 3 we test the Kalman filter model and the nonextrapolating model on the signature interruption experiment used by Fencsik et al. (2007) and Keane and Pylyshyn (2006).

General methods

Computational framework for MOT

The computational experiments described below utilize three models of multiple object tracking. Each of these models is a formal proposal about how the task of multiple object tracking is performed. This task, and the associated components of the model observers, can be usefully decomposed as follows:

1. *Measurement*. At each moment in time, the observer receives noisy measurements of the positions of all objects in the display (targets and nontargets). The inclusion of nontarget measurements is critical, as the typical MOT task would be trivial if the observer knew which portions of the moment-by-moment stimulus were due to targets. All of the observations are corrupted by independent noise; as is standard in Bayesian approaches to perception, we assume that the observer is aware of her own noise variance (e.g., Girshick, Landy, & Simoncelli, 2011; Kersten, Mamassian, & Yuille, 2004; Lee & Mumford, 2003; Ma & Huang, 2009; Maloney, 2002; van den Berg, Shin, Chou, George, & Ma, 2012; Vul, Frank, Alvarez, & Tenenbaum, 2009; in the case of motion, specifically, see Sekuler, Watamaniuk, & Blake, 2002; Warren, Graf, Champion, & Maloney, 2012).
2. *Correspondence*. After receiving a set of noisy position measurements, the observer must assign measurements to objects that are being tracked. This is known as the correspondence or data association problem for tracking (Cox, 1993; Oh, Russell, & Sastry, 2004). Because there are no non-spatial features that distinguish the items in a typical MOT experiment, all of the models attempt to solve the correspondence problem with a simple positional heuristic: Measurements are assigned to objects in a way that minimizes the total (Euclidean) distance between the predicted and observed positions. (Some models and applications include tracking of nontargets, with labeling of targets as such. Our models track only targets, however, reflecting common assumptions in the literature concerning the selection of targets via attentional mechanisms; e.g., Drew, McCollough, Horowitz, & Vogel, 2009; Pylyshyn, 2006).
3. *Position inference*. Once each target has been assigned a measurement, its current position is inferred according to Bayes' theorem. This inference combines two sources of information, the current measurement and the predicted (Bayesian prior) position of the object, each with its own degree of uncertainty. Because predictions are derived from previous inferences and observations, they will always be noisier than new observations, inclining a rational observer to weight new observations more highly than prior predictions.
4. *Extrapolation (velocity inference)*. At this point, it is worth drawing a distinction between extrapolation and prediction (as we use them in this

paper). Specifically, we will use *extrapolation* to refer to an expectation about an object's future position based entirely on recent inferences of its velocity. In contrast, we use *prediction* to refer to any expectation about future position, however it is derived. The models that extrapolate do so, in our analysis, on the basis of combining consecutive observations of position with a prior distribution on velocity. In particular, if $z(t-1)$ is the position measurement assigned to an object at time $t-1$, and $z(t)$ is the position measurement assigned to the same object at the immediately subsequent time t, then the vector $z(t)-z(t-1)$ provides a measurement of the velocity (change in position per unit time) of the object. We assume that this velocity measurement is corrupted by noise that is independent of that for spatial position and due to the properties of velocity channels (Burr & Thompson, 2011; Stocker & Simoncelli, 2006). Bayes' theorem is used to infer the velocity at time t from this measurement and the velocity prior.

5. *Prediction of position and velocity*. The correspondence and inference steps above depend on prior predictions about position and velocity. These predictions are in turn based on measurements and inferences made earlier in the course of tracking. The three initial models evaluated here differ primarily in the way that predictions are computed (see further discussion below). Discussions of extrapolation in MOT have implicitly assumed that an appropriate prediction method would be to simply add inferred velocity vectors to previous positions. In addition to exploring such a model, we examine one that places predictions closer to previously perceived positions than would be expected under pure extrapolation.

Tracking with the Kalman filter

The Kalman filter is a Bayesian model that tracks stochastic linear dynamical systems observed through noisy sensors. It operates on a stream of noisy input data to produce a statistically optimal moment-by-moment estimate of the underlying system state (i.e., positions and velocities). The Kalman filter is a recursive estimator, which means that it makes successive predictions and then corrects these predictions in light of new observations. This amounts to a form of feedback control: The model predicts the system state at some time and then obtains feedback in the form of (noisy) measurements. Accordingly, equations for the Kalman filter can be classified as either prediction equations or measurement equations. Prediction equations use probabilistic beliefs about the current state and recent past to obtain prior estimates for the immediate future. Measurement equations are responsible for the feedback—for using new measurements to obtain posterior state estimates that may differ from the priors. In the description below we adopt standard notations so that readers can refer to widely available sources deriving and describing the Kalman filter more exhaustively (Kalman, 1960; Murphy, 2012; Welch & Bishop, 2006; Yilmaz, Javed, & Shah, 2006).

Measurement

Suppose some number of targets, N_T, and non-targets, N_D, adding up to a total number of objects, N_A. At a given moment in time t, z_t^m denotes the mth observation of position—including vertical and horizontal coordinates—from an object in the display, $m = 1,\ldots,N_A$. An observation is derived from the object's true position, denoted l_t^m, the position of object m at time t, as follows:

$$z_t^m = \mathbf{l}_t^m + r_t^m. \quad (1)$$

Here r_t^m is noise, assumed to be zero-mean Gaussian white noise with measurement noise covariance $R_t^m = \sigma_z^2 \mathbf{I}_2$. We tested models with three different values for σ_z^2, derived from relevant literature and intended to reflect a reasonable range of human spatial precision in location perception (Bays & Husain, 2008). (Here and throughout $\mathbf{I}_2$ stands for the identity matrix of size two.)

Position inference

Given an observation that has been assigned to a particular target m at time t, the model estimates a posterior for the target's current position, denoted $\widehat{\mathbf{l}}_t^m$. This estimate is obtained by the weighted combination of a prior position estimate assigned to time t, $\widetilde{l}_t^m$, and the observation:

$$\widehat{\mathbf{l}}_t^m = (\mathbf{I}_2 - K_t^m)\widetilde{l}_t^m + K_t^m z_t^k. \quad (2)$$

Note that the index for the observation, z_t^k, need not be the same as that of the object (i.e., the observer may have associated the wrong measurement with a target being tracked). K_t^m is the weight matrix, also called the Kalman gain, which determines the relative weight of the prior and the current observation in determining the posterior estimate. The value for K_t^m is selected to minimize the error covariance in the posterior, denoted $\widehat{P}_t^m$ (Jacobs, 1993). Similarly, $\widetilde{P}_t^m$ denotes the error covariance in the prior at time t. K_t^m and $\widehat{P}_t^m$ are thus obtained via the following pair of equations:

$$\mathbf{K}_t^m = \widetilde{\mathbf{P}}_t^m(\widetilde{\mathbf{P}}_t^m + \mathbf{R}_t^m)^{-1} \quad (3)$$

$$\widehat{\mathbf{P}}_t^m = (\mathbf{I}_2 - \mathbf{K}_t^m)\widetilde{\mathbf{P}}_t^m. \quad (4)$$

Prediction of position and velocity

To understand how the model obtains prior estimates, consider time $t + 1$. The expected position of the object should depend on basic motion kinematics, projecting forward from the posterior estimated at time t, $\widehat{\mathbf{l}}_t^m$, and utilizing a posterior estimate of velocity, *also* obtained at time t, denoted $\widehat{\mathbf{v}}_t^m$. Additionally, the model weights the prior to some degree towards its current posterior to reflect discrepancies between priors and observations in the past, and with the intention of reducing such discrepancies over time. We will call this the *adjusted prior*. Accordingly, an adjusted prior for time $t+1$, denoted as $\widetilde{\mathbf{l}}_{t+1}^m$, is obtained via the following equation:

$$\widetilde{\mathbf{l}}_{t+1}^m = (1 - \beta_t^m)(\widehat{\mathbf{l}}_t^m + \widehat{\mathbf{v}}_t^m) + \beta_t^m \widehat{\mathbf{l}}_t^m. \quad (5)$$

Here, β_t^m is the average of the main diagonal value in the Kalman gain matrix of target m, K_t^m. A high β_t^m value draws the adjusted prior closer to the most recent posterior so that it relies less on the velocity estimate. In contrast, a value of one would amount to not adjusting the prior at all. The value β encodes the relative weight of extrapolation assigned by the model. As will be clear in Computational Experiment 2, fixed β models do not perform as well as the model presented here.

When the model makes a prediction about an object's future position, it also projects forward an expected error covariance in the prior (utilized in Equations 3 and 4), which is denoted as $\widetilde{P}_{t+1}^m$, to be used at time $t + 1$. This estimate is derived from the difference between the prior and the posterior position estimates at the previous time point (Bishop, 2006):

$$\widetilde{P}_{t+1}^m = \widehat{\mathbf{P}}_t^m + \left[\widehat{\mathbf{l}}_t^m - \widetilde{\mathbf{l}}_t^m\right]\left[\widehat{\mathbf{l}}_t^m - \widetilde{\mathbf{l}}_t^m\right]^T. \quad (6)$$

Because this variance is estimated from quantities themselves derived from noisy observations, priors and adjusted priors will always be noisier than new observations, inclining an observer to value new observations more when inferring new posteriors. We discuss this point in the General discussion in the section Why doesn't extrapolation help?

Velocity inference

The estimate $\widehat{\mathbf{v}}_t^m$ is obtained in a way similar to the position estimate $\widehat{\mathbf{l}}_t^m$, that is, by Bayesian inference dependent on a prior expectation about velocity and an observation of velocity. Observations of velocity are derived by subtracting the two most recent position observations assigned to a target, and adding error to this value to reflect independent noise in human velocity perception channels[1]:

$$\mathbf{z}_t^m - \mathbf{z}_{t-1}^m + ov_t^m. \quad (7)$$

Here ov_t^m is drawn from a noise distribution for velocity, assumed to be zero-mean Gaussian white noise with covariance $\sigma_v^2\mathbf{I}_2$. The value of σ_v^2 was derived from relevant literature on the precision of velocity perception ($\sigma_v = 0.28°\text{s}^{-1}$; Gegenfurtner, Xing, Scott, & Hawken, 2003). With this observation value, the posterior for velocity is calculated as follows:

$$\widehat{\mathbf{v}}_t^m = (\mathbf{I}_2 - \mathbf{G}_t^m)\widetilde{\mathbf{v}}_t^m + \mathbf{G}_t^m(\mathbf{z}_t^m - \mathbf{z}_{t-1}^m + ov_t^m). \quad (8)$$

Here $\mathbf{G}_t^m$ is the Kalman gain for velocity, again derived by Bayes' theorem to minimize the error covariance in $\widehat{\mathbf{v}}_t^m$.

In Equation 7 the prior on velocity for target, m, $\widetilde{\mathbf{v}}_t^m$, is the difference between the posterior position estimates at moments t and t-1, $\widehat{\mathbf{l}}_t^m - \widehat{\mathbf{l}}_{t-1}^m$ with the addition of perceptual noise:

$$\widetilde{\mathbf{v}}_t^m = \widehat{\mathbf{l}}_t^m - \widehat{\mathbf{l}}_{t-1}^m + pv_t^m. \quad (9)$$

Again, pv_t^m is drawn from a distribution assumed to be zero-mean Gaussian white noise with covariance σ_v^2 I_2 intended to reflect error within velocity channels.

Correspondence

In typical computer vision applications, the correspondence problem—the problem of linking measurements with objects being tracked—is not solved on a purely spatiotemporal basis. This is because only one object is tracked, or knowing the identity of an object is not important for the task, or because differences among objects in surface appearance (such as color or shape) can be utilized. In the MOT paradigm, however, the identities of multiple objects are important (at least at the level of the target vs. nontarget distinction), and perceptual differences other than position are not available to inform correspondence inferences.

Our models address correspondences in the following way. We denote $p(T_t^m = k)$ as the probability that the kth observation at time t corresponds to target m. The model attempts to solve the correspondence by assuming that a new observation for target m will be drawn from a Gaussian distribution centered on the adjusted prior expectation about the position of m, $\widetilde{l}_t^m$. Thus:

$$p(T_t^m = k) = N(z_t^n; \widetilde{l}_t^m), 1 \le k \le N_A, 1 \le m \le N_T. \quad (10)$$

Assuming that the new observations are generated independently, and incorporating the principles of

mutual exclusivity and exhaustive association for targets, the optimal correspondence can be obtained by maximizing the probability in Equation 11.

$$\{k_m | 1 \leq m \leq N_T\} = \underset{1 \leq k_m \leq N_A}{argmax} \left[\prod_{1 \leq m \leq N_T} p(T_t^m = k_m) \right]. \quad (11)$$

This is equivalent to minimizing the sum or product of the Mahalanobis distances (equivalently, the Euclidean distances) of new observations and the expected positions of the targets they are assigned to. There are other heuristic approaches to the correspondence problem, based on nearest neighbor matching or specific validation regions (BarShalom et al., 2009; Murphy, 2012), that could be explored in future models.

To compare with the Kalman filter described above, we implemented two additional models, one that we will call the spatial working memory model, and one that we will call the 50/50 prediction model. The reasons for implementing each of these models will become clear during the discussion of each individual experiment.

Technically, however, each model variant can be derived by setting the Kalman gain of Equation 2 and the associated weighting term β in Equation 5 to fixed values. For the spatial working memory model, we set the Kalman gain matrix to **0** (hence β is fixed to 0), meaning that correspondence decisions at time $t + 1$ are based entirely on the posterior position estimates obtained at time t. Similarly, in the 50/50 prediction model, the Kalman gain was fixed at **0.5** (hence β is fixed to 0.5); in this model, correspondence decisions are based on adjusted priors that are always an equal mixture of pure extrapolations (unadjusted priors) and previous posterior positions. Additionally, for this model, the Kalman gain for velocity was permanently set to **0.5**, such that all posterior estimates of velocity were an average of the most recent prior and an observational value. (Note that we use bold type to indicate matrices, for example **0.5** is the 2×2 matrix with 0.5 on the diagonal).

Model parameters

Several parameters and variables play critical roles in the overall performance of these models. First, the variance of spatial measurements received by the model should clearly constrain performance and influence extrapolation. In previous MOT and spatial working memory research it has been proposed that human precision (inverse variance) declines as more items are stored or tracked (Bays & Husain, 2008; Vul, Frank, Alvarez, & Tenenbaum, 2009). This issue is outside the scope of the current project, because we sought to identify general features of extrapolation that apply across the range of human precision. Accordingly, we compared models with three different standard deviations intended to capture the range of reported psychophysical performance with small and large tracking/memory loads: 0.2°, 0.435°, and 0.6° (Bays & Husain, 2008). Similarly, observers should possess uncertainty about the time differences between measurements of objects, and they obviously do not engage the display frame-by-frame at the rate that the monitor refreshes. To capture human temporal limits, we tested models that sampled at three different rates thought to reflect the low to high end of human ability (5 Hz, 12 Hz, and 20 Hz, Howard, Masom, & Holcombe, 2011; Landau & Fries, 2012; Latour, 1967; Lichtenstein, 1961; VanRullen & Koch, 2003; White & Harter, 1969). We also used a value of $0.28°^{-1}$ (Gegenfurtner, Xing, Scott, & Hawken, 2003) as the standard deviation of velocity observations; this was previously shown to be an average human discrimination threshold for an object moving at a speed of $4°s^{-1}$, the average speed at which objects moved in these trajectories (and frequently in other MOT studies).

Trajectories

The models were tested with a variety of trajectory types typical of the wider literature on human MOT abilities. Each trial began with two to eight target disks (0.43° radius) and an equal number of nontargets. No object could begin a trial closer than 1.25° to any other object (measured center to center). Each object began moving $4°s^{-1}$ in a randomly determined direction. On each frame, each object had an independent probability of changing its speed by $\pm 0.13°s^{-1}$. An object's speed could never move below $1°s^{-1}$ or above $7°s^{-1}$, and as an object's speed approached these limits, changes were more likely to adjust toward the starting speed of $4°s^{-1}$. The probability of a speed change on each frame was fixed in a trial at either 0, 0.25, 0.5, 0.75, or 1, with equal numbers of trials in each condition. Similarly, each object had an independent probability of changing its bearing on each frame of a trial. Bearing changes could take on any value within 359° relative to the original bearing. The probability of a bearing change on any given frame was fixed within a trial at 0, 0.02, 0.04, or 0.06, with equal numbers of trials in each condition. When objects collided with the boundary of the display or with the center point of the display they deflected according to Newtonian principles. When objects came within 1.25° of one another, they each changed their bearing randomly to avoid colliding. Each trial lasted 10 s. The display as a whole subtended 27° × 20°.

A total of seven tracking loads (2–8), five speed change conditions, and four direction change conditions resulted in 140 conditions. We generated 10 trials for each of these conditions (1,400 total). When a model was tested, it performed each trial 10 independent times, performing differently each time because of randomly generated measurement noise. (This method of testing is parallel to the one used by Girshick, Landy, & Simoncelli, 2011, though we use fewer simulations for reasons elaborated below.)

Analysis

In Experiments 1 and 2 we analyze results by treating each trial as a subject. By simulating each trial 10 times, we obtain the average performance of each subject. Because we generated 10 different trials for each specific condition, we end up with 10 simulated subjects for each condition (similar to the number of measurements we typically obtain with human participants). Because all of the models are tested on the same set of trajectories, model type is treated as a within-subject factor, and target load is treated as a between-subject factor. We evaluate main effects of model type and target load with split-plot factorial analysis of variance (ANOVA) for each combination of temporal resolution and spatial standard deviation separately.

Behavioral results: Comparing model and human performance

Because the purpose of the computational experiments was to compare models with the same psychophysical limits, but with different approaches to extrapolation, we first sought to validate empirically that the range of psychophysical limits imposed on the models produces performance in the range of human observers. To do this, we tested human observers in a subset of representative trajectories intended for use in the computational experiments. (This subset included trials with perfect inertia). We then compared simulated model performance (obtained as described in the General methods) with that of the human observers.

Methods

Participants

Ten Johns Hopkins University undergraduates participated for course credit. All had normal or corrected-to-normal visual acuity. The protocol of this experiment was approved by the Homewood Institutional Review Board of Johns Hopkins University.

Apparatus

Stimuli were presented on a Macintosh iMAC computer with a refresh rate of 60 Hz. 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 and presented with MATLAB and the psychophysics toolbox (Brainard, 1997; Pelli, 1997). The trajectories were a subset of those used in the computational experiments. They included only target loads of three to eight. In order to cover trajectories with a range of inertia, while conducting an experiment of reasonable length, we included trials with the following four combinations of parameters: (a) trials with perfect inertia (probability of both speed and bearing change at zero); (b) 0.75 probability of speed change, zero probability of bearing change; (c) zero probability of a speed change and 0.04 probability of a bearing change; (d) 0.75 probability of a speed change and a 0.04 probability of a bearing change. Trials were randomly selected for each target load and setting combination producing an experiment that comprised 120 trials for each participant.

All stimuli were presented in a black square subtending 25.38° × 19.98°. Each trial started with six to 16 blue discs (diameter 0.94°) along with a white fixation cross (0.47° × 0.47°) in the center (which remained present throughout a trial). After 0.5 s, a subset of between three and eight discs turned yellow for 1.5 s, indicating that these were the targets. Finally, all discs turned blue again. After another 0.5 s, all of the discs moved, following preselected trajectories (see above) for 10 s. At the end of the motion, participants were prompted to click on the discs that they thought were the targets. When a participant clicked on a disc, it turned yellow. After the participant selected as many discs as there were targets, the true targets flashed in red to provide feedback.

Results

We compared human and model performance in a variety of ways. Figure 2 first displays general performance as a function of tracking load for human participants and the Kalman filter model. Human performance clearly declines as a function of tracking load. Similarly, Kalman filter performance generally declines as a function of tracking load. More to the point however, we plot performance for the model with a 0.2° spatial resolution, as well as the model with a 0.435° spatial resolution (in each case with a temporal sampling rate of 12 Hz). Our intention was that this range of resolution would allow us to capture the range of likely human performance. Some have proposed that

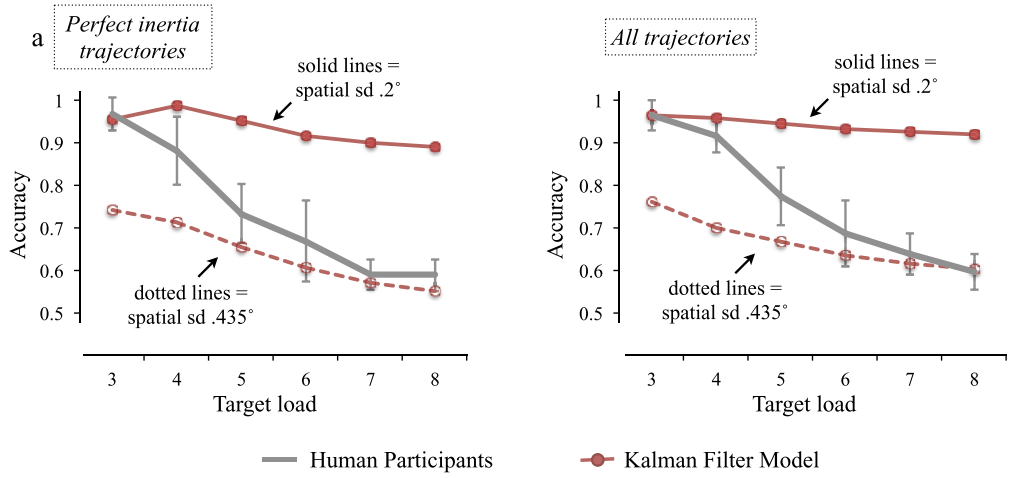

Figure 2. Comparison between performance of human observers ($N = 10$) and the Kalman filter model. Results from perfect inertia trajectories (a) and all trajectories (b) are plotted separately.

human performance declines as a function of tracking load because spatial precision declines (Bays & Husain, 2008; Vul, Frank, Alvarez, & Tenenbaum, 2009). We sought to remain neutral on the issue, instead hoping to identify features of extrapolation that might be true within the range of human ability, whatever the exact causes of changes in ability might be (both between and within individuals).

As will become clear in the forthcoming computational experiments, all the models we tested performed relatively similarly to one another. Thus to the extent that the Kalman filter's performance traced the boundaries of human performance as a function of memory load, all the models performed in this range. Moreover, our computational experiments included models with 0.6° spatial precision and also with 5 Hz and 20 Hz sampling rates (in addition to the 0.2/0.435° 12 Hz models shown). This ensured that any generalizations about extrapolation obtained from the computational experiments would apply across a relatively wide range of baseline abilities, including those most typically estimated for human observers.

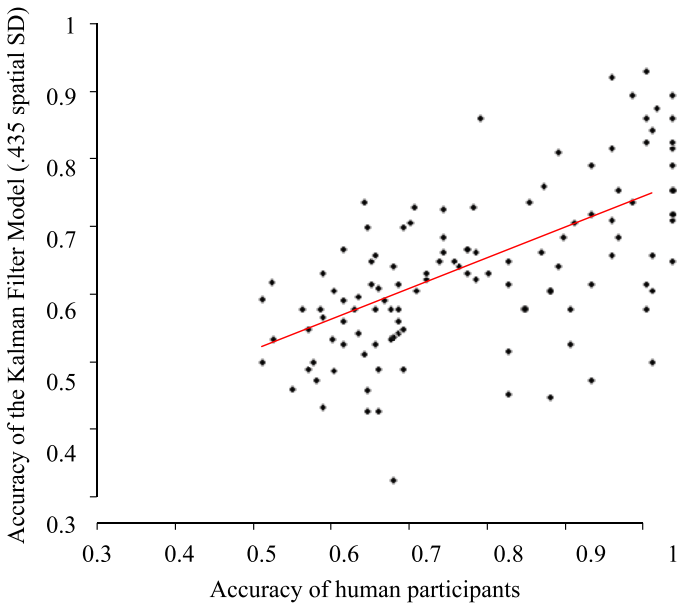

Figure 3. Trial-by-trial correlation between tracking performance of the Kalman filter model (shown with a spatial *SD* of 0.435) and human observers ($N = 10$).

To further investigate the relationship between the models and human performance, we considered the extent to which model performance correlated with that of human participants on a trial-by-trial basis. Specifically, trials generated randomly should vary randomly in terms of their difficulty, because some trials should, by chance, include more nearby interactions between targets and nontargets of the sort that cause tracking errors (Bae & Flombaum, 2012). Perhaps large enough differences between trials leads to systematic variability in performance that can be captured by the models.

In total, the 10 human observers completed 120 trials that we could compare with model performance. Each of the models completed each of these trials in 10 simulations. In this way we could correlate human performance averaged across 10 observers for each trial with model performance in each trial averaged across 10 simulations. Figure 3 shows the correlation between human observers and the 0.435° model. The correlations were significant for all the models tested (largest $p = 0.002$).

Some of this correlation likely arose from systematic effects of target load on performance. Accordingly, we recomputed correlations with target load as a control variable. Correlation coefficients and corresponding p values are listed in Table 1. All of the models showed significant correlations, though not always strong ones. This should be expected, because model errors (and, we

	Spatial *SD* 0.2°			Spatial *SD* 0.435°			
	Spatial working memory	Kalman filter	50/50 prediction	Spatial working memory	Kalman filter	50/50 prediction	Within human
r	0.206	0.235	0.327	0.340	0.404	0.266	0.563
p	0.025	0.010	<0.001	<0.001	<0.001	0.003	<0.001

Table 1. Partial correlations between human performance and the performance of the three kinds of models investigated, on a trial-by-trial basis and with target load as the control variable. All models shown include a sampling rate of 12 Hz. Results were similar for all models investigated.

presume, those made by human observers) occur probabilistically; an error is not necessarily made just because a target and nontarget pass closely by one another. To provide a sense of scale for the model correlations with human performance, we split the human data in half (based upon the order by which the participants had arrived in the lab), and we correlated trial-by-trial performance for the first five observers with the remaining five. The correlation was significant with a correlation coefficient of 0.563. Though larger than the correlations with model performance, this demarcates a kind of ceiling on the ability to predict trial-by-trial variability, at least with this many participants and trials.

Overall, the fact that the models correlated significantly with observers on a trial-by-trial basis after controlling for target load suggests that the algorithms and assumptions built into the models characterize at least some important aspects of the mechanisms underlying human abilities, and afford a reasonable tool for investigating the effectiveness of extrapolation during multiple object tracking. Moreover, that all three model incarnations (Kalman filter, spatial working memory, and 50/50 prediction) produced similar and significant correlations evidences their viability as candidate models for capturing algorithmic aspects of human tracking abilities. In general, we believe that predictions of trial-by-trial variance in performance, as opposed exclusive reliance on load-dependent variance, should become a standard way of comparing models of tracking in future work.

Computational Experiment 1: The advantage of prediction?

In many ways, the main implication of the following experiments is that extrapolations from noisy measurements may not improve tracking performance in MOT. Expecting extrapolations to be inaccurate has not played a role in previous discussions of human multiple object tracking. Many studies have sought evidence of extrapolation in the form of exact and accurate knowledge of where an object should be, for example, in the interruption experiments described earlier (Fencsik et al., 2007; Keane & Pylyshyn, 2006).

One of the models in the current experiment is a Kalman filter that weights extrapolations and current observations relative to one another. We contrast the performance of this model with a spatial working memory model that does not extrapolate, instead tracking entirely on the basis of where objects were last observed. As different as these models seem, in principle, note that with the Kalman gain fixed to **0**, the Kalman filter model reduces to the spatial working memory model; though the spatial working memory model enjoys computational savings by dispensing with the calculations associated with extrapolation. With a Kalman gain fixed at **1**, the Kalman filter model would fully expect objects to appear at newly calculated positions given currently available information about velocity. Weights between **0** and **1** reflect the models' confidence in its extrapolations. Accordingly, we can ask two questions in the current experiment. Is there a marginal advantage to using extrapolated predictions in a dynamically weighted way? And if predictions are made at all, how strongly should extrapolations be weighted, given human-like perceptual uncertainty?

Results and discussion

Figure 4a displays the proportion of targets tracked correctly by the Kalman filter model (drawn in red), and the spatial working memory model (drawn in blue), in only the trials with high inertia trajectories—the trajectories in which objects did not change their speed or bearing randomly. Figure 4b compares the Kalman filter model with the spatial working memory model, averaged across all trajectories tested (amounting to 1,400 unique trials).

There was a main effect of target load on tracking performance for all models with high inertia trajectories and in all other trajectories as well. (For high inertia trajectories lowest $F[6, 63] = 3.81$; for all trajectories combined, lowest $F[6, 1393] = 37.8$; all $p < 0.05$). For some parameter settings, there was also a significant main effect of model type. Statistics for these comparisons are shown in Tables 2a and 2b. Whenever

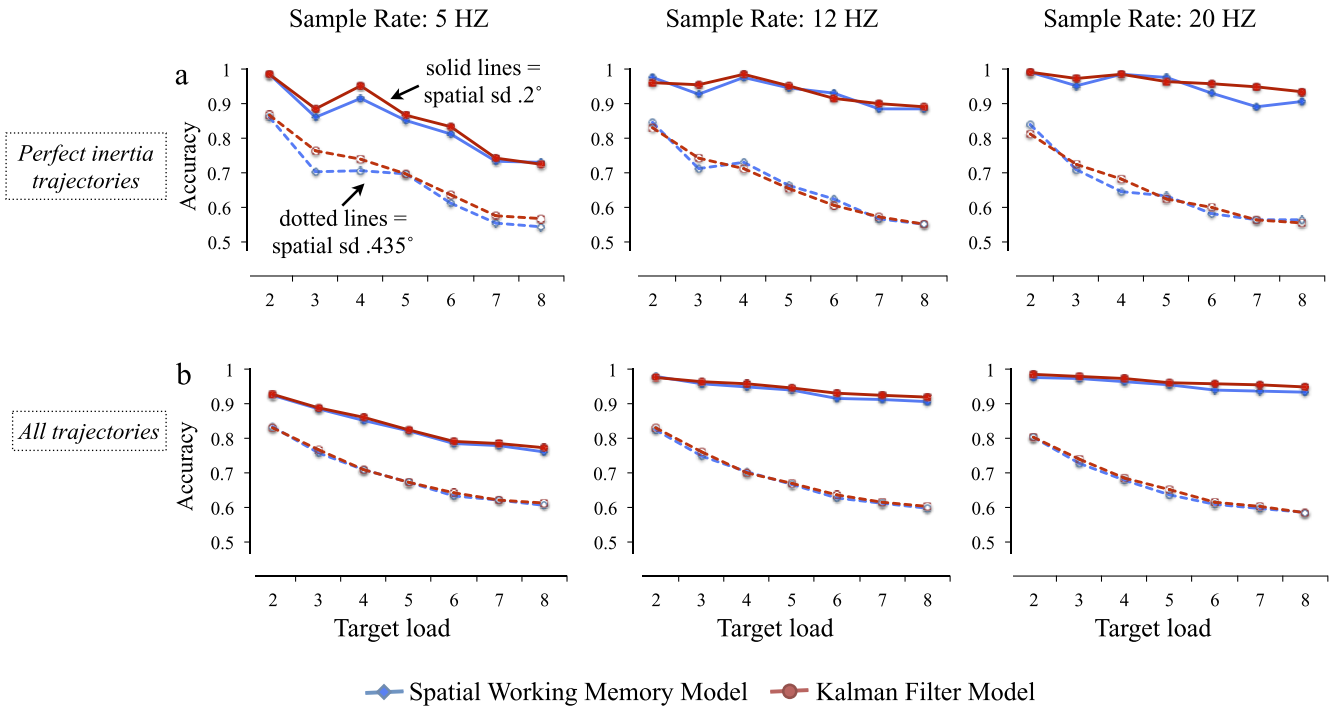

Figure 4. Results of Experiment 1. Models tested differed in terms of their sampling rates (separated by panels left to right) and spatial standard deviation (dotted lines = 0.435° models, solid lines = 0.2° models). There were no large differences in performance between the Kalman filter model (red) and the spatial working memory model (blue). This was true both for trajectories with perfect inertia (a), and averaged across the wide range of trajectories tested (b). (For simplicity models with 0.6° precision are not shown, though results were similar).

there was a statistically significant effect, the Kalman filter outperformed the spatial working memory model, but as is clear from the figures and the tables, any effects were small when present (largest mean difference between models = 2.4%).

In summary, the Kalman filter model enjoyed a small but significant advantage with some model parameters. But practically, performance was mostly comparable across model variants. Even for trials in which extrapolation should have conferred the greatest advantage (Howe & Holcombe, 2012), the marginal advantage of extrapolating—compared to just remembering where things were—was extremely small (e.g., compared to differences in performance as a function of target load).

The weights that the models placed on extrapolations (to produce adjusted priors) make the reason for these results apparent. Across all models, target loads, and trajectory types—even those with perfect inertia—the highest adjustment weight assigned (β) was 0.13, and usually it was closer to 0.095. (Note that the maximum possible weight is 1.0; the weights assigned by the models are much closer to the minimum possible value, which is zero). For illustrative purposes, Figure 5 graphs the weights assigned by the 12 Hz / 0.2° version of the model. We selected the model with the smallest *SD* tested because its predictions should have been the most accurate, and as a result, the most highly weighted.

	Spatial *SD* = 0.2°	Spatial *SD* = 0.435°
Sample rate = 5 HZ	$p = 0.03$ $\Delta = 1.4\%$	$p = 0.011$ $\Delta = 2.4\%$
Sample rate = 12 HZ	$p = 0.31$ $\Delta = 0.50\%$	$p = 0.71$ $\Delta = -0.30\%$
Sample rate = 20 HZ	$p < 0.001$ $\Delta = 1.8\%$	$p = 0.76$ $\Delta = 0.30\%$

Table 2a. *p* values for main effects of model type with high inertia trajectories, $F(1, 63)$, along with mean differences in performance (Δ = Kalman filter model – spatial working memory model).

	Spatial *SD* = 0.2°	Spatial *SD* = 0.435°
Sample rate = 5 HZ	$p < 0.001$ $\Delta = 0.60\%$	$p = 0.61$ $\Delta = 0.30\%$
Sample rate = 12 HZ	$p < 0.001$ $\Delta = 0.90\%$	$p = 0.02$ $\Delta = 0.50\%$
Sample rate = 20 HZ	$p < 0.001$ $\Delta = 0.11\%$	$p = 0.004$ $\Delta = 0.60\%$

Table 2b. *p* values for main effects of model type across all trajectory types, $F(1, 1393)$, along with mean differences in performance (Δ = Kalman filter model – spatial working memory model).

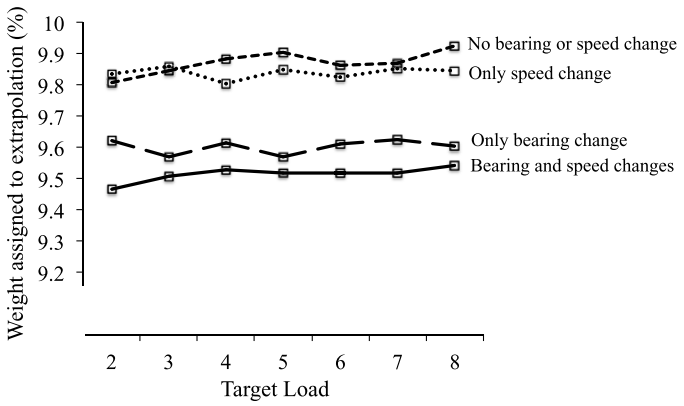

Figure 5. Results of Experiment 1. Weights assigned to extrapolation by the Kalman filter model as a function of trajectory type (and target load). For illustrative purposes, we only show weights assigned by the 12 Hz version of the model with a spatial standard deviation of 0.2° (but results were similar for all model variations). Though differences were small, higher weights were assigned to extrapolations in trajectories without bearing changes compared to those with bearing changes. In general, weights assigned to extrapolations were between 9% and 13%.

The results were similar for all variations tested. The figure divides these results across four types of trajectories: those in which speed and bearing remained constant for an object throughout a trial, those in which only speed could change, those in which bearing could change, and those in which both speed and bearing could change. Differences are very small but instructive. Extrapolations were weighted less when object bearings were subject to random changes. Recall that the Kalman filter adjusted its own weights over time (the weights shown are averages across trials and time points). Thus the model responded differently to trials with bearing changes in comparison to those without; but it nonetheless assigned a low weight to extrapolation—less than 10%—even for trials without frequent bearing changes.

Computational Experiment 1b

To follow up on these results, we again compared performance between the Kalman filter model and the spatial working memory model in one new group of trials. In particular, we were concerned that even our high inertial trials were not high inertia enough, owing to trajectory changes whenever objects approached one another (to prevent overlaps and occlusion). We thus generated a new set of 70 trials in which each object had a 0% chance of changing speed or bearing, except when colliding with one of the boundaries of the display. When objects approached one another, they maintained their trajectories as though nothing were in the way, essentially passing through one another.

Furthermore, for our Kalman filter model, which made linear predictions, it is possible that collisions with the boundary could also create confusion by producing relatively sudden changes. Specifically, collisions with the boundary in our trajectories followed Newtonian principles, with the angle of reflection equaling the angle of incidence. But at the boundaries, the Kalman filter model continued to make linear predictions, which were fundamentally useless and practically put the model in the same position as the spatial working memory model. We therefore ran the simulations of the newly generated trajectories with two changes to the Kalman filter model. First, whenever the model made a prediction that placed an object beyond the boundary of the display, those predictions were automatically adjusted to conform to the actual nature of the boundary collision mechanics. No additional noise was injected into this process, allowing the model to make predictions that were as accurate as possible given only the noise that went into the linear predictions it would have made otherwise.

This adjustment amounts to what is perhaps an unrealistic advantage compared to human observers. That is, the model possessed perfect knowledge of the rules that governed collisions in these displays.

The second adjustment made to this model involved the uncertainty associated with the perception of velocity. In the Kalman filter model, as we implemented it initially, predictions were corrupted by noise in both observations of object position and observations of velocity (Equation 7) This was done to capture the possibility that there is independent noise in neural channels associated with motion perception, which could be involved in inferences about velocity (Burr & Thompson, 2011, Stocker & Simoncelli, 2006). Out of concern that this made the model excessively uncertain about extrapolations in particular, the model implemented in Experiment 1b did not include velocity noise. Predictions concerning velocity where made on the basis of Equation 7, but omitting the term ov_t^m.

With these two adjustments, the model still did not outperform the spatial working memory model, as shown in Figure 6.

To summarize Experiment 1b: We tested a model with considerably less prediction uncertainty than the model previously implemented, both by removing independent noise in the perception of velocity and endowing the model with advance knowledge about how collisions with display boundaries should evolve. Still, the pattern of results was similar to those in Experiment 1. The Kalman filter model rarely outperformed its spatial working memory counterpart, and when it did, effects were relatively minor and limited to smaller tracking loads (see Figure 6). Again the

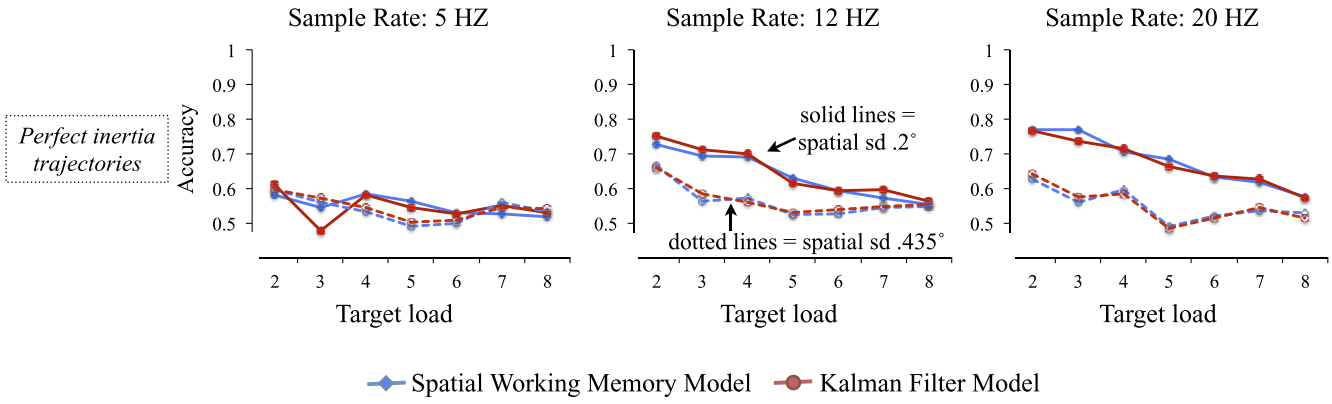

Figure 6. Results of Experiment 1b. Comparison between the Kalman filter model and the spatial working memory model. In these simulations, the Kalman filter model made correct predictions about object deflections during collisions with the boundary, and it also did not include any independent noise in velocity channels, only the noise resulting from observations of spatial position.

Kalman filter ultimately placed relatively low weights on extrapolation. The 20 Hz / 0.2° model—in this case, with no extra noise in velocity channels—arrived at the highest value among all the models tested so far (15%).

Summary

Two insights follow from this first computational experiment. The first is that extrapolation only appears beneficial to a small extent and within a limited scope—with very dependable trajectories and when tracking few objects, perhaps only when tracking two. This is consistent with the clearest evidence of extrapolation found in behavioral work on MOT (Fencsik et al., 2007; Howe & Holcombe, 2012).

By testing models known to employ different strategies, and by comparing them across a wide array of trials that included incrementally more trajectory changes, we can add that in many contexts, extrapolation may provide very little advantage. This is important because in order to employ a strategy only when trajectories are dependable, an observer would need some way of determining that she finds herself in the relevant context. How would she discover this? How good can we expect observers to be at discriminating between dependable and only slightly less dependable trajectories?

Surely an observer would use the same noisy inputs that form the basis of extrapolation to determine the dependability of object trajectories. Indeed, one can interpret the weights placed on extrapolation as an estimate of how dependable an observer thinks the trajectories are. An observer might reasonably think that the context is reliable to the extent that she can make reliable predictions. From this perspective the low weights placed on extrapolation by the Kalman filter on all occasions evidences the difficulty of discriminating between trajectory types. The just noticeable difference (JND) for trajectory inertia, as it were, is apparently very large. This is also consistent with previous behavioral work wherein observers controlled object inertia from trial to trial in an attempt to identify trials that felt easiest. Responses were extremely noisy, and performance was largely invariant with respect to inertia (Vul et al., 2009).

The second insight is that even an observer who is trying to extrapolate can behave a lot like an observer who is not trying. This is what the Kalman filter model did by weighting observations highly relative to extrapolations. It is difficult to know exactly what human observers do, as evidenced by the mixed results in the behavioral literature. This is perhaps because realistic psychophysical limits constrain the inputs to extrapolation, so that a rational observer should severely bias any predictions she makes towards her most recent observations. If human observers employ an extrapolation strategy that accounts for their own psychophysical uncertainty, they could be trying to make predictions, but still resemble observers who do not make predictions at all. This is especially so in the context of MOT, where a difference between predicting and not would only appear if it precludes target and nontargets confusions frequently enough to produce a noticeable performance difference.

Computational Experiment 2: Extrapolating rigidly versus not at all

Perhaps continuous adjustment of one's extrapolation strategy is counterproductive in the context of

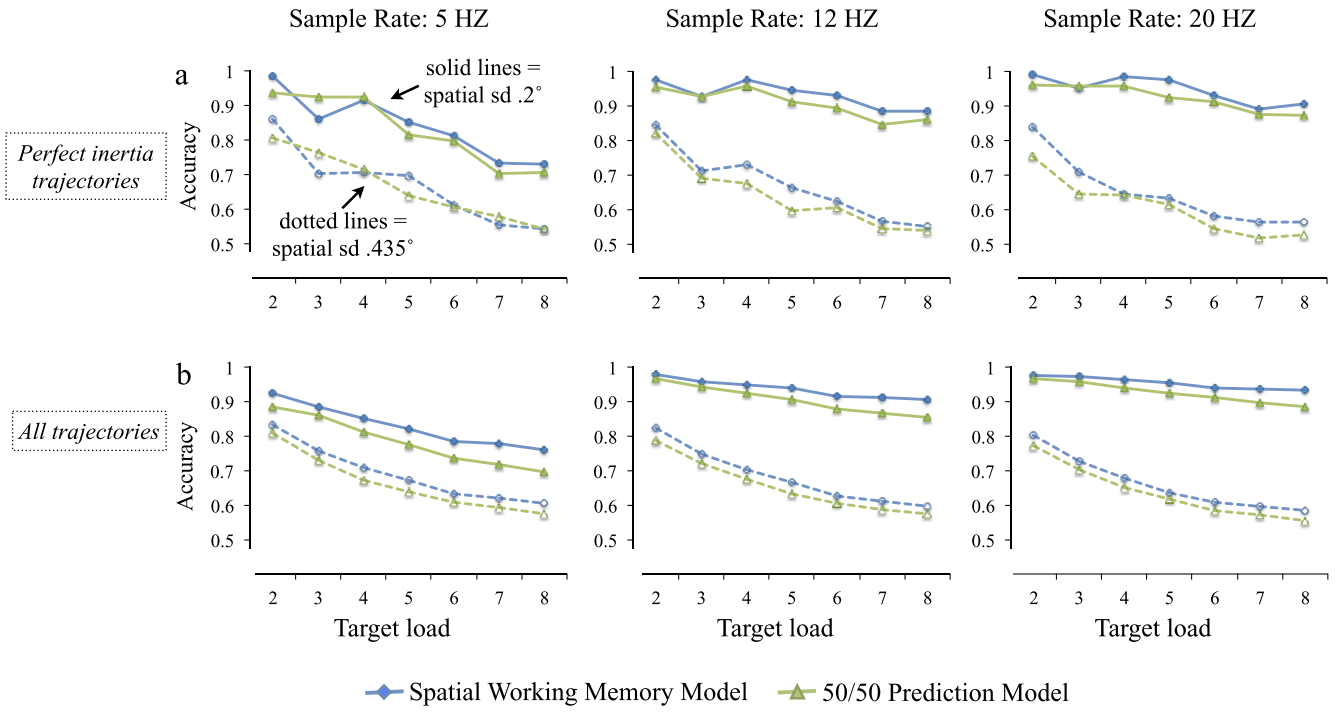

Figure 7. Results of Experiment 2. Models tested differed in terms of their sampling rates (separated by panels left to right), spatial standard deviations (dotted lines = 0.435° models, solid lines = 0.2° models), and whether they extrapolated (green) or not (blue). The spatial working memory model that did not extrapolate outperformed comparable 50/50 prediction models. This was true both for trajectories with perfect inertia (a), and averaged across the wide range of trajectories tested (b). (For simplicity models with 0.6° precision are not shown, though results were similar).

MOT, excessive micromanagement? Perhaps one would be better off behaving more rigidly—adopting a strategy and sticking to it? From the perspective of typical research on Kalman filters and related algorithms, this seems unlikely. But as we noted before, many such applications do not involve a correspondence problem that relies exclusively on spatiotemporal data. Accordingly, we investigated the performance of a third tracking model, one that utilizes a rigid strategy to rely on extrapolations. This rigid model is also closer in spirit to typical discussions of extrapolation in the MOT literature, which have not explicitly considered the possibility of an adjustably weighted strategy (e.g., Fencsik et al., 2007; Keane & Pylyshyn, 2006; Vul et al., 2009).

In particular, we test a model that permanently weights observations and predictions equally, the 50/50 prediction model. This model is a special case of the Kalman filter model, with the Kalman gain permanently set to **0.5**. (We did not test a model with gain set to **1** because pilot experiments suggested that such a model performs very poorly). Moreover, a 50/50 model is probably a good choice for an inflexible strategy, assuming one does not know the best weighting in advance. With no a priori reason to weight one source of evidence more than another, one should weight them equally.

Overall, the question of interest was whether the 50/50 prediction model would outperform the spatial working memory model from Experiment 1, and by how much. Small differences—or better performance for the spatial working memory model—would suggest that rigid extrapolation is not a good strategy in MOT, and that the Kalman filter of Experiment 1 does not suffer for adjusting its weights too much or too frequently.

Methods

This experiment was identical to Experiment 1, testing the 50/50 prediction model on all the same trajectories as the two models described previously.

Results and discussion

Figure 7a displays the proportion of targets tracked correctly by each tested model on only the trajectories with very high inertia —the trajectories in which objects did not change their speed or bearing randomly. Green lines identify the 50/50 prediction model and for comparison, blue lines identify the spatial working memory model. The spatial working

	Spatial $SD = 0.2°$	Spatial $SD = 0.435°$
Sample rate = 5 HZ	$p = 0.26$	$p = 0.74$
	$\Delta = 1.2\%$	$\Delta = 0.30\%$
Sample rate = 12 HZ	$p < 0.001$	$p = 0.007$
	$\Delta = 2.4\%$	$\Delta = 3.1\%$
Sample rate = 20 HZ	$p < 0.001$	$p < 0.001$
	$\Delta = 2.4\%$	$\Delta = 4.2\%$

Table 3a. p values for main effects of model type in high inertia trajectories, $F(1, 63)$ along with mean differences in performance (Δ = spatial working memory model − 50/50 prediction model).

	Spatial $SD = 0.2°$	Spatial $SD = 0.435°$
Sample rate = 5 HZ	$p < 0.001$	$p < 0.001$
	$\Delta = 4.6\%$	$\Delta = 2.9\%$
Sample rate = 12 HZ	$p < 0.001$	$p < 0.001$
	$\Delta = 3.1\%$	$\Delta = 2.7\%$
Sample rate = 20 HZ	$p < 0.001$	$p < 0.001$
	$\Delta = 2.8\%$	$\Delta = 2.6\%$

Table 3b. p values for main effects of model type in all trajectories, $F(1, 1393)$ along with mean differences in performance (Δ = spatial working memory model − 50/50 prediction model).

memory model outperformed at most tracking loads. Figure 7b compares the same models, averaged across all trajectory variations employed (amounting to 1,400 unique trials). Here the pattern is very clear: The spatial working memory models always outperformed their 50/50 counterparts by between 1% to 5% depending on the tracking load and spatial precision.

Statistically, there was a main effect of target load on tracking performance for all models with high inertia trajectories and in all other trajectories as well. (For high inertia trajectories lowest $F[6, 63] = 4.5$; for all trajectories combined, lowest $F[6, 1393] = 66.3$; all $p < 0.05$). For some parameter settings, there was also a significant main effect of model type. Statistics for these comparisons are shown in Tables 3a and 3b. Whenever there was a statistically significant effect, the spatial working memory model outperformed the 50/50 prediction model.

These results belie the intuition that extrapolation is always beneficial. Even with relatively predictable trajectories, using extrapolations almost never led to improved performance. Instead, implementing a simple proximity heuristic (Franconeri et al., 2012) seemed equal to the task of accomplishing MOT, at least compared to the rigid prediction model tested.

In retrospect, it seems clear why. Consider how the 50/50 model made predictions: it compared inferred object positions at one moment in time with positions at a previous moment in time to estimate speeds and bearings. But these estimates were subject to noise in observations of position, and as a result, they were noisy as well. This is why the Kalman filter model in Experiment 1 consistently reduced its reliance on extrapolations. We will come back to this point in the General discussion.

Replications of Experiment 1

The results of Experiment 1 were surprising, though perhaps not in light of the mixed observations in the experimental literature. In particular, it appeared that making predictions about object trajectories, even when those trajectories were linear and highly reliable, did not improve tracking performance very much, defined here as the ability to tell apart targets and nontargets. Experiment 3 will investigate whether these results can potentially explain human behavior in a canonical extrapolation experiment.

But first, we were concerned that the results of Experiment 1 were in some way artifacts of the kinds of choices one needs to make when building an algorithm. Specifically, Kalman filters are not typically used in situations with more than one state to estimate at a time, nor in situations where there is a data association problem—what we have called a correspondence problem—with respect to linking noisy data and the states one wants to estimate. As a result, we were forced to make choices in our implementation, concerning which variable to use as the basis for addressing the correspondence problem. We decided on what we had termed an *adjusted prior*, a prediction that was weighted by β, combining the current posterior, and the extrapolated prior. We did this for a number of reasons, among which was the finding that fixed β models, like the one used in Experiment 2, did not perform well. But the concern is that the actual velocity estimate now used by the model is a quantity derived outside of the standard components of the Kalman filter, an adjustment on what would otherwise be a straightforward subtraction of the most recent posterior estimates (Equations 7 and 8). Perhaps, as a result, our mongrel Kalman filter is no longer optimized? More generally, perhaps slightly different choices about how to address state inference and data assignment in the same algorithm would simply work more effectively?

To investigate these concerns, we implemented two additional models, both of which extrapolated on the basis of position and velocity covariance exactly as typically done in a Kalman filter. The relevant variables to consider in these models are the quantity used to make correspondence assignments, and the

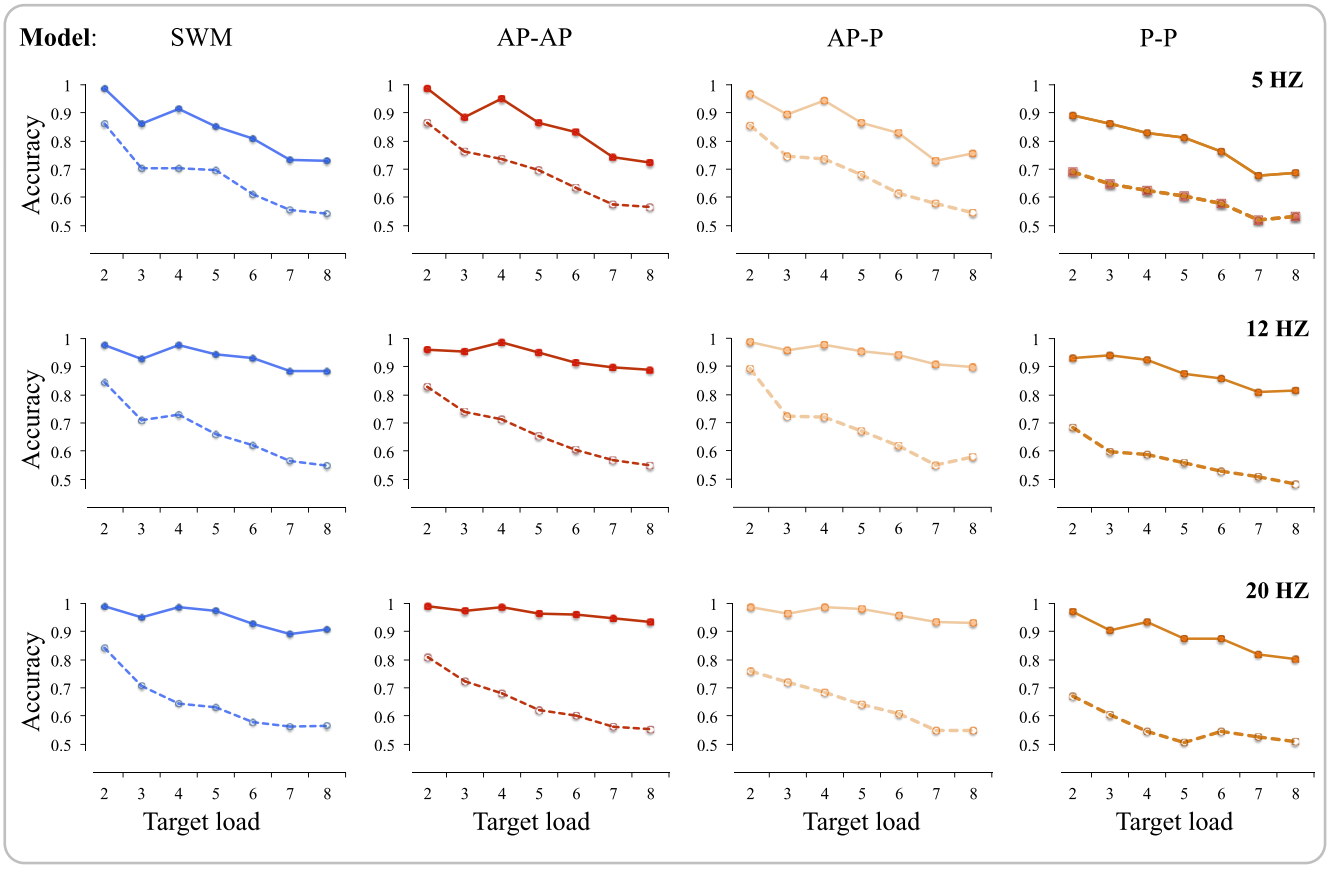

Figure 8. Performance comparison among models in trajectories with perfect inertia. All models presented possess the basic skeleton from the main Kalman filter model in Experiment 1b. SWM refers to the spatial working memory model. AP-AP refers to the model in Experiment 1b. The first position in each model name refers to the variable used to address correspondences and the second refers to the variable that is combined with assigned observations to infer new posterior estimates. AP refers to adjusted prior, and P refers to unadjusted prior. Figure 9 summarizes these results in terms of performance collapsed across target load.

quantity combined with an assigned observation to make inferences about an object's current position—to infer a posterior. The model from Experiment 1 used the adjusted prior for both. So here, we will call it the AP-AP model, where the first AP refers to adjusted prior for correspondence and the second for posterior inference. The two new models we call AP-P and P-P, referring, respectively, to adjusted prior for correspondence, prior for motion inference, and prior for both. Prior refers to the extrapolated position of the object, without adjustment, on the basis of a current, dynamic process. Thus the two new models combine assigned observations and unadjusted priors to infer new posteriors. Both of the models are faithful to the standard Kalman filter optimization. The difference between the two is only in the quantity used to infer correspondences with observations, a process not usually built into the Kalman filter, and for which we are unaware of relevant optimization theorems.[2]

Results and discussion

We replicated Experiment 1 with each of the new models. The results are shown in Figures 8, 9, 10, and 11, for comparison, with the performance of the SWM and the AP-AP models from Experiment 1b. Statistical analyses revealed that the P-P model significantly underperformed all others ($p < 0.001$ for all comparisons). In only one case (20 Hz / 0.2°), the AP-P model significantly outperformed the SWM and AP-AP models, but by no more than 0.3%. These results suggest that our main discovery—relatively similar performance for a predictive filter and a non-predictive one—is not the consequence of estimating position and velocity separately, thus extracting the inference of posteriors from the standard measurement-prediction loop. This does not challenge previous work on Kalman filter optimization however, because the filter

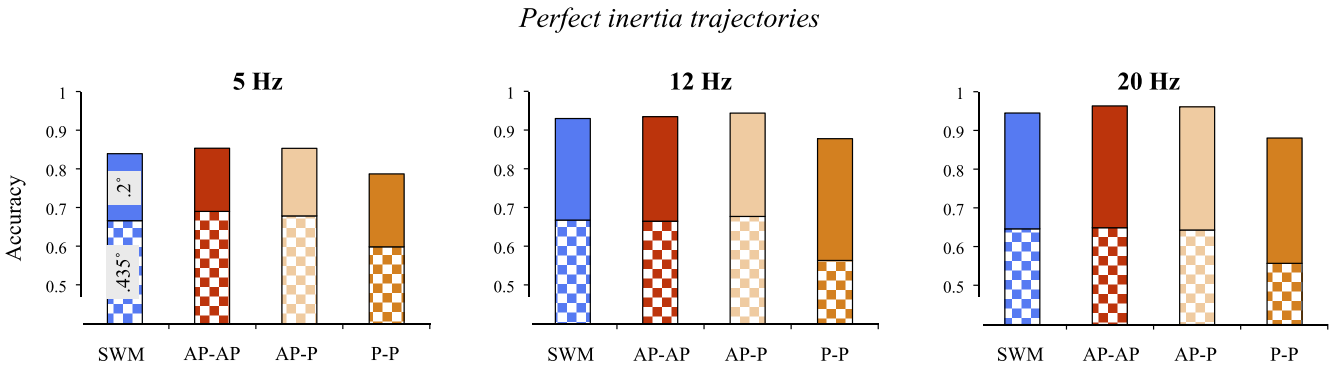

Figure 9. Comparison among models, collapsed across tracking loads (perfect inertia trajectories only). Dotted portions of each graph designate performance of 0.435° models, and solid portions that of 0.2° models.

is not designed for situations with a correspondence or data assignment problem in the first place.

Given the results of Experiments 1 and 2, and this replication, the question arises, "why does extrapolating not seem to confer an advantage in MOT?" We return to this issue in the General discussion after first considering the models in a slightly modified MOT task previously used to investigate the possibility that human participants extrapolate.

Computational Experiment 3: Identifying targets following a global interruption

In the literature on human MOT abilities, the intuition that observers should and do extrapolate was initially tested using a global interruption paradigm (Fencsik et al., 2007; Keane & Pylyshyn, 2006). In the

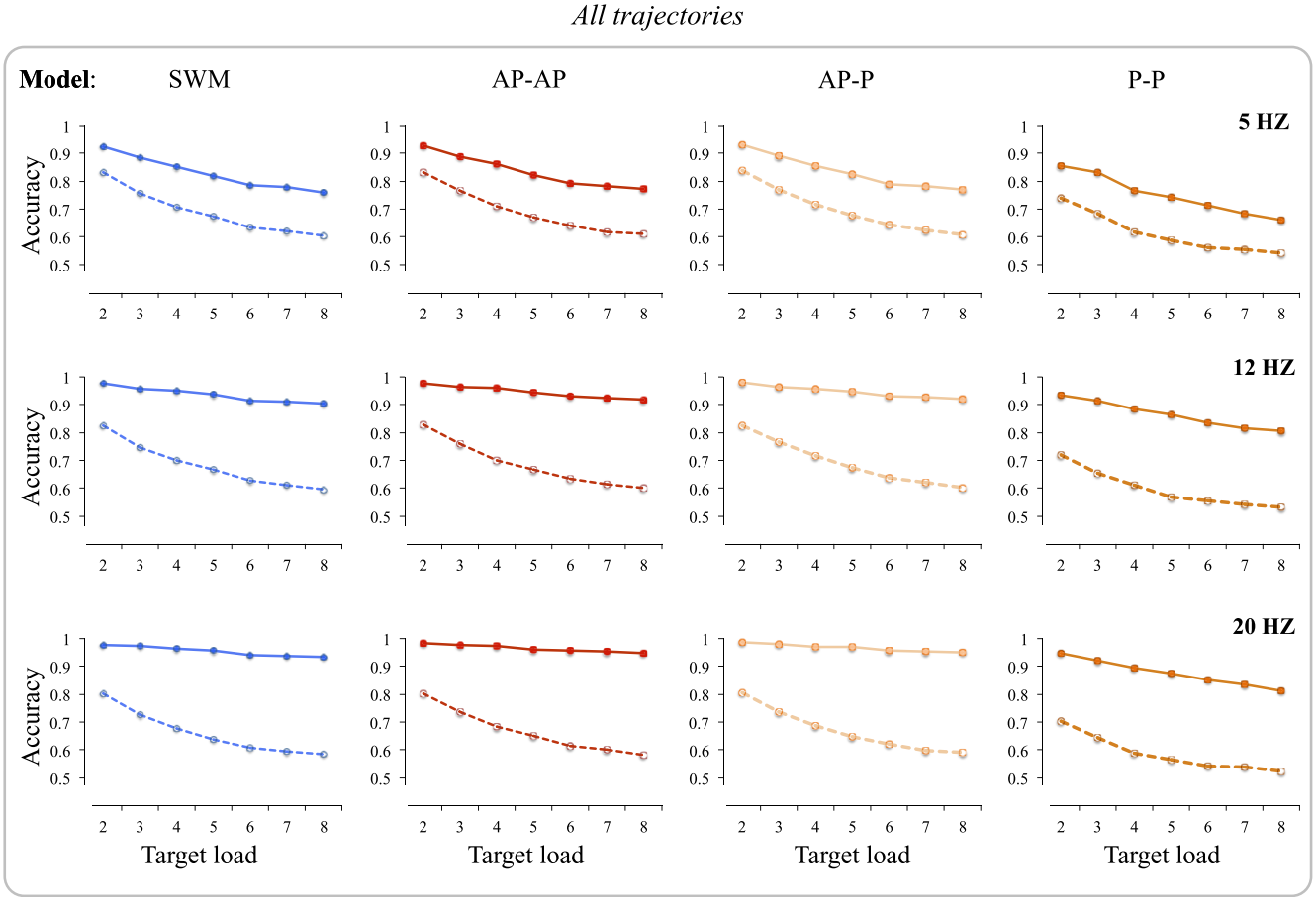

Figure 10. Performance comparison among models in all trajectories.

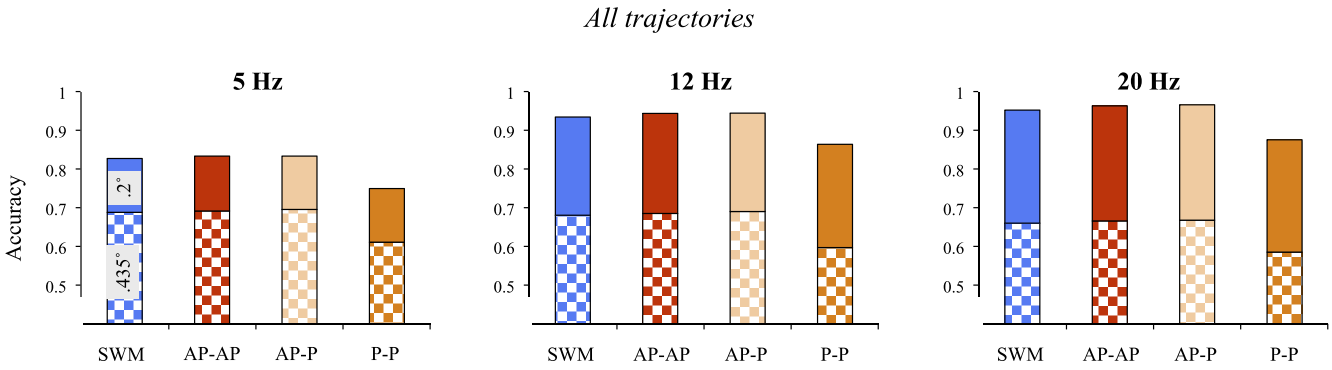

Figure 11. Comparison among models, collapsed across tracking loads (all trajectories together). Dotted portions of each graph designate performance of 0.435° models, and solid portions that of 0.2° models.

middle of a trial, the display became blank for a short duration, and following the interruption, observers were required to continue tracking, eventually identifying the targets. The results of these studies have been surprising. Participants perform better when objects reappear at their disappearance locations compared to when they reappear in their trajectory-determined locations. This has been interpreted to suggest that participants mostly do not extrapolate during MOT (Fencsik et al., 2007; Keane & Pylyshyn, 2006).

The results of Experiments 1 and 2 point to a potential explanation. These interruption studies have sought evidence of extrapolation by placing objects exactly where they should be given their trajectories prior to an interruption. Recognizing that a human observer starts with noisy inputs and that she would not start with prior knowledge of the kinds of trajectories employed, it becomes clear that expectations about where objects should appear next could deviate considerably from where they actually appear. Moreover, if they extrapolate by evaluating and then weighting extrapolations, observers may ultimately employ a strategy that looks quite a bit more like not extrapolating, and which produces tracking performance that is essentially indistinguishable—that is, if they behave something like the models we implanted that utilize an adjusted prior for correspondence assignments. Of course an observer who does not extrapolate would not perform better in a global interruption experiment when objects reappear at new, extrapolated positions compared to when they reappear where last seen. But perhaps the same can be said of an observer who does extrapolate, in particular an observer who behaves like our adjusted prior models, and learns to extrapolate conservatively?

In the current computational experiment, we used the models from the previous two experiments to replicate what was reported as Experiment 1 by Fencsik et al. (2007) and Keane and Pylyshyn (2006). The models tracked targets in trials with a global interruption. In one condition, the objects reappeared where they disappeared. In the other condition, they reappeared at their trajectory-determined positions. Model performance was compared across the two experimental conditions.

Recall that Experiment 1 included trials in which objects always ended up where they were going —trials with reliable trajectories. Yet the model that embodied such an expectation did not perform better than the model without it. In this experiment, we can again compare models with different kinds of expectations. But we can also compare the models to themselves in different trial types. In the case of the Kalman filter this allows us to ask whether an observer who does expect objects to move where they are headed actually performs better when the objects conform to this expectation compared to when the objects do not.

Methods

We tested the apatial working memory model and the Kalman filter model from Experiment 1. The trajectories and conditions employed were nearly identical to those used by Fencsik et al. (2007) in their experiment 1.

Specifically, we used the same trajectories as in our Experiment 1 with the following modifications to match those in Fencsik et al. (2007). First, we only included trajectories with perfect inertia—trajectories in which objects did not change their speed or bearing randomly. Second, we used a total tracking duration of 5 s for each trial. At a random point in each trial, all of the stimuli disappeared for 307 ms (this was at least 2 s after tracking started, and at least 1 s before it ended). There were two conditions in the experiment governing where the objects reappeared following this gap. In half of trials, the objects appeared exactly where they disappeared (no move condition). And in the remaining half, they appeared exactly where they should have, had they maintained their pre-gap trajectories (move condition).

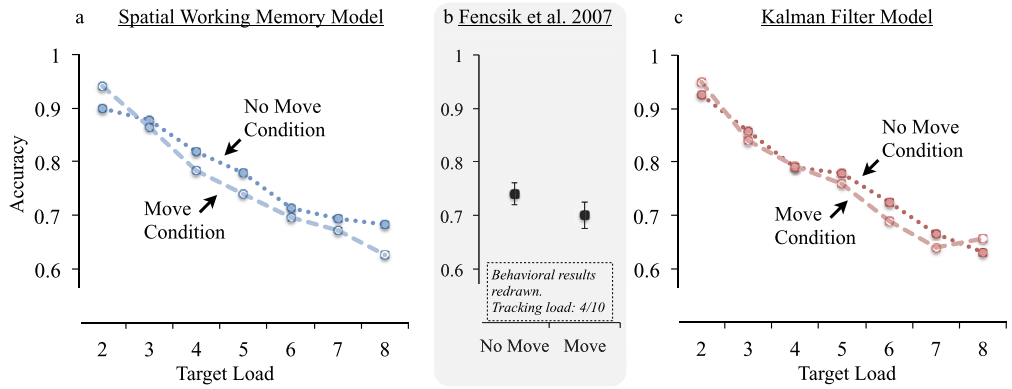

Figure 12. Results of Experiment 3. Performance of the spatial working memory model (a) and the Kalman filter (c) in a replication of experiment 1 from Fencsik et al. (2007). Results only shown for model variations with a spatial standard deviation of 0.435° and sampling rate of 12 Hz (but results were similar for all models). For comparison, (b) shows behavioral results redrawn from Fencsik et al. (2007; experiment 1). That experiment only included a tracking load of 4/10.

Because we tested models, we were able to generate both move and no move versions of each individual trial, testing the model on each, and affording trial-based comparisons. The models performed a total of 10 trials at each target load, doing a version of each trial that ended with the move condition, and a version that ended with a no move condition. The models did each trial by condition combination 10 times. Note that the Kalman filter model adjusted its Kalman gain at each sampling step in each trial. Thus there was no relationship between the weights set in move compared to no move trials, and a model could, in principle, have selected meaningfully different values.

Results and discussion

As previously found for human observers (Fencsik et al., 2007; Keane & Pylyshyn, 2006), the spatial working memory model performed better in the no move compared to the move condition. Figure 12 graphs performance as a function of target load for the spatial working memory model that utilized a spatial standard deviation value of 0.435° and a sampling rate of 12 Hz (results were similar for other model variations), the comparable Kalman filter model, and for comparison, the performance of human observers in the study by Fencsik et al. (2007).

Statistically, we analyzed these data by treating each trial as a subject, producing a simulated experiment with 10 subjects in each target load by condition. Because there were identical versions of each trial that ended with either a move or a no move manipulation, we treated reappearance position as a within-subject factor, and target load as a between-subject factor. Average model performance over 10 runs on each individual trial by condition was the dependent measure. A 2 × 7 split-plot factorial ANOVA revealed a significant main effect of reappearance position, $F(1, 63) = 4.394$, $p < 0.05$. There was also a significant main effect of target load, $F(6, 63) = 10.629$, $p < 0.001$. But there was no significant interaction between the two, $F(6, 63) = 1.437$, $p > 0.05$. Although the effects here are not very large—between 2% and 5% depending on the target load—they are comparable to those reported in the human literature. Fencsik et al. (2007) only tested a tracking load of four targets in their experiment 1 and obtained a difference of about 4% (Figure 12c). In Figure 12a the difference at tracking load four is 3.5%.

The Kalman filter model also performed better in no move compared to move trials (Figure 12b), but the difference was not significant, $F(1, 63) = 0.528$, $p > 0.05$. These results are certainly consistent with the theory that human observers use only spatial working memory during MOT (Franconeri et al., 2012; Keane & Pylyshyn, 2006). But we do not encourage the use of this computational experiment to draw that conclusion. Note that performance differences for the Kalman filter model were in the right direction and even close in magnitude to those observed in humans. But we tested a wider range of tracking loads than have been explored with human observers. Ultimately, using these models to make confident inferences about what humans do would require models that are known to perform similarly to humans across even a wider range of testing conditions. In particular, extrapolation weights and performance depend on the spatial precision parameters and sampling rates used, factors that will vary by individual within a group of participants. We compared ranges of parameter values, but without between subject individual differences, and with the same parameter values across tracking loads. This was done so that our results would not apply narrowly to any selected or fit values. But the tradeoff is that the computational experiments lack the kind of between-subject variability found in a behavioral

experiment. Future work should include model fitting, individual differences within a (modeled) experimental group, and/or effects of tracking load on precision and sampling rate, which would be better suited to identifying an effect of a specific experimental manipulation on tracking performance.

The point to emphasize, however, pertains to the possibility of a difference between internal algorithm and outward behavior. Specifically, the experiment demonstrates that an observer who does extrapolate could perform similarly to one who does not in the relevant experiment. The Kalman filter model in this experiment was extrapolating. It expected objects to end up where they were going.

Yet it was no worse off when the objects did not conform to this expectation. Comparing Figures 10a and 10c it is also clear that the Kalman filter performed no differently with respect to tracking success than the spatial working memory model in either type of interruption condition. Like in Experiment 1, there were no significant performance differences between the two models: move condition, $F(1, 63) = 0.018$; for no move condition, $F(1, 63) = 2.79$; both $ps > 0.1$.

Thus a model that was trying to extrapolate failed to outperform a model that did not try, even when unobservable moments of motion conformed to expectations that only the Kalman filter possessed. Practically, the experiment illustrates the challenge of seeking evidence of extrapolation in MOT by looking for performance advantages when objects maintain trajectory. Those advantages could fail to materialize, even when extrapolation is employed. Extrapolating conservatively as a response to noisy measurements can produce behavior that closely resembles not extrapolating at all.

General discussion

Whether or not human observers extrapolate when performing multiple object tracking has been a recent focus of research. But results have led to mixed conclusions (Atsma, Koning, & van Lier, 2012; Franconeri et al., 2012; Iordanescu, Graboweky, & Suzuki, 2009; Keane & Pylyshyn, 2006; St. Clair et al., 2010). In an attempt to better understand these results, we investigated an algorithm by which extrapolation could be implemented given noisy inputs, which in turn allowed us to address two important and related questions: (a) Is there an advantage to extrapolating? (b) How highly should an observer weight extrapolations in order to use them effectively? Succinctly, should an observer extrapolate, and if so, how? These questions may have been difficult to answer via an experimental approach; it would be challenging to compare human performance while extrapolating or not given that it remains unclear whether extrapolation is ever used under typical MOT conditions.

The experiments compared several nearly identical models, including one that did not extrapolate at all, one that extrapolated in a relatively rigid way, and ones that extrapolated in a computationally flexible way, weighting predictions in light of feedback. In Experiment 1, we discovered that weighting extrapolations confers a relatively small advantage, at best, and generally only in trials with very dependable object trajectories. We also replicated this result several times with slightly different models in terms of the variables that they treated as predictions with respect to assigning correspondences and estimating new posteriors. In Experiment 2 we discovered that extrapolating rigidly comes at a cost to performance compared to not extrapolating at all.

With respect to the second question—concerning how to weight extrapolations—we discovered that extrapolations should be weighted very modestly compared to recent observations. Expectations about object positions—even if they include a component driven by extrapolations—should be biased heavily toward recently observed positions. In retrospect this is at least partially obvious, because if extrapolations are derived from noisy observations, their expected noise cannot be less than the noise expected of a new observation.

These results stand to reconcile mixed previous results with respect to prediction and multiple object tracking. More broadly, they illustrate the computational complexities of the MOT task given psychophysical limitations and they illustrate the value of using algorithmic computational models to investigate tracking. We discuss these implications below, after first discussing potential explanations for the primary, counterintuitive result, that extrapolating seems not to confer an advantage in multiple object tracking.

Why doesn't extrapolation help?

Frankly, we initiated this study with the expectation that extrapolation would support tracking performance, but only under very limited circumstances and only with very high temporal and spatial resolution. But what we found was that in MOT, it seems not to help across a wide range of model variations and psychophysical parameters. Upon reflection, we suspect that there are two broad reasons for this.

The first reason amounts to a kind of lesson about MOT: that it is won and lost at the margins. What we mean by this is that inevitably, some tracking errors will be unavoidable, and caused by the fact of noisy target and nontarget representations, regardless of

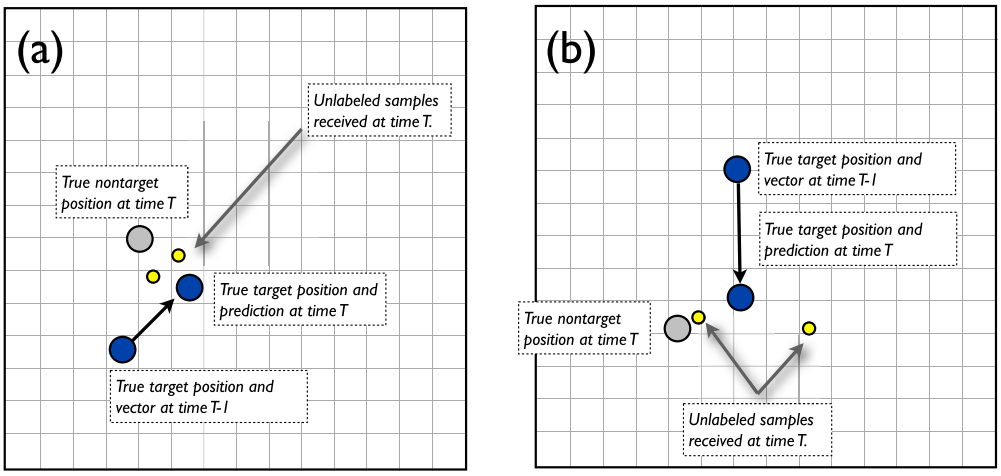

Figure 13. Two schematic situations in which perfect prediction would not preclude a tracking error. The true locations of a target (blue) and a distractor (gray) are shown, together with unlabeled samples (yellow). (Note that in MOT, the targets and nontargets share the same color). (a) Noisy observations lead to two potential correspondences (one correct and one incorrect) with equal probability. (b) Noisy observations lead to an incorrect correspondence appearing more likely.

prediction quality. Consider two situations, as examples, schematized in Figure 13—both situations in which possession of a perfectly accurate prediction would not preclude a tracking error. In Panel A, suppose an observer made the correct prediction about where the target would appear at time T on the basis of perfectly accurate knowledge of time T-1. The noisy, and unlabeled samples received from the target and nontarget at time T would still be indistinguishable. In Panel B, the sample from a nontarget just happens be closer to the target's true position, and thus the nontarget sample is more likely to be labeled as the target, even assuming a perfect prediction. Importantly, situations like this should arise reliably by chance, that is, situations in which the signal from a target is on the noisier end of the spectrum. The purpose of these contrived schematics, which of course rely on the construct of a noisy sample, is to illustrate that prediction will only help in MOT if it helps one to tell targets and nontargets apart when they are close enough to become confused (see also Bae & Flombaum, 2012). If a good portion of typical MOT errors arises in situations that are genuinely ambiguous for probabilistic reasons, then even accurate predictions will have relatively thin margins to impact.

This perspective is consistent with experimental work demonstrating that speed effects in MOT are largely a function of the number of close encounters between targets and nontargets (Bae & Flombaum, 2012; Franconeri et al., 2010). And it is consistent with our empirical observation that trial-by-trial correlations in performance between participants had a ceiling close to 0.5. All trials should include occasions wherein targets and nontargets are more or less likely to be confused because of their arrangements. But if whether or not they become confused is largely determined by chance—unusually bad signals, unusually well-placed wrong signals, lapses of attention, or blinks and saccades at just the wrong time—then between subject error correlations will have a relatively low ceiling, and even excellent predictions will not save the day.

The second reason that extrapolation may not help has to do with the critical role of velocity, and bearing in particular, for generating accurate predictions. With an inaccurate bearing, a prediction will become increasingly farther from accurate as object speeds (or the lag between encounters) increases. So without accurate representation of bearing, it is not possible to make good predictions.

Figure 14 attempts to schematically make these points, illustrating how a prediction based on noisy inputs can deviate considerably from actual outcomes, and how in the context of MOT, it could lead to correspondence errors.

We further conducted a small set of simulations to make this point. Specifically, we utilized the P-P model from our replications of Experiment 1 (because it is most faithful to a standard Kalman filter). The model, recall, makes predictions by comparing its most recent posterior position estimates, and then projecting the resulting vector forward. The model performed 100 simulations of one trial in which there was only a single, perfect inertia object to track (and no nontargets). Figure 15a plots the average Kalman gain of the model over the simulations, and as a function of temporal sampling rate and spatial precision. The Kalman gain is the relative weighting of a prior and a new observation when inferring a current posterior. The weight on priors was always 10%–17%. The

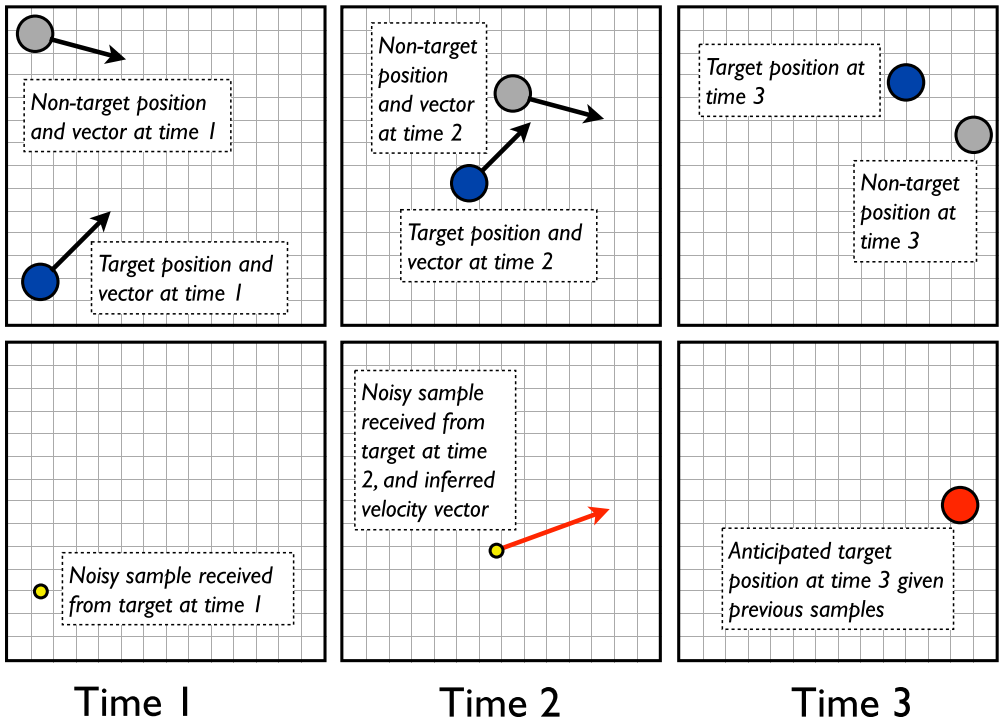

Figure 14. Schematic diagram of target and nontarget motion over three moments, contrasted with potential noisy samples and predictions of an observer. The diagram is meant to illustrate how noisy predictions can deviate from true outcomes, and how they could lead to correspondence errors in the MOT task. Objects in the top row indicate the actual positions of a target (blue) and a nontarget (gray) at three successive moments, in addition to their actual and stable motion vectors. Small yellow circles (bottom row) designate noisy samples received by an observer from the target at Moments 1 and 2. The red arrow at Time 2 designates the most-likely motion vector an observer would infer on the basis of those samples, and the larger red circle at Time 3 designates the best prediction the observer would make, as a result, about the object's expected position at Time 3.

challenge here was not to keep targets and nontargets distinct, just to generate good expectations about target position. But when these expectations relied on bearing estimates derived entirely from noisy observations of position, they were not very accurate. (Note that this model did not include independent velocity noise, and it computed expected covariance following the standard Kalman filter approach).

We then conducted the same set of simulations again, but this time, with a model that we endowed with the correct velocity estimate. It still received noisy observations of position, it still inferred posteriors by combining priors and observations, and it still generated priors by projecting its velocity estimates from its most recent posterior estimate. Only the velocity estimates were perfectly accurate. As shown in Figure

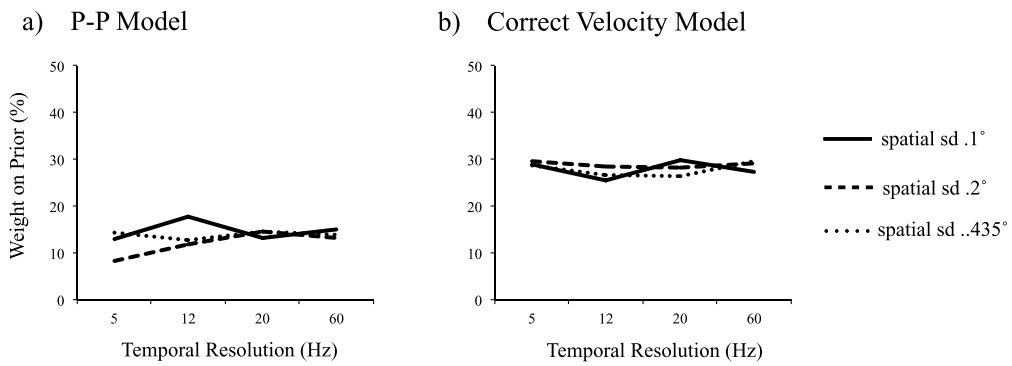

Figure 15. Weights assigned to priors (compared to new observations) when inferring a posterior. Comparison is between the P-P Kalman filter model (a; see Replications of Experiment 1) and a new model (b) that was endowed with accurate knowledge of velocity. Results are from 100 simulations of a single trial including only a single target to track and no nontargets.

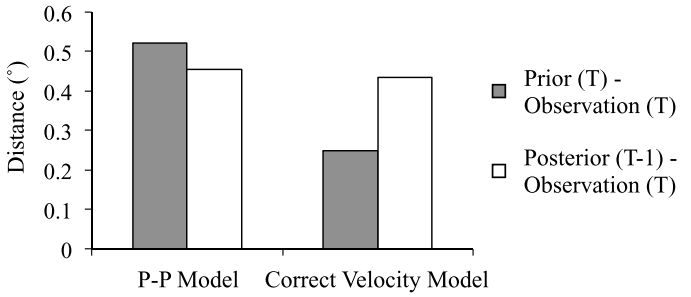

Figure 16. Average distance between prior at time T and observation at time T (gray bars) and between posterior at time T-1 and observation at time T (white bars), for the P-P Kalman filter model and the correct velocity model, respectively. Results are from models with a 0.2° spatial resolution and 12 Hz sampling rate. Patterns were the same for models with other parameters. The correct velocity model made predictions that were closer to new observations than the most recent posterior. But for the P-P model, the opposite was true.

15b, this model ultimately placed weights between 20% and 25% on priors. Moreover, unlike the P-P model, on average this model's priors at a given point in time (T) were closer to the target-linked observation received at that point in time (T) than were the model's most recent posterior estimates at T-1. This is the definition of a good prediction. The model did acquire better expectations about the future than the mere assumption that it should look like the past. But the same was not true of the P-P model (Figure 16).

Making good predictions requires a mechanism for acquiring accurate measurements of velocity. As we discuss below, in the context of the wider literature on extrapolation, it is likely that humans possess such mechanisms, and ones that are independent from mechanisms that estimate position. But they may only be available in the context of tracking a single object, and they may rely on eye movements.

Do human observers extrapolate in MOT?

In an effort to understand mixed behavioral results concerning whether human observers extrapolate in MOT, we investigated how extrapolation could be implemented at the algorithmic level. We discovered that noisy inputs could lead an observer to extrapolate conservatively, expecting objects to be found closer to where they were recently perceived than to where they are headed. And we also discovered that such an observer would not outperform an observer who does not make predictions, under many trajectory types, in an interruption experiment (like that of Fencsik et al., 2007), and especially with more than two or three targets to track. Because this was primarily a computational study, we cannot conclude that human observers make predictions similarly to one of our adjusted prior models. But we can propose an alternative to the binary framing of the question that has prevailed in the literature. The expectation has generally been that observers extrapolate fully, and perhaps accurately, or not at all. The alternative is that they *try* to extrapolate, but in a measured way, ultimately conservatively, and perhaps somewhat inaccurately. This possibility is not only important as a viable alternative. It also stands to reconcile the mixed findings in previous work.

Specifically, making conservative predictions is consistent with work demonstrating that observers are sensitive to and can report object bearing accurately, on average, but with a great deal of noise in many situations (Horowitz & Cohen 2010; Shooner et al., 2010; St. Clair et al., 2010). At the same time, we have shown that using noisy estimates of velocity, especially if derived entirely from estimates of position over time, would not lead to a performance advantage (at least not with more than two or three targets) in MOT, in conditions with entirely reliable object trajectories (perfect inertia). Howe and Holcombe (2012) found no advantage among human observers for such trajectories, compared to more unreliable ones, except when tracking only two objects. Similarly, we showed that making weighted predictions to accommodate noisy estimates would lead an observer to preform similarly in interruption conditions in which objects do or do not proceed along their paths (Experiment 3). Fencsik et al. (2007) and Keane and Pylyshyn (2006) found similar effects among human observers. (Indeed, we utilized their experimental design.) Finally, Howard, Masom, and Holcombe (2011) asked observers to localize tracked targets, and they found that knowledge of object positions lagged behind true positions, but ahead of recent ones. And two studies that have shown attentional biases in the direction of target motion, one via probe detection while tracking (Atsma et al., 2012), and one via localization responses following an abruptly ended trial (Iordanescu, Graboweky, & Suzuki, 2009). By placing low weights on extrapolation—biasing expectations towards recent observations—the adjusted prior models naturally stay ahead of recent positions while lagging behind actual trajectories.

Importantly, our findings can also be reconciled with evidence suggesting that people do extrapolate advantageously in some MOT settings, and the larger literature on accurate extrapolation more generally (e.g., Bennett et al., 2010; Diaz, Cooper, & Hayhoe, 2013; Diaz, Cooper, Rothkopf, & Hayhoe, 2013; Howe & Holcombe, 2012; Spering et al., 2011; Warren et al., 2012). The crucial factor in that work, from our perspective, is that it has been concerned with tasks and situations involving only a single target, and without

nontargets (e.g., Bennett et al., 2010; Diaz, Cooper, & Hayhoe, 2013; Diaz, Cooper, Rothkopf, & Hayhoe, 2013; Spering et al., 2011; Warren et al., 2012). In those settings, the utility of extrapolating is not measured in terms of distinguishing targets and nontargets, but instead, in terms of the effectiveness of actions that require anticipation, such as catching or striking. If an observer has a more accurate sense of where an object just was than where it is going, she should still act on her sense of where it is going. You can't catch a ball where it just was. In MOT, in contrast, extrapolation may be carried out more conservatively (or avoided) because of a small or no marginal utility in terms of task-specific performance.

Moreover, mechanistically, tracking only one object makes it possible to use eye-movements and even smooth pursuit. Many studies of sports (which always involve only one ball) suggest that smooth-pursuit eye movements are beneficial for extrapolation (along with controlled, saccadic eye movements), for example, in baseball, basketball, cricket, squash, volleyball, and table tennis (Bahill & LaRitz, 1984; Lee, 2010; Land & Furneaux, 1997; Land & McLeod, 2000; McKinney, Chajka, & Hayhoe 2008; Ripoll, Bard, & Paillard, 1986). Target-referenced eye movements are generally thought to be the starting point for effective extrapolation. It is also known that smooth pursuit eye movements enhance prediction of visual motion in a laboratory-based task (Spering et al., 2011). And beyond improving visual acuity with respect to a moving target (Bahill & LaRitz, 1984), advantages are thought to accrue from an eye-motion signal generated internally by the occulomotor system. Similarly, saccades to future ball positions—in racquetball, for instance—are known to enhance performance, to arise spontaneously in observers (Diaz, Cooper, & Hayhoe, 2013; Diaz, Cooper, Rothkopf, & Hayhoe, 2013), and to enhance predictions about the timing of contact between a target and other objects (Bennett et al., 2010).

The problem for MOT is that target-directed eye movements cannot be referenced with respect to more than one object at a time. They may be used during multiple object tracking, but their effects should be weakened. This is probably why no evidence for accurate extrapolation has been found in MOT with more than two targets in the display. With two or fewer, however, eye movements may be beneficial. The two studies that have shown evidence of extrapolation did not enforce or monitor eye fixation. Thus there remains a distinct possibility that the presence of effective extrapolation in those studies reflects the benefits of target-directed eye movements (either saccades or smooth pursuit).

One final potential point of contact between our results and the broader literature on eye-gaze mediated extrapolation comes from the simulations we reported in the preceding section, demonstrating that extrapolations become accurate when they utilize independent and accurate knowledge of bearing, even while obtaining noisy observations of position and imperfect posterior estimates. Fixating a target, and potentially tracking it with one's eyes may supply a channel for acquiring independent and relatively speaking, less noisy estimates of bearing, perhaps by taking advantage of both higher and lower level motion systems (Cavanagh, 1992; Lu & Sperling, 2001), and perhaps through estimates derived from internal signals (Spering & Montagnini, 2011). Future work in the context of well-understood and effective extrapolation should investigate effects of increased tracking load, and the precision of velocity estimates.

As we already noted: we are hesitant to conclude that what observers do is weight extrapolations in exactly the way our adjusted prior models do. For the time being, the implication of our results is that extrapolating while weighting by reliability can look very similar to not extrapolating at all. The mixed evidence in the prevailing literature may reflect this state of affairs. Our results also suggest that one of the major barriers to effective extrapolation in MOT is the difficulty of acquiring accurate velocity estimates about multiple objects at once, consistent with highly noisy bearing estimates reported in a previous MOT experiment (Horowitz & Cohen, 2010).

The Kalman filter: A computational framework for multiple object tracking

Twenty-five years of behavioral research has produced a great deal of data, and a good understanding of the basic mental and neural resources involved in carrying out multiple object tracking. For example, we know that attention appears importantly involved in tracking (Scholl, 2009), that working memory plays a role as well (Postle, D'Esposito, & Corkin, 2005; Zhang, Xuan, Fu, & Pylyshyn, 2010), that eye movements are not necessary for tracking (Intriligator & Cavanagh, 2001)—but can facilitate performance if used strategically (Fehd & Seiffert, 2008; Fehd & Seiffert, 2010)—and that the difficulty associated with tracking load can be mapped to neural activity (Drew, Horowitz, & Vogel, 2013).

But computationally, these effects, phenomena, and results have not been incorporated into algorithmically explicit models. Actually, relatively few computational models have been implemented in the first place; we are aware of only one oscillator model (Kazanovich & Borisyuk, 2006) and two Bayesian

models (Ma & Huang, 2009; Vul et al., 2009). These models have been designed to capture basic aspects of performance and, to a certain extent, assumed features of neural implementation. But they have not yet been built to incorporate the full suite of mechanisms known to be involved in multiple object tracking. Another way to put this is that a complete model will ultimately accommodate not only the fact that human performance declines with memory load, but also the role of attention, the role of working memory, the possibility of utilizing eye movements, and so on.

Our models are certainly not complete in this regard. But they reaffirm the potential for using the Kalman filter as a modeling framework. Specifically, we extended previous models in three important ways. First, the models reported here contended with limited sampling rates, between 5 and 20 Hz. Previous models (Ma & Huang, 2009; Vul et al., 2009) sampled the relevant displays at the rate of presentation. As a result, those models were limited in terms of their spatial resolution, but not in terms of their temporal resolution. Temporal resolution is known to be limited in human observers (Landau & Fries, 2012; Latour, 1967; Lichtenstein, 1961; VanRullen & Koch, 2003; White and Harter, 1969), and in the case of MOT specifically several behavioral studies have investigated the impact of limited temporal sampling (Holcombe & Chen, 2012; Holcombe & Chen, 2013). Sampling rate is a parameter with which the Kalman filter framework can incorporate the implications of this work.

Second, the models discussed here tracked only targets. That is, the models only sought correspondences for target objects when they received unlabeled samples from all of the items in the display. In previous work, models either dealt with a different tracking task that does not involve nontargets (Ma & Huang, 2009), or they tracked all items in a display, labeling targets and nontargets categorically (Vul et al., 2009). While this is the optimal approach—benefiting from mutually exclusive correspondence decisions when unlabeled samples are received—it does not comport intuitively with the pervasive opinion that MOT involves selective attention to targets (and the extensive data corroborating this opinion; e.g., see Cavanagh & Alvarez, 2005; Pylyshyn & Annan, 2006; Scholl, 2009). Moreover, it is a challenging assumption for theories that also assume architecturally limited tracking capacity of either a fixed or continuous nature (e.g., Alvarez & Franconeri, 2007; Pylyshyn & Storm, 1988; Vul et al., 2009), because presumably tracking nontargets consumes those resources and human performance declines when the number of targets increases but the total number of objects remains constant. Via selective tracking, the Kalman filter supplies a vehicle for implementing selective attention within a computational framework. We implemented selective attention as tracking of all targets to the exclusion of all nontargets. But one could easily explore variants that would capture features of alternative theories, for instance, by incorporating a fixed capacity limit on the number of items tracked or a serial selection strategy (Drew et al., 2013; Oksama & Hyönä, 2008).

Finally, and most directly relevant to the topic of this investigation, our Kalman filter model implemented extrapolation in importantly different ways than the only other Bayesian model to address MOT. (Other kinds of modeling approaches have not implemented extrapolation directly). Specifically, Vul et al. (2009) implemented prediction by including a global inertia parameter in their model: the model made an assumption about how much object trajectories changed in general, and it applied this general assumption to all objects equally on each frame. In their initial simulations, Vul et al. endowed the model with the true inertia term; to the extent that object trajectories were more or less dependable, the model knew exactly how so in advance. This is importantly different from our model, which made no assumptions about trajectory dependability a priori. Instead, the Kalman filter that we implemented weighted new extrapolations relative to recently inferred conclusions about each individual object's position, in a sense, developing its own theory about the dependability of the trajectories.

These contributions jointly constitute a more general framework for tracking than those described previously. And they intuitively supply the best chances for a model to perform better when making extrapolations. In a series of computational experiments, Vul et al. (2009) found that with prior and fixed expectations about object inertia, models that expected object positions to change slowly performed better than models that expected fast changes. Our results in Computational Experiment 2 converge with their conclusions. The model that placed a fixed weight of 0.5 on extrapolation performed worse than the spatial working memory model. Experiments 1 and 3 demonstrate that even a model that can change is weights will have a difficult time outperforming a simple spatial working memory approach, and also that a model can discover on its own to prefer conservative predictions. A model that can discover on it is own how to make predictions provides a foundation for studying related tracking tasks, for example those which involve circular, orbiting, and other forms of regular—i.e., dependable—but nonlinear motion (Holcombe & Chen, 2012; Tombu & Seiffert, 2011). Future research should investigate the Kalman filters and related applications in these contexts.

Conclusion

In many ways, the main implication of the reported computational experiments is that the challenge in multiple object tracking is not predicting the future, so much as interpreting the present. Telling targets and nontargets apart is difficult because they are perceived noisily. As a result, making predictions and not making them can appear very similar in outward behavior and performance.

Keywords: multiple object tracking, Kalman filter, attention, spatial working memory

Acknowledgments

This work was supported, in part, by a generous seed grant from the Johns Hopkins University Science of Learning Institute.

Commercial relationships: none.
Corresponding author: Jonathan Isaac Flombaum.
Email: flombaum@jhu.edu.
Address: Department of Psychological and Brain Sciences, The Johns Hopkins University, Baltimore, MD, USA.

Footnotes

[1]Experiment 1b utilizes a model that omits this step to ensure that the addition of independent noise does not account for the main results of the study.

[2]Note that in Equation 5, $\tilde{I}^{m}_{t+1}$ denoted what is actually the adjusted prior, because Experiment 1 employed the AP-AP model that only uses that quantity. The AP-P model is unique in these experiments because it uses both an adjusted and unadjusted prior. Both are computed from Equation 5, but for calculating the unadjusted prior $\boldsymbol{\beta}$ is set to zero.

References

Adelson, E. H., & Bergen, J. R. (1986). The extraction of spatiotemporal energy in human and machine vision. *Proceedings of IEEE, Workshop on Visual Motion*, 151–155.

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track?: Evidence for a resource-limited attention tracking mechanism. *Journal of Vision, 7*(13):14, 1–10, http://www.journalofvision.org/content/7/13/14, doi:10.1167/7.13.14. [PubMed] [Article]

Anstis, S. M. (1974). A chart demonstrating variations in acuity with retinal position. *Visual Research, 14,* 589–592.

Atsma, J., Koning, A., & van Lier, R. (2012). Multiple object tracking: Anticipatory attention doesn't "bounce". *Journal of Vision, 12*(13):1, 1–11, http://www.journalofvision.org/content/12/13/1, doi:10.1167/12.13.1. [PubMed] [Article]

Bae, G. Y., & Flombaum, J. I. (2012). Close encounters of the distracting kind: Identifying the cause of visual tracking errors. *Attention, Perception, & Psychophysics, 74*(4), 703–715.

Bahill, A. T., & LaRitz, T. (1984). Why can't batters keep their eyes on the ball? *American Scientist, 72,* 249–253.

BarShalom, Y., Daum, F., & Huang, J. (2009). The probabilistic data association filter. *Control Systems, IEEE, 29*(6), 82–100.

BarShalom, Y., & Fortmann, T. (1988). Tracking and data association. *Academic Press,* 56.

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.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10):7, 1–11, http://www.journalofvision.org/content/9/10/7, doi:10.1167/9.10.7. [PubMed] [Article]

Bennett, S. J., Baures, R., Hecht, H., & Benguigui, N. (2010). Eye movements influence estimation of time-to-contact in prediction motion. *Experimental Brain Research, 206*(4), 399–407.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. *1*, p. 740). New York: Springer.

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

Boykov, Y., & Huttenlocher, D. (2000). Adaptive Bayesian recognition in tracking rigid objects. *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2*, 697–704.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Burr, D., & Thompson, P. (2011). Motion psychophysics: 1985–2010. *Vision Research, 51*(13), 1431–1456.

Cavanagh, P. (1992). Attention-based motion perception. *Science, 257,* 1563–1565.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences, 9*(7), 349–354.

Cox, I. J. (1993). A review of statistical data association techniques for motion correspondence. *International Journal of Computer Vision, 10*(1), 53–66.

Dawson, M. R. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological Review, 98*(4), 569–603. doi:10.1037/0033-295X.98.4.569.

Diaz, G., Cooper, J., & Hayhoe, M. (2013). Memory and prediction in natural gaze control. *Philosophical Transactions of the Royal Society, London B Biological Science, 368*(1628).

Diaz, G. J., Cooper, J., Rothkopf, C. A., & Hayhoe, M. M. (2013). Saccades to future ball location reveal memory-based prediction in a natural interception task. *Journal of Vision*, *13*(1):20, 1–14, http://www.journalofvision.org/content/13/1/20, doi:10.1167/13.1.20. [PubMed] [Article]

Drew, T., Horowitz, T. S., & Vogel, E. K. (2013). Swapping or dropping? Electrophysiological measures of difficulty during multiple object tracking. *Cognition, 126*(2), 213–223. doi:10.1016/j.cognition.2012.10.003.

Drew, T., McCollough, A., Horowitz, T., & Vogel, E. (2009). Attentional enhancement during multiple object tracking. *Psychonomic Bulletin & Review, 16,* 411–417.

Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition, 108*(1), 201–209. doi:10.1016/j.cognition.2007.11.008.

Fehd, H. M., & Seiffert, A. E. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision, 10*(4):19, 1–13, http://www.journalofvision.org/content/10/4/19, doi:10.1167/10.4.19. [PubMed] [Article]

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

Flombaum, J. I., Scholl, B. J., & Pylyshyn, Z. W. (2008). Attentional resources in tracking through occlusion: The high-beams effect. *Cognition, 107,* 904–931.

Franconeri, S., 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.

Franconeri, S. L., Pylyshyn, Z. W., & Scholl, B. J. (2012). A simple proximity heuristic allows tracking of multiple objects through occlusion. *Attention, Perception, & Psychophysics, 74,* 691–702. doi:10.3758/s13414-011-0265-9.

Gegenfurtner, K. R., Xing, D., Scott, B. H., & Hawken, M. J. (2003). A comparison of pursuit eye movement and perceptual performance in speed discrimination. *Journal of Vision, 3*(11):19, 865–876, http://www.journalofvision.org/content/3/11/19, doi:10.1167/3.11.19. [PubMed] [Article]

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience, 14*(7), 926–932.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383*(6598), 334–337.

Holcombe, A., & Chen, W. (2012). Exhausting attentional tracking resources with a single fast-moving object. *Cognition, 123,* 218–228.

Holcombe, A., & Chen, W. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to <3 Hz when tracking three. *Journal of Vision, 13*(1):12, 1–19, http://www.journalofvision.org/content/13/1/12, doi:10.1167/13.1.12. [PubMed] [Article]

Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception and Psychophysics*, *72*(7), 1765–1775.

Howard, C. J., Masom, D., & Holcombe, A. O. (2011). Position representations lag behind targets in multiple object tracking. *Vision Research, 51,* 1907–1919.

Howe, P. D., & Holcombe, A. O. (2012). Motion information is sometimes used as an aid to the visual tracking of objects. *Journal of Vision*, *12*(13):10, 1–10, http://www.journalofvision.org/content/12/13/10, doi:10.1167/12.13.10.

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

Iordanescu, L., Graboweky, 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]

Jacobs, O. L. R. (1993). Introduction to control theory (2nd ed.). Oxford: Oxford University Press.

Kalman, R. E. (1960). A new approach to linear

filtering and prediction problems. *Journal of Basic Engineering, 82*(1), 35–45.

Kazanovich, Y., & Borisyuk, R. (2006). An oscillatory neural model of multiple object tracking. *Neural Computation, 18*(6), 1413–1440.

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

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55,* 271–304.

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *352*, 1231–1239.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nature Neuroscience, 3,* 1340–1345.

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. *Current Biology, 22*(11), 1000–1004.

Latour, P. L. (1967). Evidence of internal clocks in the human operator. *Acta Psychologica, 27,* 341–348.

Lee, S. M. (2010). Does your eye keep on the ball? The strategy of eye movement for volleyball defensive players during spike serve perception. *International Journal of Sports Science, 22,* 128–137.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 20*(7), 1434–1448.

Lichtenstein, M. (1961). Phenomenal simultaneity with irregular timing of components of the visual stimulus. *Perceptual and Motor Skills, 12,* 47–60.

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., ... Name. (2005). Multiple object tracking is based on scene, not retinal, coordinates. *Journal of Experimental Psychology: Human Perception & Performance, 31,* 235–247.

Lu, Z., & Sperling, G. (2001). Three-systems theory of human visual motion perception: Review and update. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, *18*(9), 2331–2370.

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]

Maloney, L. T. (2002). Statistical decision theory and biological vision. In D. Heyer & R. Mausfeld (Eds.), *Perception and the physical world: Psychological and philosophical issues in perception* (pp. 145–189). New York: Wiley.

Mazyar, H., van den Berg, R., & Ma, W. J. (2012). Does precision decrease with set size? *Journal of Vision, 12*(6):10, 1–16, http://www.journalofvision.org/content/12/6/10, doi:10.1167/12.6.10. [PubMed] [Article]

McKinney, T., Chajka, K., & Hayhoe, M. (2008). Proactive gaze control in squash. *Journal of Vision*, *8*(6): 111, http://www.journalofvision.org/content/8/6/111, doi:10.1167/8.6.111. [Abstract]

Murphy, K. P. (2012). Machine learning: A probabilistic perspective. Cambridge, MA: The MIT Press.

Oh, S., Russell, S., & Sastry, S. (2004). Markov chain Monte Carlo data association for general multiple-target tracking problems. *Proc. IEEE Conf. Decision and Control, 1,* 735–742.

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

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

Postle, B. R., D'Esposito, M., & Corkin, S. (2005). Effects of verbal and nonverbal interference on spatial and object working memory. *Memory and Cognition, 33,* 203–212.

Pylyshyn, Z. W. (2006). Some puzzling findings in multiple object tracking (MOT): II. Inhibition of moving nontargets. *Visual Cognition, 14,* 175–198.

Pylyshyn, Z. W., & Annan, V. (2006). Dynamics of target selection in multiple object (MOT). *Spatial Vision, 19*(6), 485–504.

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

Ripoll, H., Bard, C., & Paillard, J. (1986). Stabilization of head and eyes on target as a factor in successful basketball shooting. *Human Movement Science, 5,* 47–58.

Scholl, B. J. (2009). What have we learned about attention from multiple object tracking (and vice versa). In D. Dedrick & L. Trick (Eds.), *Computa-*

tion, Cognition, and Pylyshyn (pp. 49–78). Cambridge, MA: MIT Press.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38,* 259–290.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*(1/2), 159–177.

Sekuler, R., Watamaniuk, S. N., & Blake, R. (2002). Perception of visual motion. *Stevens Handbook of Experimental Psychology, 1.*

Shooner, C., Tripathy, S. P., Bedell, H. E., & Öğmen, H. (2010). High-capacity, transient retention of direction-of-motion information for multiple moving objects. *Journal of Vision*, *10*(6):8, 1–20, http://www.journalofvision.org/content/10/6/8, doi:10.1167/10.6.8. [PubMed] [Article]

Spering, M., & Montagnini, A. (2011). Do we track what we see? Common versus independent processing for motion perception and smooth pursuit eye movements: A review. *Vision Research, 51,* 836–852.

Spering, M., Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Keep your eyes on the ball: Smooth pursuit eye movements enhance prediction of visual motion. *Journal of Neurophysiology, 105,* 1756–1767.

Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review, 102,* 503–532.

St. Clair, R., Huff, M., & Seiffert, A. E. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision, 10*(4):18, 1–13, http://www.journalofvision.org/content/10/4/18, doi:10.1167/10.4.18. [PubMed] [Article]

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience, 9*(4), 578–585.

Tombu, M., & Seiffert, A. (2011). Tracking planets and moons: Mechanisms of object tracking revealed with a new paradigm. *Attention, Perception, & Psychophysics, 73,* 738–750.

Warren, P. A., Graf, E. W., Champion, R. A., & Maloney, L. T. (2012). Visual extrapolation under risk: Human observers estimate and compensate for exogenous uncertainty. *Proceedings of the Royal Society B: Biological Sciences, 279*(1736), 2171–2179.

Welch, G., & Bishop, G. (2006). An introduction to the Kalman filter. *An introduction to the Kalman filter*. University of North Carolina: Chapel Hill, NC.

White, C. T., & Harter, M. R. (1969). Intermittency in reaction time and perception, and evoked response correlates of image quality. *Acta Pcychologica, 30, Attention and Performance, II,* 368–377.

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. *Proceedings of the National Academy of Sciences, USA, 109*(22), 8780–8785.

VanRullen, V., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, *7*(5), 207–213.

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

Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: A survey. *ACM Computing Surveys (CSUR), 38*(4), *13*, 1–45.

Zhang, H., Xuan, Y. M., Fu, X. L., & Pylyshyn, Z. W. (2010). Do objects in working memory compete with objects in perception? *Visual Cognition*, *18*(4), 617–640.