Title

Attention dynamics in multiple object tracking

Permalink

https://escholarship.org/uc/item/4ss6w1r5

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Authors

Srivastava, Nisheeth
Vul, Edward

Publication Date

2015

Peer reviewed

Attention dynamics in multiple object tracking

Nisheeth Srivastava (nsrivastava@ucsd.edu)
Department of Psychology, UC San Diego
La Jolla, CA 92093 USA

Edward Vul (edwardvul@ucsd.edu)
Department of Psychology, UC San Diego
La Jolla, CA 92093 USA

Abstract

We present a computational model of multiple object tracking that makes trial-level predictions about the allocation of visual attention and the resulting performance. This model follows the intuition of allocated resources modulating spatial resolution, but it implements it in a specific way that leads to accurate predictions in multiple task manipulations. Experiments on human subjects, guided by the model's predictions, demonstrate that observers tracking multiple objects use low-level computations of target confusability to adjust the spatial resolution at which the target needs to be tracked, and that the resulting allocation closely approximates the rational solution. Whereas earlier models of multiple object tracking have predicted the big picture relationship between stimulus complexity and response accuracy, our approach makes accurate predictions of both the aggregate effect of target number and velocity and of the variations in difficulty across individual trials and targets arising from the idiosyncratic within-trial interactions of targets and distractors.

Keywords: multiple object tracking; visual cognition; attention; hierarchical Bayesian models

Introduction

A wealth of multiple object tracking (Z. Pylyshyn & Storm, 1988) experiments have documented a rich set of phenomena that have yet to be explained in a unified manner. Many behavioral patterns of object tracking arise when the load (the number of targets to be tracked) is constant: objects are harder to track when they move faster (Alvarez & Franconeri, 2007), are closer together (Franconeri, Lin, Pylyshyn, Fisher, & Enns, 2008), have less reliable identifying features (Makovski & Jiang, 2009), are intermixed among more distracters (Feria, 2012), or must be tracked for longer periods of time. These effects can be explained by a simple ideal observer model solving the tracking correspondence problem under uncertainty (Vul, Frank, Alvarez, & Tenenbaum, 2009), and reflect the information available for the task, rather than the limitations of human cognitive resources. However, such ideal observer models cannot account for a second class of phenomena that arise when varying the number of targets to be tracked: under fixed conditions, participants can only track a small number of objects (Z. Pylyshyn & Storm, 1988), but more can be tracked when they move slower (Alvarez & Franconeri, 2007). These effects provide a glimpse at the limitations and tradeoffs imposed by the cognitive machinery that humans employ to track objects, and theories designed to capture these phenomena postulate either a limited pool of pointers to tracked objects ("slot" models (Luck & Vogel, 1997)) or a finite pool of resources that is spread thin when too many objects must be tracked (Alvarez & Cavanagh, 2004) ("resource" models). To date, these models have been largely descriptive and do not engage with the phenomena that arise from the difficulty of solving the correspondence problem under uncertainty. Our aim in this paper is to unify ideal observer models of object tracking with cognitive resource limitations and allocation to capture both classes of object tracking phenomena, and more generally, to generate insights about how cognitive constraints and low-level uncertainty are coupled in human cognition.

We propose a hierarchical model of human performance on the MOT task that uses recursive Bayesian estimation of position coordinates to model the consequences of perceptual uncertainty, and controls the effective length scales on which these estimators work as a function of the amount of *attention* resource allocated to them by a higher-level controller. Our model follows the phenomenological intuition that humans are able to make finer-grained judgments of relative position when they attend more to a particular location, and that such targeted covert attention is a scarce resource - resolution gain in the attended patch is bought at the expense of coarser-grained resolution elsewhere. We demonstrate that adding a hierarchical controller that assigns spatial resolution to each of the lower-level trackers out of a common pool of attention resource permits us to model MOT phenomena that reflect flexible cognitive resources, e.g. the number of objects that can be tracked, and the profile of most common errors made by subjects. Furthermore, we show that people track different targets with variable spatial precision over time, following our models' predictions of strategic and dynamic allocation of cognitive resources, and that our model distinguishes between "dropping" and "swapping" errors (Drew, Horowitz, & Vogel, 2013) in a novel behavioral characterization.

Overview of flexible-resolution spatial tracking

We work within the framework of rational analysis, wherein models are strongly characterized by their computational goals. Thus, our tracking model is based on a low-level controller that iteratively solves the correspondence problem of observed objects across the movie, and a higher-level controller that allocates a finite resource that improves localization precision. The overall outline of the model we use is shown in graphical form in Figure 1b.

The computational goal of the lower-level controllers is to estimate individual object positions with statistical optimality given the noise/uncertainty of localizing objects in individual frames (Vul et al., 2009). This assumption is entirely in line with existing ideas in Bayesian studies of visual perception (Knill & Richards, 1996) and simply suggests that the low-level controllers behave as ideal Bayesian observers. We supplement this low-level controller with a finite resource the allocation of which modulates the behavior of the ideal observer by changing localization uncertainty.

The finite "resource" in our model is based on the assumption that humans can actively control the spatial resolution/uncertainty of individual percepts. Intuitively, if we want to be able to make finer discriminations of spatial position for an object, we will 'attend' to it more than if we were simply concerned with coarse estimates of its position relative to other objects. We incorporate this intuition into a hierarchical model of inference (as illustrated in Figure 1(b)), where low-level percept-tracking controllers learn the dynamics of individual objects and emits bottom-up signals identifying the likelihood of their tracking labels being lost, and a high-level meta-cognitive module uses these signals to rationally allocate attention to these controllers from a limited global pool, with the constraint that greater attention allocation permits finer spatial resolution. The top-down attention allocation, in turn, determines the uncertainty associated with lower-level position measurements.

The computational goal of the higher-level controller is to greedily reduce correspondence uncertainty, constrained by the total amount of attention resource available. While this is certainly not the only possible goal for metacognitive attention dynamics, constrained greedy optimization is rational in the context of dynamic resource allocation when the underlying demand distribution is non-stationary.

Bayesian object tracking

We model individual object tracking as an ideal Bayesian observer learning a linear dynamical system. Given a state equation,

$$x_{t+1} = Hx_t + \mathcal{N}(0, Q), \tag{1}$$

and a measurement equation,

$$z_t = Cx_t + \mathcal{N}(0, R), \tag{2}$$

where Q is process noise, and R is measurement noise, we implemented a Kalman filter that learns $\{H, C, Q, R\}$ at every time step using expectation-maximization based parameter estimation (Ghahramani & Hinton, 1996). This filter serves as our perceptual ideal Bayesian observer for a single moving object. It takes the two dimensional coordinates as the state observation $\{x, y\}$, predicts the future value of the latent state variable s, and thus generates predictions about future coordinates x, y.

A model completely faithful to the computational requirements of the MOT task would explicitly solve the correspondence problem: which observation should be associated with which filter, as in (Vul et al., 2009). However, to account for human behavior, a simplification is possible: rather than solving the correspondence problem at every time step, we can simply predict the ambiguity of correspondence at each time-step, and swap labels accordingly. This approach permits us to treat particle-filter bindings as known, instead of unknown, by default at every iteration, which greatly reduces the computational complexity of the model.

Rational attention allocation

The top-level attention model assumes that subjects possess a fixed amount of total attention, which can be represented as the scalar integer A. Following indexing-based ideas of object tracking (Z. Pylyshyn, 1989), the model assigns indices p to all objects on the screen; and the amount of attention assigned to each object location at time t is a function $a_t(p)$, where $\sum_p^{\mathcal{P}} a_t(p) = A$.

In every iteration, the model first determines the list of targets for which it will preferentially allocate attention[1] by propagating the list of particles marked as targets (henceforth, the "target list") forward across time.

At every time step, the model evaluates the potential *confusability* of all targets based on the object states the low-level Kalman filters. We approximate the probability of confusion as a logistic sigmoid decreasing with the distance between the target and its nearest distracter, but critically, this distance is scaled by the spatial resolution that each tracker's allocated attention resource permits it to have. These convergent desiderata inform our formal definition of confusability as,

$$c(p) = \exp(-K \times a_{t-1}(p) \times d_t^*(p)), \tag{3}$$

where, K is a scaling parameter, and $d_t^*(p) = \min d_t(p)$, and $d_t(p)$ is the estimated distance at model iteration t between p and all distracters if p is a target or between p and all targets if p is a distracter.

If a target is easily confusable with a distracter and vice versa, the two will swap target/distracter labels with a probability determined by the magnitude of their confusability. Once all possible swaps have been resolved, the particle possesses a new list of targets (which could be the same as the old list if no swaps occurred).

Since the model's current top-level attention allocation to all trackers is based on the previous iteration's distance estimates, it now determines a new attention allocation for each of A 'units' of attention. Each unit is assigned to an object p by sampling an object index from a mixture model: With probability τ an object index is sampled from a distribution obtained by normalizing the confusability of all objects, and with probability $1 - \tau$ an object index is sampled from the targets with probability proportional to their confusability. The parameter τ controls the extent of inhibition of distracter particles. A value of 1 would mean that the model treats targets and distracters equally while dividing up attention. A value of 0 would mean that the model ignores all distracters and attends only to the targets[2].

[1] While earlier indexing-based models of MOT have tried to retain the individual identities of each of the target particles, empirical results (Z. Pylyshyn, 2004) show that humans find it much easier to track target/non-target compared to tracking numbered target identities across the same trial duration. In light of this observation, we use a binary target/non-target identification for all particles.

[2] A very rough grid search in parameter space suggested that a useful value of τ would be 0.4; this is the value we have used throughout our experiments.

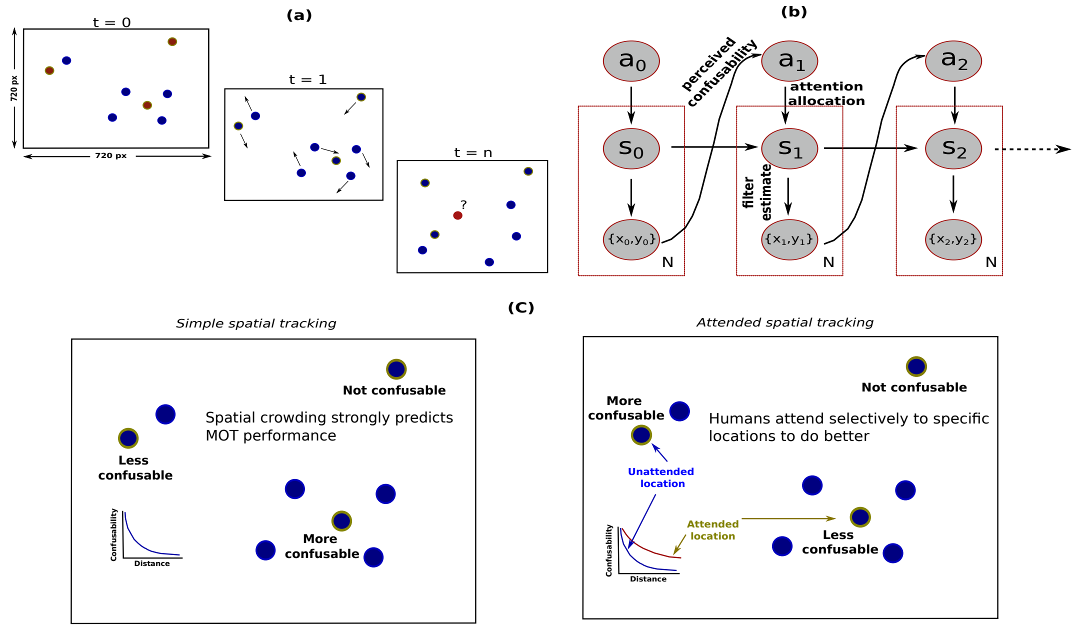

Figure 1: (a) Schematic representation of a typical multiple object tracking (MOT) task. (b) Graphical description of a hierarchical model for tracking N objects simultaneously. The lower level state estimate s is computed using a bank of Kalman filters which predict particle locations with an accuracy that is influenced by their spatial resolution. (c) The scale of spatial resolution for a filter at any time step is determined by the attention allocated to it by the top-level model of attention dynamics. This model obtains information about the confusability of tracking targets from the filter predictions and rationally allocates attention to minimize overall confusability constrained by its attention budget.

Finally, reflecting sensitivity to cognitive processing costs, we assumed the model would possess some degree of inertia to changing its attention allocation, so that,

$$a_t(p) = \lambda a_{t-1}(p) + (1-\lambda)\hat{a}_t(p), \qquad (4)$$

where $\hat{a}_t$ is the allocation computed for the present iteration as above.

Experiment design

The basic MOT task is illustrated in Figure 1a. After initial presentation of 12 objects, k of which were red (targets) and the rest (distracters) blue at the beginning of each trial, the subject presses a key to set them in motion. The objects all turn blue and move on the screen following the dynamics outlined in Equation 5. After 5 seconds, the objects stop moving and one of them, sampled from among the set of targets and set of distracters with equal probability, turns red. The subject must indicate, by pressing ‘y’ or ‘n’, if the red object was red at the beginning of the trial too.

Participants were allowed to practice the task they were to perform until they verified that they understood the objective and were accustomed to the keyboard controls. Practice data was discarded from subsequent analysis in all cases. All experiments were IRB-approved and 50 undergraduate students volunteered as subjects for course credit. Participants viewed the MOT display on a 17-inch PC monitor, and used mouse and keyboard for inputs.

Position and velocity for x and y of each object evolve independently according to an Ornstein-Uhlenbeck process:

$$\begin{aligned} x_t &= x_{t-1} + v_t, \\ v_t &= \lambda v_{t-1} - k x_{t-1} + w_t, \qquad (5) \\ w_t &\sim N(0, \sigma_w) \end{aligned}$$

where x and v are the position and velocity at time t; λ is a friction parameter constrained to be between 0 and 1; k is a spring constant which pulls the particles mildly to the center of the screen; and w_t is random acceleration noise added at each time point which is distributed as a zero-mean Gaussian with standard deviation σ_w.

In two dimensions, this stochastic process describes a randomly moving cloud of objects; the spring constant assures that the objects will not drift off to infinity, and the friction parameter assures that they will not accelerate to infinity. Within the range of parameters we consider, this process converges to a stable distribution of positions and velocities.

Results

More targets become harder to track

We replicate the finding that object tracking becomes harder both with increasing velocity of the particle swarm, and with increasing number of targets (Alvarez & Franconeri, 2007). Unlike the original experiment, where subjects were allowed to adjust their own speed to what they felt was subjectively comfortable, we used a 3-up-1-down staircase, varying the parameter σ_w from Equation 5 in steps of 0.5, thereby objectively measuring a $\sim 79\%$ accuracy threshold for participants.

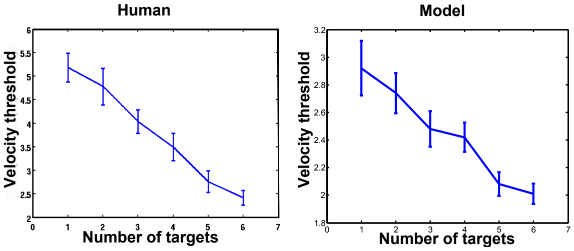

Figure 2: The speed at which observers can maintain a particular accuracy threshold decreases as the number of objects to be tracked increases both for *Left:* 14 human subjects tested using a 3 up-1 down staircase experiment varying object velocity and *Right:* simulations of our model performing the same staircase task.

Results for 14 subjects are shown in Figure 2a and qualitatively match those from (Alvarez & Franconeri, 2007). An *in silico* replication of this experiment using a same-sized agent pool yields identical results, shown in figure 2b, demonstrating that our model replicates aggregate human performance limitations arising out of both increasing velocity and target count. This overall pattern of behavior cannot be captured by a simple ideal observer without a constrained resource.

Predicting individual trial errors

While replicating aggregate predictions forms a useful baseline for assessing model validity, our model provides performance predictions for individual MOT trials, thereby providing a way to examine the limitations that humans face in doing this task at a much finer resolution. Pursuant to our interest in limitations to MOT performance, we are interested more in examining if our model gets the same trials wrong as humans. An algorithm that has difficulty solving the same MOT trials that humans find difficult to solve is more interesting from a scientific standpoint than one that merely captures overall performance trends.

We conducted this analysis in the form of a binary classification study - using multiple (N=11) simulations of model performance on an individual trial as a predictor for human performance. Perfect correlation between human and model predictions would equivalent to perfect binary classification of human errors/non-errors using model predictions. As illustrated in Figure 3, our model outperforms a static spatial tracking model on two comprehensive criteria of classification performance: F-measure and area under the ROC curve. We obtained the ROC for our analysis by varying the threshold count of number of times the model got a trial correct (out of N) for us to label it positive between 1 and 11. The F-score reported is from the middle of the ROC, corresponding to a threshold count of 6.

In a separate experiment, we asked 22 students to perform the MOT task on 150 pre-set trials, with object velocity set at the average of our earlier sample. We then used classification with 20-fold cross-validation to calculate how well the performance of half the subjects on a trial predicted that of the other half, thereby obtaining a theoretical upper bound on classification performance (illustrated in Figure 3a). This upper bound places the extent of improvement in within-trial prediction performance engendered by our model in proper perspective - our model is clearly a considerable improvement over the static case.

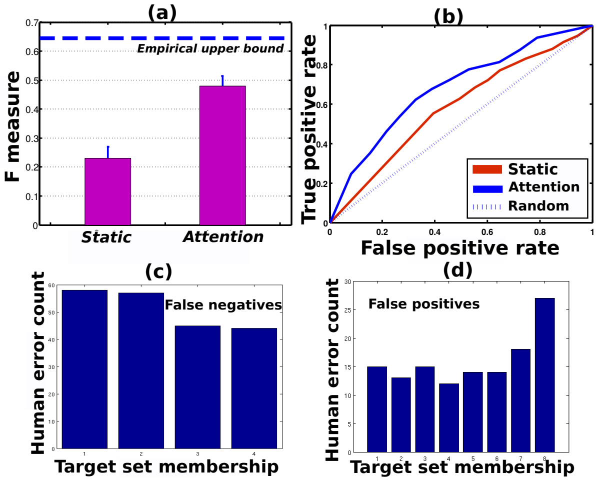

Figure 3: Treating model performance (correct/incorrect) per trial as a binary classifier of human performance shows that attention-gated spatial tracking predicts trial-level human accuracy better than static spatial interference based models, (a) with a considerably greater F-measure and (b) higher area under the ROC curve. Not only does our model make mistakes on the same trials as humans do, it also makes substantially the same mistakes that humans do, both for (c) trials where humans mistakenly identify a distractor is a target and (d) trials where they mistakenly identify a target as a distractor.

Finally, since our model simulates objects movements throughout the trial, it generates predictions for which objects it considers to be targets at the end of each simulation run. By measuring the congruence of these final target sets predicted by the model with the frequency with which humans made mistakes on probed objects, we can get a sense for whether the model *makes the same mistakes* the humans did, not just mistakes on the same trials the humans did. Panels (c)-(d) in Figure 3 present quantitative evidence for congruence between human and model errors. In both figures, the x-axis plots the probability rank with which the model assigns a probed object to the target set, measured across 11 simulation runs; the y-axis counts the number of times the probed object occurred in all error trials across 30 subjects. For false negative trials, where humans, when probed with a target, said that it wasn't, panel (c) shows that the targets that humans mistook for distractors were less likely to be members of the model's target set. For false positive trials, where humans, when probed with a distractor, said it was a target, panel (d) shows that such distractors were more likely to be members of the target set in our simulation runs.

Assessing meta-cognitive attention control

The model we have proposed augments flexibility in spatial resolution to existing Bayesian accounts of multiple object tracking, and our simulation experiments show that it does indeed improve trial-level predictions. Here we further test some more specific predictions of the strategic-allocation MOT model: does precision of tracking follow the predictions our model makes about allocated resources? and does the model distinguish between qualitatively different types of target-identification errors?

Crowded locations are tracked better The key non-trivial prediction of our strategic allocation model of multiple object tracking is that subjects will localize easily confusable objects with greater precision, because they will selectively attend to them more to resolve the possible ambiguity. In contrast, a bottom-up theory of tracking would predict no relationship between crowding and localization error - location errors in such models would reflect either constant perceptual uncertainty or might even increase for more confusable objects due to crowding (Whitney & Levi, 2011).

We directly tested this prediction by making a simple manipulation to the basic design. We interleaved trials probing the identity of one of 4 targets with ones wherein, once the dots stop moving, one of them disappears, and participants were prompted to click on its latest position using a mouse. Participants were instructed to focus on getting the probe trials correct, and respond on the location trials as best they could. This was done to ensure that subjects did not stop attending to targets in order to focus more generally on the entire viewing area to better minimize location errors. We further expect that the randomly interleaved presentation of both types of trials (controlled by a Bernoulli parameter $p = 0.5$) also dissuaded such task switching.

Unlike in the other experiments, where trials were generated *de novo* for each participant, all 29 participants saw the same 150 pre-selected trials in this experiment. These trials were selected to hold the distance between the probed/disappearing particle and its nearest neighbor fixed at five separate values, 30 trials per distance value.

While the data are noisy, the results in Figure 4a, plotting the localization error (in pixels) that subjects make against the category of trial (sorted by distance to nearest neighbor), show a clear trend favoring our hypothesis ($\rho = 0.91, p = 0.03$), and supporting related observations from (Iordanescu, Grabowecky, & Suzuki, 2009). Objects that disappeared in crowded locations were localized with greater precision than objects in less crowded locations. This empirical result supports our work's basic assumption - that rational attention allocation influences MOT performance via flexibility in spatial resolution.

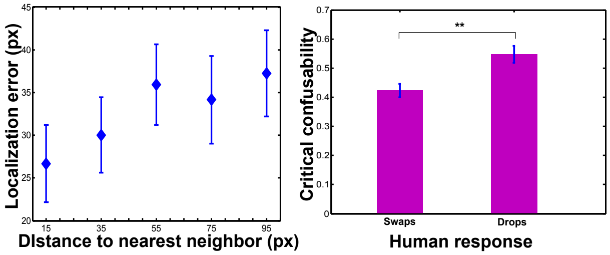

Figure 4: Indirect measurements supporting the role of metacognitive attention dynamics in the MOT task. *Left:* Subjects were more precise in localizing objects that were in crowded locations than those that were more isolated. *Right:* The model's overall confusability load was significantly (p = 0.0017) higher at peak confusability in 'drop' trials than in 'swap' trials as measured behaviorally, suggesting that 'dropping' could be a rational response in such situations. All error bars represent ±1 s.e.m.

Drop/swap predictions People do not always track all of the objects they were asked to, with errors arising from swapped labels between targets and distracters, instead, they sometimes simply drop a target and stop tracking it (Drew et al., 2013). For our purposes, swaps are erroneous identifications of target-distracter labels, as uncovered in the probe trials. Drops are erroneous identifications that participants knew would likely be erroneous before responding because they knew they had dropped a target. Therefore, we can estimate whether a given error was a swap or a drop by asking participants if they were surprised by the error. When an error arises from a participant swapping the probed target for a distracter, or vice versa, they would be surprised when told that they are wrong. Conversely, participants who knew that they had dropped one or more targets would express little surprise at being wrong.

We attempted to elicit precisely this information in a third experiment. The protocol for this study followed the same staircase design used in the first experiment; we collected data only for 14 subjects and with trials involving 4 targets amidst 8 distracters. Also, every time a subject responded incorrectly to the probed particle, they were required to indicate with a keypress whether they were surprised at being wrong before proceeding to the next trial.

Even though our model does not include an explicit mechanism for dropping targets, it is possible to construct hypotheses about situations within trials that would promote drops, and operationalize them testably. In particular, we expect that subjects would drop objects from the target list if their attention resources were over-stretched, causing irreducible confusability among objects. In our model the overall demand for attention might exceed capacity if there are many potentially confusable targets. Therefore, sensitivity to the drop-swap distinction in our model would predict that the cumulative confusability of all the targets would be larger at critical points in the trial for instances where errors would occur due to drops than for instances where errors would occur due to swaps. This prediction is borne out in our data, as shown in Figure 4b, where we show that the critical[3] confusability for

[3] Since errors in MOT happen at critical junctures, and cannot be characterized by statistics averaged across the trial, we have used the largest value of confusability obtained within a trial as our definition

trials labeled as 'drop' errors from our behavioral characterization is consistently higher than for trials labeled as 'swap' errors.

Discussion

Patterns of aggregate behavior in multiple object tracking as a function of the average speed and spacing of objects, the duration of tracking, and the number of distractors can be explained by an ideal observer iteratively solving the correspondence problem (Vul et al., 2009). However, such ideal observers cannot capture the critical effects of tracking load – how many targets must be tracked – indicating that some sort of cognitive resource constraints limit human performance. We combined these two features to model human object tracking performance as Bayesian ideal tracking with a resource constraint, and showed that such an agent exhibits the same tradeoffs between speed and number of targets tracked as people. We go further to show that this limited resource is not allocated to targets according to a fixed, static division, but is instead allocated strategically depending on the prospective costs and benefits of possible allocations.

Strategic, dynamic allocation of cognitive resources can better predict across-trial variation in performance. Furthermore, such strategic meta-cognitive allocation accounts for differences between trials when targets are dropped from consideration, rather than merely mis-associated and swapped with distracters, differences that we were able to behaviorally elicit using a novel experimental manipulation. Finally, the specific combination of our presumed resource (spatial resolution – potentially mediated by attention), and our dynamic allocation policy, predicts that variation in the precision of position estimates for individual targets (localization errors increase for less crowded objects), and we show that this holds for human observers. Together, these results represent proof-of-concept for how we can capture the interaction between bottom-up uncertainty and human cognitive resources using task-sensitive meta-cognitive policies: in multiple object tracking, spatial resolution is allocated to reduce uncertainty for the correspondence problem.

Since our computational model is strongly predicated upon the ability of observers to consistently index objects, it fails in the same directions as indexing theory, e.g. it cannot explain why humans find it easier to differentiate targets from distractors than to identify which target is which (Z. W. Pylyshyn, 2006). Future work could replace our indexing assumption with more realistic models of generating attention foci given visual stimulus (Trommershäuser, Maloney, & Landy, 2003) to accommodate these results.

Acknowledgments

NS and EV are both supported by NSF CPS grant #1239323.

of critical confusability.

References

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

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

Drew, T., Horowitz, T., & Vogel, E. (2013). Swapping or dropping? electrophysiological measures of difficulty during multiple object tracking. *Cognition*, *126*, 213-223.

Feria, C. (2012). The effects of distractors in multiple object tracking are modulated by the similarity of distractor and target features. *Perception*, *41*, 287-304.

Franconeri, S., Lin, J., Pylyshyn, Z., Fisher, B., & Enns, J. (2008). Evidence against a speed limit in multiple object tracking. *Psychonomic Bulletin & Review*, *15*, 802-808.

Ghahramani, Z., & Hinton, G. E. (1996). *Parameter estimation for linear dynamical systems* (Tech. Rep.). Technical Report CRG-TR-96-2, University of Totronto, Dept. of Computer Science.

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision*, *9*(4), 1.

Knill, D., & Richards, W. (1996). *Perception as bayesian inference*. Cambridge: MIT Press.

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

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

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

Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking (mot): I. tracking without keeping track of object identities. *Visual Cognition*, *11*, 801-822.

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

Pylyshyn, Z. W. (2006). Some puzzling findings in multiple object tracking (MOT): II. inhibition of moving nontargets. *Visual Cognition*, *14*(2), 175–198.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003). Statistical decision theory and the selection of rapid, goal-directed movements. *JOSA A*, *20*(7), 1419–1433.

Vul, E., Frank, M., Alvarez, G., & Tenenbaum, J. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. In *Advances in neural information processing systems* (Vol. 22).

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