

The role of kinematic properties in multiple object tracking

Yang Wang

Department of Psychology, University of California, San Diego, San Diego, California, USA



Edward Vul

Department of Psychology, University of California, San Diego, San Diego, California, USA



People commonly track objects moving in complex natural displays and their performance in the multiple object tracking paradigm has been used to study such visual attention for more than three decades. Given the theoretical and practical importance of object tracking, it is critical to understand how people solve the correspondence problem to track objects; however, it remains unclear what information people use to achieve this feat. In particular, although people can track multiple moving objects based on their positions, there is ambiguity about whether people can track objects via higher order kinematic information, such as velocity. We designed a paradigm in which position was rendered uninformative to directly examine whether people could use higher order kinematic information to track multiple objects. We find that people can track via velocity, but not acceleration, even though observers can reliably detect the acceleration cue that they cannot use for tracking. Furthermore, we show a capacity constraint on using higher order kinematic information—people perform worse when required to use velocity to resolve correspondence for multiple object pairs simultaneously. Together, our results suggest that, although people can use higher order kinematic information for object tracking, precise higher order kinematic information is not freely available from the early visual system.

Introduction

Parents track their children running in a playground, drivers track cars weaving through traffic, and spectators track players moving across a football pitch. In all of these cases people must dynamically solve the correspondence problem (Marr & Poggio, 1976; Ullman, 1979) to determine which visual elements at a given point in time correspond with objects they have been tracking (Vul et al., 2009). Because object identities and their persistence over time are established via the correspondence problem, identifying which features are used to solve the correspondence problem in object tracking is of paramount importance to

explaining how we establish a coherent worldview over time. However, basic questions about which kinematic features are used to solve the correspondence problem in object tracking remain unsettled.

In the multiple object tracking paradigm (Pylyshyn & Storm, 1988), observers are first shown a display consisting of a small set of identical objects and a subset of them are randomly selected and briefly highlighted as targets. The objects then move across the display while observers attempt to keep track of the targets. After some period of time, the objects stop and observers are asked to identify the targets. This task captures the key elements and cognitive processes underlying naturalistic tracking in driving or team sports (Meyerhoff et al., 2017).

Several aspects of object dynamics influence tracking performance. Most important, it is harder to track objects when targets and distractors have close encounters that cause spatial interference (Franconeri, Jonathan, & Scimeca, 2010) or provide opportunities for target–distractor confusion (Vul et al., 2009). Moreover, tracking performance declines as the speed of the objects increases (e.g., Liu et al., 2005; Fencsik, Urrea, Place, Wolfe, & Horowitz, 2006; Alvarez & Franconeri, 2007; Bettencourt & Somer, 2009; Huff, Papanmeier, Jahn, & Hesse, 2010; Tombu & Seiffert, 2011). Part of the decline in performance with increasing speed is attributable to an increased rate of close encounters when objects move faster (Franconeri, Lin, Pylyshyn, Fisher, & Enns, 2008; Franconeri, Jonathan, & Scimeca, 2010). However, there is a further effect of speed itself, such that faster speeds impede tracking performance independent of the rate of close encounters between targets and distractors (Tombu & Seiffert, 2011; Ferial, 2013). What processes give rise to these tracking phenomena?

Early research on multiple object tracking advanced a theory of visual indexes (e.g., Pylyshyn, 1989, 2001, 2003)—early visual mechanisms that provide an association between a target label and position coordinates. According to the visual index theory, little information is extracted from each tracked

Citation: Wang, Y., & Vul, E. (2021). The role of kinematic properties in multiple object tracking. *Journal of Vision*, 21(3):22, 1–15, <https://doi.org/10.1167/jov.21.3.22>.



stimulus—indexes mark the current spatial position of each stimulus and no further information is used to differentiate targets from distractors. Subsequent work has shown that, despite the significant role of location information during tracking, other visual features can also be used to track objects, such as color, spatial frequency (Blaser, Pylyshyn, & Holcombe, 2000), or image identity (Makovski & Jiang, 2009). However, the extent to which people can track objects via velocity, let alone higher order derivatives of position, remains controversial. Velocity can be separated into two components: speed and direction. There is some evidence that stable object speeds make tracking easier, although the reasons may be due to variable speeds' effects on rates of spatial interference (Meyerhoff, Papanmeier, Jahn, & Huff, 2016). The use of motion direction has been mired in more controversy. It seems reasonable to expect that velocity direction would be helpful in solving the correspondence problem because people expect smooth and uniform motion trajectories (Ramachandran & Anstis, 1983; Anstis & Ramachandran, 1987; Watamaniuk & McKee, 1995; Watamaniuk, McKee, & Grzywacz, 1995; Verghese, Watamaniuk, McKee, & Grzywacz, 1999). However, studies examining whether velocity direction is used to track objects have yielded mixed results. On the one hand, observers can report the direction of motion when tracking multiple targets (Horowitz & Cohen, 2010; Shooner, Tripathy, Bedell, & Ogmen, 2010), suggesting that velocity direction is at least available for motion extrapolation to aid tracking. Furthermore, people use velocity direction to estimate extrapolated object positions (e.g., Fencsik et al., 2007; Iordanescu, et al., 2009), and can track continuously visible objects better when their velocity direction is more stable (Howe & Holcombe, 2012; Luu & Howe, 2015). On the other hand, target recovery studies—wherein a target disappears part way through tracking and reappears either at its original or its extrapolated position—find that people do not extrapolate position using velocity and instead are more likely to solve the correspondence problem in favor of the original observed position (e.g., Keane & Pylyshyn, 2006; Franconeri et al., 2012). Moreover, extrapolation using velocity may not even be beneficial (Zhong et al., 2014). On the whole, the existing literature is in conflict about the use of velocity direction to solve the correspondence problem during multiple object tracking.

The incongruous results about the use of velocity direction to track objects might be caused by a violation of a fundamental assumption in the prior literature: the existing work has assumed that, if velocity is used to track, it will be used to completely extrapolate position. Insofar as this is the case, then velocity will effectively modify the expected position of an invisible object by an amount that is equal to the duration of disappearance times the last seen velocity. However,

velocity may be used to solve the correspondence problem in slightly different ways that do not amount to complete extrapolation. Position may be only partially extrapolated in the velocity direction (i.e., sublinear extrapolation), in which case the extent of the extrapolation will determine whether the original or the fully extrapolated position ought to be a better resolution of the correspondence problem in the target recovery paradigm. Another possibility is that velocity is used directly as a noisy feature like color, spatial frequency, or position; under this account, the original position is preferred to an extrapolated position, even though velocity is used to solve correspondence if position is ambiguous. In short, target recovery paradigms that compare the original position to a fully extrapolated position may be pitting velocity against position, rather than providing a pure test of whether velocity is used at all to solve the correspondence problem.

In the current study, we explicitly test whether people can use instantaneous velocity direction to track multiple objects when positional information is completely ambiguous, and only velocity information may resolve correspondence. Furthermore, we apply the same design logic to ask whether average acceleration direction can be used for tracking when position and velocity information are both uninformative. In [Experiment 1](#), we show that, when velocity is necessary to track objects, people do use it; however, under the same conditions, they fail to use acceleration. In [Experiment 2](#), we replicate the results in [Experiment 1](#) and further confirm that position is more useful for solving the correspondence problem than velocity with our presentation parameters. [Experiment 3](#), we show that the failure to use acceleration for tracking is not owing to insufficient low-level information about acceleration. In [Experiment 4](#), we show that the extent to which velocity, but not acceleration, is used varies with “kinematic load”—the number of object pairs require velocity information simultaneously for tracking. Finally, we provide several accounts that can explain our success, as well as the previously reported failures, of using velocity direction to solve the correspondence problem in object tracking. Together, these results indicate a hierarchy of kinematic information for tracking—position is most precise and most useful, but velocity direction can still be used; average acceleration direction, in contrast, seems to be largely not used for multiple object tracking.

Experiment 1

Can velocity or acceleration be used for tracking when position is completely ambiguous? To answer this question in as pure a paradigm as possible, we designed

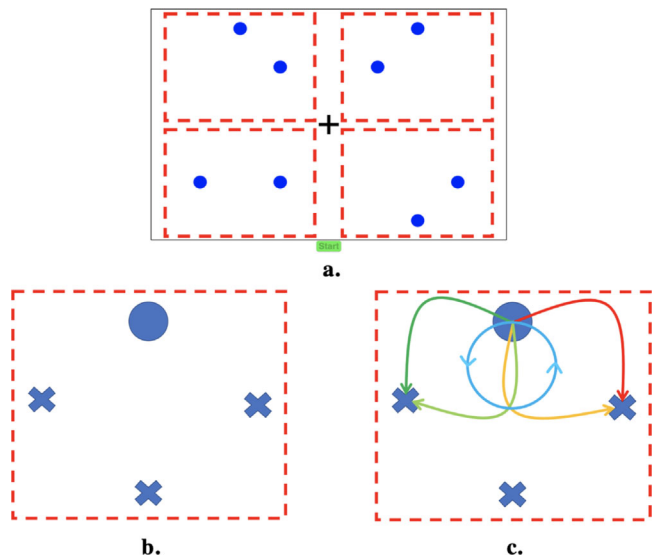


Figure 1. The general set-up of the [Experiment 1](#). (a) Eight objects were displayed as four pairs and each pair had one target and one distractor. (b) Within each quadrant, the target and the distractor were randomly initialized at one of the four fixed virtual positions. (c) There are five possible transitions for objects at each virtual position: four parabolic paths and a circular path.

trajectories in which tracking using only position would result in chance performance, and tracking via velocity direction and acceleration was the only way to track correctly. We treat velocity as it is treated in physics: a signed displacement vector. This means that velocity vectors with the same magnitude, but different directions, are very different. This contrasts with some uses of the term “velocity” to refer to the object’s speed, which would correspond with the magnitude, but not the direction, of the velocity vector.

Method

Stimuli

For each trial, eight objects were displayed as four distractor–target pairs in four quadrants ([Figure 1a](#)). Within each quadrant, the target and the distractor were initialized randomly at one of the four predetermined positions ([Figure 1b](#)). The movement of the objects proceeds in discrete transitions between these four predetermined positions (subject to the constraint that the two tracked objects end in different positions). The separation of pairs of objects into quadrants provides an unambiguous between-quadrant positional cue so that targets are only plausibly confused with their paired distractor. This strategy allows us to control the paths and object interactions and maintain chance performance at 0.5. The trajectory between two adjacent positions was a parabolic path, with

the exception of a circular transition to the same location ([Figure 1c](#)). The paths of all the objects are precalculated and cached to minimize the need for computations during the animation. In each trial, participants were asked to track the targets through 8 transitions, with each transition taking 1 second (for 8 seconds of tracking total).

To accommodate participants’ various displays and interfaces, the size of the canvas was scaled to the size of the participants’ window, and stimuli were scaled to the size of the canvas. The vertical and horizontal offsets of quadrant centers from central fixation was one-quarter of the canvas height and width, respectively. All other aspects of the display were scaled to the canvas height, and we will describe the display in units of object diameters. Objects had a diameter of 1/18th of the canvas height. The four possible object positions in each quadrant were arranged in a square, each offset from the quadrant center by three-object diameters (one-sixth of the canvas height). Transitions between adjacent object positions followed a parabola, with a total trajectory length equal to six-object diameters on-third of the canvas height). Because each transition took 1 second, objects moved at an average speed of 6 of their own diameters per second.

There were three different types of interactive transitions between the target and distractor of each pair: position, velocity, and acceleration. During position transitions ([Figure 2a](#)), the target and distractor were always spatially separated. During velocity transitions ([Figure 2b](#)), the two objects started from diagonally opposite positions and their trajectories intersected at the center of the quadrant; at the intersection, the target and distractor have exactly the same position, but opposite instantaneous velocity directions and opposite average accelerations. Consequently, during velocity transitions, solving correspondence using position alone would lead to chance performance, but using velocity and/or acceleration information would allow accurate tracking. Finally, during acceleration transitions ([Figure 2c](#)), two objects started in adjacent positions and intersected at the center of the quadrant; at the intersection, the target and distractor had identical positions and velocity directions, but opposite average acceleration directions. Thus, during acceleration transitions, using position and velocity alone would yield chance performance, but using acceleration information would allow accurate tracking. Both velocity and acceleration transitions are referred to as critical transitions.

For each trial, all four distractor–target pairs underwent eight simultaneous transitions at a rate of one transition per second. There were three different trial conditions: position, velocity, and acceleration. In the position condition ([Figure 3a](#)), all eight transitions for the four pairs of objects were position transitions; thus, position information was sufficient throughout

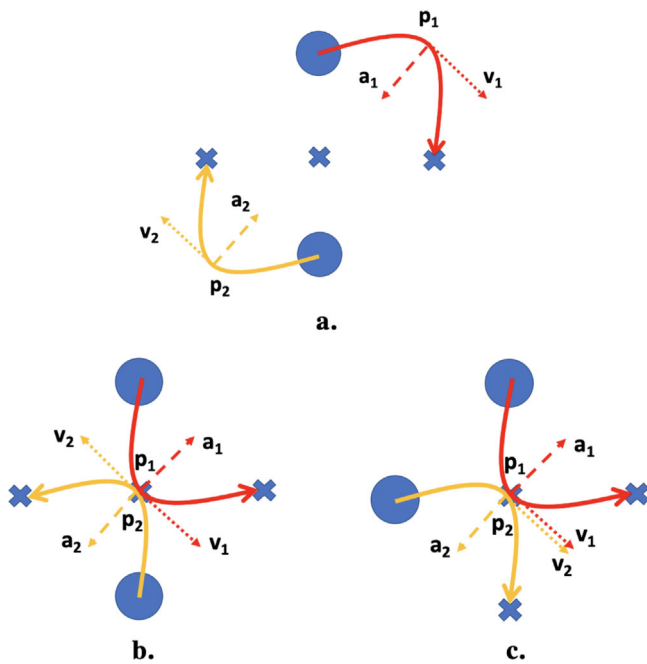


Figure 2. The three types of transitions. (a) Position transition. Objects are well-separated throughout tracking with possibly distinctive velocity and average acceleration directions. Here we show one of three possible position transitions with these specific starting and end positions. From a pair of diagonally opposite starting positions, a total of 17 position transitions are possible. From a pair of adjacent starting positions, a total of 15 position transitions are possible. (b) Velocity transition. The pair of objects intersect at the center, making the position ambiguous and instantaneous velocity direction and average acceleration direction informative for tracking. (c) Acceleration transition. The pair of objects intersect at the center making the position and velocity direction both ambiguous and acceleration direction is informative for tracking.

tracking. In the velocity condition (Figure 3b), one of the transitions was selected randomly to be a velocity transition, whereas all seven of the other transitions were position transitions. In Experiment 1, all four distractor–target pairs underwent a velocity transition at the same time. The velocity transition could not occur at either the first or second transition, and it was randomly placed into one of the third through eighth transitions. The transition before the velocity transition was constrained to end with the target and distractor in opposite positions, to allow a velocity transition to occur next. The insertion of velocity transition means that position was insufficient to accurately track through a whole velocity condition trial, and accurate performance required using velocity to track through the velocity transition. Finally, in the acceleration condition (Figure 3c), seven transitions were position conditions and one of the third through eighth transitions was randomly selected to serve as the acceleration transition for all four pairs. The transition

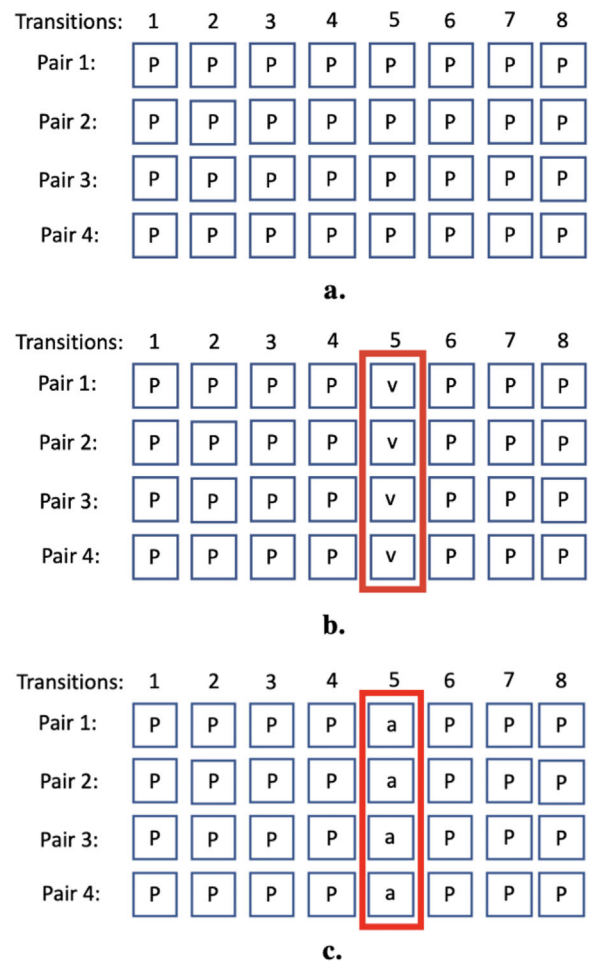


Figure 3. Transition diagrams for three conditions. (a) Position condition. All four pairs undergo eight position transitions. (b) Velocity condition. At a random transition, all four pairs synchronously go through a velocity transition and the other seven transitions are all position transitions. (c) Acceleration condition. At a random transition, all four pairs simultaneously go through an acceleration transition and the remaining seven transitions are all position transitions.

before the acceleration transition was constrained to end with the target and distractor in adjacent positions. Thus, to track accurately through an acceleration trial, participants must accurately track through an acceleration transition, which required relying on acceleration to solve the correspondence problem.

Participants

Fifty-three undergraduates from the University of California San Diego (UCSD) completed the study online, via their web browser, for course credit.

Procedure

Participants completed 15 trials for each of the three types of conditions—position, velocity, and

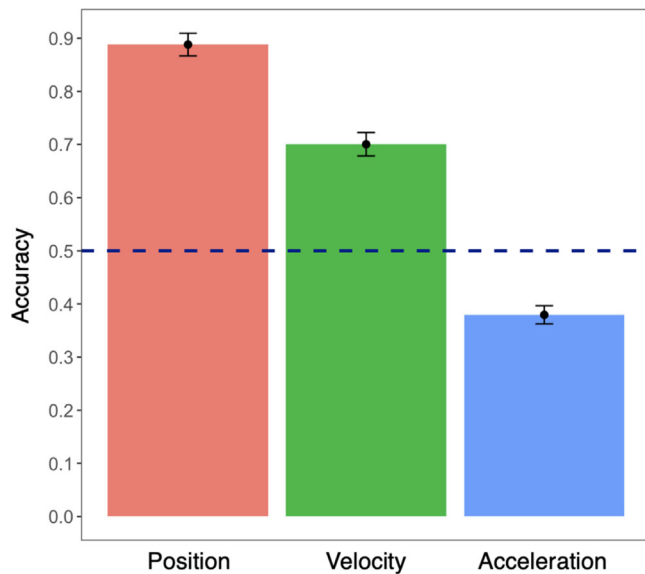


Figure 4. Tracking accuracy for three trial conditions in Experiment 1. Accuracies (*y*-axis) are plotted as a function of conditions (*x*-axis). Error bars show the between-observer standard errors, and the dashed line indicates chance accuracy. Observers can track well above chance in position and velocity conditions but significantly below chance in the acceleration condition.

acceleration—and the order of trial conditions was randomized. After reading the instructions and starting the trial, eight objects appeared on the display in four pairs. One target in each pair was highlighted for three seconds in black circles, and then they moved for the next 8 seconds (as described in the Stimuli section). Participants were asked to fixate at the fixation point throughout tracking. After the objects completed the eighth transition, participants were asked to click the target in each pair. After submitting their response, participants received feedback wherein correctly identified targets were highlighted with green circles and incorrect ones were highlighted with red circles.

Results and discussion

If observers can use direction of velocity to track objects, then they should be able to consistently solve the correspondence problem in the velocity condition. If they only use position, then they should be at chance (50%), because position alone is completely ambiguous in velocity trials. Figure 4 shows that people performed well above chance in the velocity condition, $M = 71.5\%$, $t(52) = 12.826$, $p < 0.001$, indicating that observers can and do use velocity information and, possibly, acceleration information to track multiple targets.

We also evaluated performance in the acceleration condition in which both position and velocity direction were ambiguous at the path intersection, but average

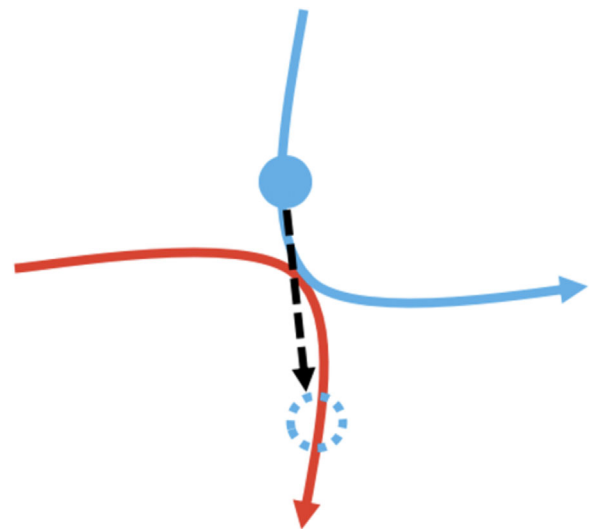


Figure 5. The systematically below-chance performance (swapping) in the acceleration condition is consistent with people using lagged (delayed) velocity while disregarding acceleration information. If observers disregard acceleration information, they should expect the lagged velocity (black arrow) to continue through the ambiguous intersection, and thus might expect the target (blue) to end up on the distracter's trajectory (red curve). Such a reliance on lagged velocity, and a disregard of average acceleration direction, would yield a reliable misidentification of the distracter as the target.

acceleration direction could establish correspondence. Accuracy in the acceleration condition ($M = 38.2\%$) was significantly below chance, $t(52) = -8.994$, $p < 0.001$, indicating that people not only fail to use acceleration direction to track targets, but also tend to consistently swap the targets and distractors. This result indicates that velocity, rather than acceleration, is used in the velocity condition. Although instantaneous velocity is ambiguous at the path intersection in the acceleration condition, lagged (or delayed) velocity is systematically misleading (Figure 5). If people rely on slightly delayed velocity direction to track objects and ignore acceleration information, then we would expect systematically below chance performance in the acceleration condition. It is possible that, when people sample positions less frequently at higher loads, position estimates lag behind the physical stimulus (Howard & Holcombe, 2008; Howard et al., 2011), and extrapolating from these lagged positions yields the apparent use of lagged velocity direction and systematic misassociation after object intersections.

In addition, accuracy in the position condition ($M = 88.3\%$) was significantly higher than the velocity condition ($M = 71.5\%$) with ambiguous positional information, $t(52) = 12.186$, $p < 0.001$, suggesting that position is better than velocity for solving the correspondence problem at the resolutions and separations we tested. However, the position condition

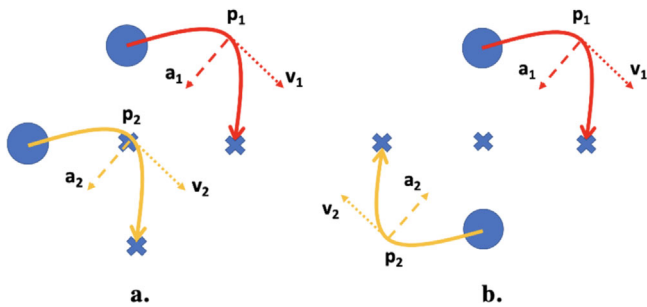


Figure 6. Two different types of position conditions. (a) An example of transition in the “position only” condition. The target and distractor within each quadrant differ only in position while having matched velocity and acceleration. (b) An example of the “mixed position” transition from [Experiment 1](#). The target and distractor are always spatially separated with possibly distinctive velocity and average acceleration directions.

we ran was a mixture of heterogeneous transitions. Although all position transitions included spatial separation, some also had unique velocities and unique accelerations for the target and distractor. Consequently, the superior performance in the position condition might arise from observers having access to more kinematic cues to track in the position condition (position, velocity, and average acceleration) than in the velocity condition (velocity and acceleration). We address this issue in [Experiment 2](#).

Experiment 2

Was accuracy higher in the position condition than the velocity condition because position provided more information than velocity did, or because some position transitions provided three cues for tracking (position, velocity, and acceleration)? To test whether position is more useful for tracking than velocity, we must compare the velocity condition to a position only condition. Thus, in [Experiment 2](#), we introduced a position only transition ([Figure 6a](#)), wherein the target and the distractor within each quadrant differed only in position while having identical velocity and acceleration information.

Method

[Experiment 2](#) is a replication of [Experiment 1](#) except an additional position-only condition. Forty-four undergraduates from UCSD completed the study online, via their web browser, for course credit.

Results and discussion

Insofar as the position condition outperforms the velocity condition owing to the reliance on velocity and

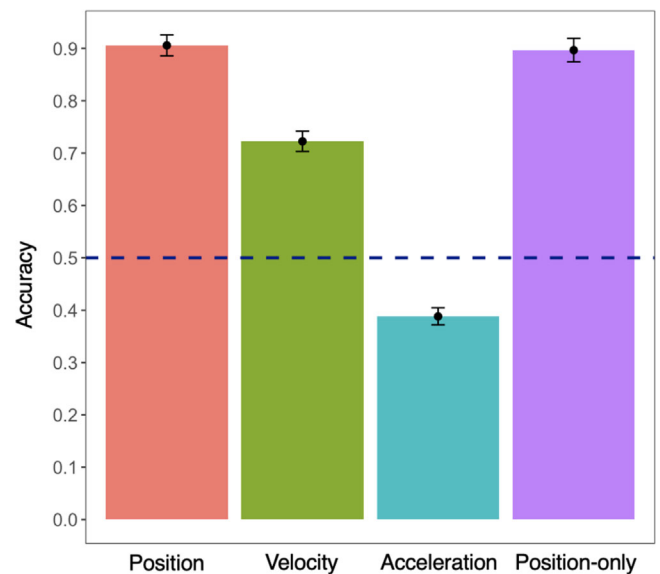


Figure 7. Tracking accuracy in a replication of [Experiment 1](#) with “position only” condition. Accuracies (y-axis) are plotted as a function of conditions (x-axis). Error bars show the between-observer standard errors, and the dashed line indicates chance accuracy. The tracking accuracy of the “position only” condition is comparable to that of the “mixed position” condition from [Experiment 1](#) and higher than that of the velocity condition.

acceleration in the position trials, we would expect that the position-only condition would perform similarly to velocity, and worse than the mixed position condition. We found the accuracy of this position-only condition ($M = 0.897$) was no different from the original position condition, $M = 0.906$, $t(43) = 0.853$, $p = 0.398$, and was significantly higher than the velocity condition, $M = 0.722$, $t(43) = 7.87$, $p < 0.001$ ([Figure 7](#)).

The matched performance in position-only condition and mixed position condition indicates that people seem not to use velocity and acceleration information at all when position information is unambiguous ([Zhong et al., 2014](#)). Thus, the mixed position condition is effectively equivalent to position only condition, because the unambiguous position information completely dominates tracking and velocity and acceleration provide no further assistance. These results confirm that at the resolutions and separations we tested, position is better for solving the correspondence problem than velocity direction.

Experiment 3

[Experiment 1](#) shows that people do not use acceleration to solve the correspondence problem. One possible explanation of this failure to use direction of average acceleration to disambiguate targets from

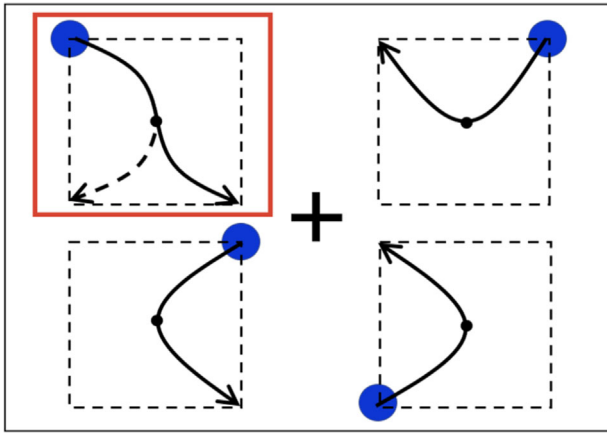


Figure 8. The display of [Experiment 3](#). There are four objects and each is randomly initialized at one of the four virtual locations in each quadrant. One of the objects is randomly selected (indicated by a red rectangle) as the “critical object” that has a distinctive path at the halfway of its transition. The other three objects complete their smooth parabolic transitions. The figure shows only one of the different possible paths.

distractors is that people are simply insensitive to acceleration—acceleration is simply not perceived and thus could not possibly be used to solve the correspondence problem. In [Experiment 3](#), we tested this explanation via a trajectory-change detection task. Observers were asked to report which of four objects had undergone a distinctive path change halfway through its parabolic path. If people were sensitive to particular motion information (i.e., position, velocity direction, and acceleration direction), they should be able to detect the change of motion information among the others that have no changes.

Method

Stimuli, procedure, and participants

The basic display structure of [Experiment 3](#) is very similar to that of [Experiment 1](#). For each trial, there were four objects and each was displayed at one of the quadrants ([Figure 8](#)). One of the four objects, later referred to as a “critical object,” was randomly selected to undergo a path change at the halfway point of its smooth parabolic path. The other three objects completed their smooth parabolic paths.

There were eight possible transitions for the critical object, which we grouped based on the four types of path change the object underwent at the halfway point ([Figure 9](#)).

Position change: The object was instantaneously translated onto a new trajectory, and continued on a path with an otherwise unperturbed velocity and acceleration profile.

A 180° velocity direction change: The sign of the instantaneous velocity vector at the center of the quadrant was flipped. There were two variations depending on what happened to the average acceleration relative to the translation of the projectile motion. If the acceleration vector remained unchanged, the object simply doubled back on its trajectory; if the acceleration vector also changed, the object moved to a new spot.

A 90° velocity change: The velocity vector rotated by 90° at the halfway point. The acceleration vector also changed by 90°. [Figure 9](#) shows the four rotations that were possible in this condition.

Acceleration change: Position and velocity were unperturbed at the halfway point, but acceleration vector was inverted (rotated 180°).

It is worth noting that, although we can describe the distal physical stimulus as undergoing an instantaneous change in velocity with no sustained change in acceleration, it is not the case that this needs to be an accurate description of the proximal percept of the stimulus. For instance, the instant change of velocity can be described as an instant of infinite acceleration. Insofar as the visual system smooths this into a sustained acceleration estimate, which means that the conditions that we describe as including only a sustained change in velocity actually have a sustained acceleration change as well. For our purposes, we adopt the distal description because we are interested in characterizing what kinds of sustained changes in the physical stimulus the visual system can detect.

A trial began with four objects displayed in their starting positions for 2 seconds. This was followed by a 1-second transition during which three objects transitioned via unperturbed parabolic arcs, and the critical object underwent one of the aforementioned trajectory changes at the halfway point. After the 1-second transition, objects came to rest and observers were asked to report which object experienced a trajectory change by selecting it with mouse click. Participants received feedback about the correct choice. Forty-eight undergraduates from UCSD completed the 25-minute study of 200 trials (25 trials for each condition) online for course credit.

Results and discussion

To examine whether observers are sensitive to kinematic information including the acceleration, we evaluated participants’ performance in detecting different kinds of trajectory changes. Particularly, if observers were sensitive to acceleration information, they should be able to detect which object experiences an acceleration change.

[Figure 10](#) shows the accuracy for all eight trajectory changes we tested. Observers were able to identify the

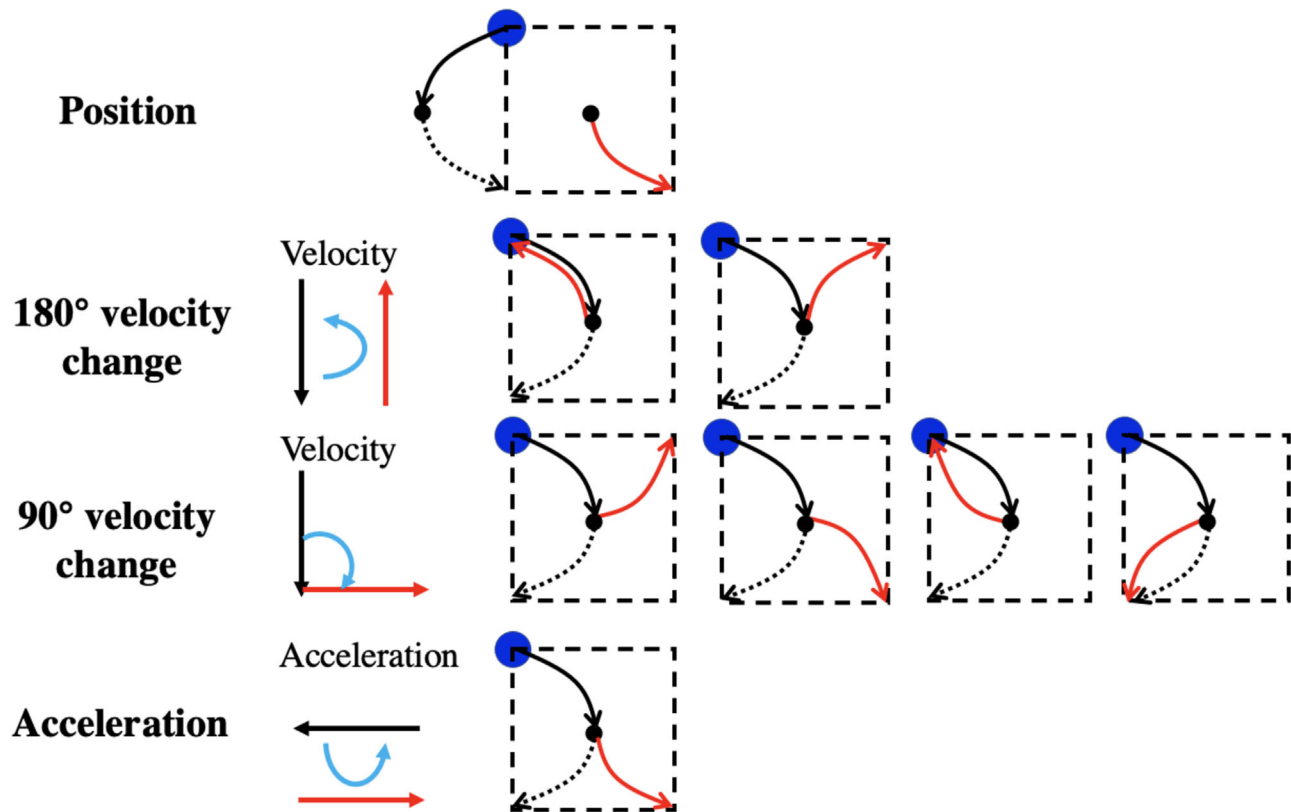


Figure 9. Transition diagrams for the critical object. The black solid arrow indicates the path of the first half. The black dotted arrow indicates the smooth parabolic path of the second half. The red arrow indicates the actual path determined by changes of kinematic properties. Position condition: The object has a sudden positional shift at the halfway point of the normal parabolic path. A 180° velocity change condition: The critical object has a change of velocity direction by 180° at the halfway point of its parabolic path. The 90° velocity change condition: The critical object has a change of velocity direction by 90° at the halfway point of its parabolic path. Acceleration condition: The critical object keeps the same velocity direction but changes its acceleration direction by 180° at the halfway point of its parabolic path.

critical objects significantly above chance (25%) in all conditions except, most notably, the acceleration condition ($M = 37.1\%$, $t(47) = 7.215$, $p < 0.001$). This finding suggests that observers are sensitive to acceleration direction, even though they do not seem to use it to solve the correspondence problem.

Although people were sensitive to changes in velocity direction ($M = 57.4\%$, $t(47) = 12.791$, $p < 0.001$), they were, on average, more sensitive to 180° changes ($M = 63.1\%$) than 90° changes (54.6%), $t(47) = 8.711$, $p < 0.001$. Furthermore, observers were more sensitive to position information ($M = 89.8\%$) than to velocity information, $M = 57.4\%$, $t(47) = 14.778$, $p < 0.001$, than to acceleration information, $M = 37.1\%$, $t(47) = 9.992$, $p < 0.001$.

Experiment 3 tested how sensitive observers were to different kinds of motion information. In general, the results illustrate that people are sensitive to the kinematic properties (i.e., position, velocity direction and acceleration direction), but with lower sensitivity to higher order derivatives of position. However, even for acceleration, people are reliably above chance,

indicating that people are sensitive to the kinds of acceleration information that they could have used to disambiguate targets and distractors in Experiment 1. Thus, Experiment 3 shows that the failure to use the acceleration direction to disambiguate targets and distractors in Experiment 1 cannot be attributed to acceleration information being unavailable at the early sensory level, and must arise from a failure to use this information for tracking. Although people are greater than chance in detecting acceleration, they are not very much above chance, leaving open the possibility that their performance is achieved by paying attention to one quadrant at a time, rather than detecting acceleration for all quadrants in parallel.

Experiment 4

Howe and Holcombe (2012) found that accuracy was higher in the predictable condition than in the unpredictable condition when observers tracked

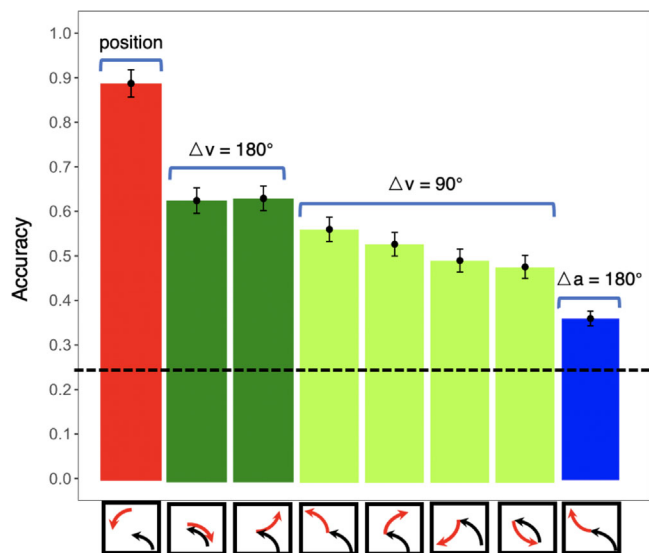


Figure 10. Accuracies for the detection task. The x-axis represents different cases and the y-axis represents accuracies. Error bars show the between-observer standard errors. The dark green bars are the cases of 180° velocity change. The light green bars are the cases of 90° velocity change. The blue bar is the case of 180° acceleration change. In general, observers are sensitive to the kinematic properties including average acceleration direction.

two targets, but not when they tracked four targets, suggesting that the ability to use kinematic information like velocity was modulated by tracking load. Thus, in [Experiment 1](#), simultaneously predicting the future positions to disambiguate the targets and the distractors using kinematic properties (i.e., velocity and acceleration) of all four pairs might impose special challenges for observers. In our study, instead of varying the tracking load which was the number of objects to be tracked, we varied the kinematic load, which controlled the number of pairs that underwent critical transitions simultaneously. For the manipulation of the kinematic load to have any effect, observers must allocate attentional resources inhomogeneously, and dynamically, across targets. In other words, observers need to know which targets are heading into trouble (i.e., will be confused with a distractor in the near future) and require more resources to solve the correspondence problem. In that sense, our simultaneity manipulation was similar to the manipulations of proximity and speed across targets in one trial that others had undertaken to estimate dynamic resource reallocation across targets ([Iordanescu et al., 2009](#); [Vul et al., 2009](#); [Chen et al., 2013](#); [Meyerhoff, Papenmeier, Jahn, & Huff, 2016](#)). In this way, we provided a more sensitive measure of using the velocity or acceleration for tracking while controlling for the total number of objects to be tracked and it allowed us to examine how

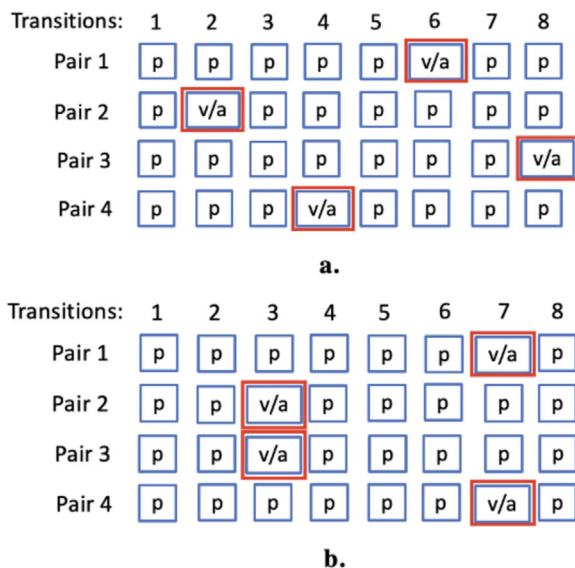


Figure 11. Transition diagrams for velocity or acceleration conditions with different simultaneities. (a) One-pair simultaneity: Only one randomly selected pair of objects takes the critical transition at a time. (b) Two-pair simultaneity: Two randomly selected pairs of objects take the critical transition at a time.

the kinematic load affected attentional reallocation for using motion information.

Method

Stimuli and procedure

The general stimuli and procedure were identical to that of [Experiment 1](#), except that the trials for the velocity and acceleration had asynchronous critical transitions and we later refer to the number of pairs that underwent critical transitions simultaneously as *simultaneity*. For the single-pair simultaneity condition, only one randomly selected pair of objects underwent the critical transition at a time and the critical transitions for the four quadrants occurred at the second, fourth, sixth, and eighth transitions ([Figure 11a](#)). Similarly, for the two-pair simultaneity condition two quadrants underwent the critical movement simultaneously during the third transition, and the remaining two quadrants had the critical movement in the seventh transition ([Figure 11b](#)). In all conditions, each of the four pairs of objects underwent a critical transition exactly once. Thus, there were seven different conditions: position condition, velocity conditions with simultaneity being one, two or four, and acceleration conditions with simultaneity being one, two or four.

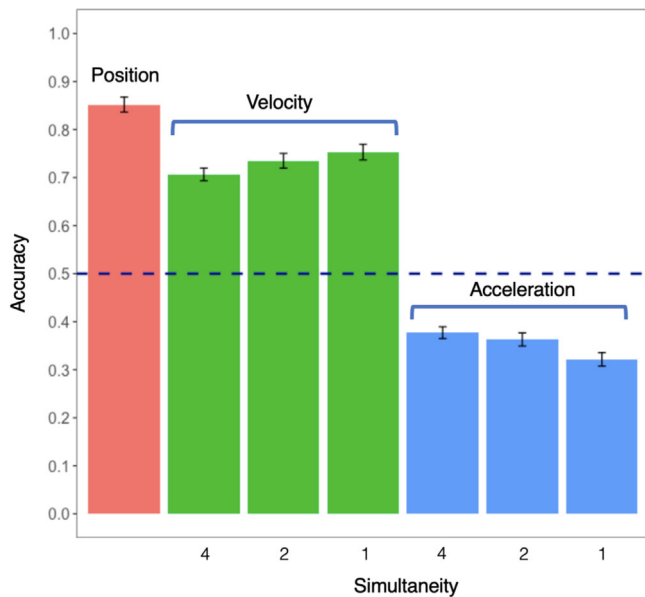


Figure 12. Tracking accuracy for varying simultaneity. The x-axis represents simultaneity and the y-axis represents accuracies. Error bars show the between-observer standard errors. The green bars are accuracies for the velocity conditions with varying simultaneity and the accuracies increase as the simultaneity decreases. The blue bars are accuracies for the acceleration conditions with varying simultaneity and the accuracies decrease as the simultaneity decreases.

Participants

One hundred twenty-four UCSD undergraduates were recruited through the online SONA system and completed the study for course credit. Participants completed 7 trials for each condition and 49 trials in total.

Results and discussion

To test if the use of kinematic information is modulated by kinematic load, we evaluated the performance when different numbers of pairs (i.e., four pairs, two pairs or one pair) simultaneously underwent the critical velocity or acceleration transitions. If observers were able to use velocity information more effectively when fewer pairs underwent simultaneous critical transitions, the accuracy for the velocity condition should increase as the simultaneity decreases. In contrast, observers seemed not to use acceleration information to solve the correspondence problem, but if they were able to use acceleration information when the acceleration condition had lower kinematic load, they should perform above chance.

Figure 12 shows subject performance in acceleration and velocity conditions as a function of simultaneity. Performance when simultaneity is four in the velocity or acceleration condition replicates the results in

Experiment 1: people were reliably above chance in the velocity condition ($M = 70.6\%$), and reliably below chance in the acceleration condition ($M = 37.7\%$). In the velocity condition, performance systematically improved as the simultaneity decreased from four pairs ($M = 70.6\%$) to two pairs ($M = 73.5\%$) to one pair ($M = 75.3\%$). A within-subject one-way analysis of variance revealed significant differences between the means, $F(2, 246) = 8.556, p < 0.001$. Pairwise t -test with Holm-Bonferroni corrections showed the mean accuracy of the one-pair velocity condition was significantly higher than that of the four-pair velocity condition, $t(123) = 3.75, p < 0.001$; the two-pair velocity condition was significantly higher than that of four-pair velocity condition, $t(123) = 2.67, p = 0.017$; there was no significant differences between two-pair velocity condition and one-pair velocity condition, $t(123) = -1.65, p = 0.1$. These results indicated that, when the kinematic load was lower, observers were able to use velocity information more effectively to disambiguate targets from distractors at the confusion points. In contrast, in the acceleration condition, accuracy systematically decreased as the simultaneity decreased from four pairs ($M = 37.8\%$), to two pairs ($M = 36.3\%$), to one pair (32.2%). A within-subject one-way analysis of variance showed that there were significant differences between the means, $F(2, 246) = 8.021, p < 0.001$. A pairwise t -test with Holm-Bonferroni corrections revealed that the mean accuracy of the one-pair acceleration condition was significantly lower than that of the four-pair velocity condition, $t(123) = -3.85, p < 0.001$; the one-pair acceleration condition was significantly lower than that of the two-pair acceleration condition, $t(123) = -2.97, p = 0.007$; there was no significant difference between the two-pair condition and four-pair condition, $t(123) = -0.94, p = 0.346$. This result confirmed that observers do not use acceleration information to solve the correspondence problem and they were even more prone to misidentify the distractors as targets (i.e., the swap phenomenon). It was also consistent with the idea that people used the lagged velocity information to erroneously disambiguate targets from distractors in the acceleration condition.

Experiment 4 showed that people perform worse when multiple targets require velocity information simultaneously, but they performed better when multiple targets require acceleration simultaneously. These seemingly opposite kinematic load effects were both consistent with people modulating attention to process velocity information with different precision. Under this account, when four objects were undergoing a velocity transition, they were each represented with limited velocity precision and this lower precision decreased observers' ability to track. When only one object was undergoing a velocity transition, it might benefit from more attention, and thus greater

precision in velocity, allowing it to be tracked more effectively. If, instead, the objects were undergoing an acceleration transition, greater precision in representing velocity would yield systematically worse performance than a scenario where velocity was represented less precisely (see [Figure 5](#)). Consequently, greater precision in representing velocity with low kinematic load predicted that performance would improve as kinematic load decreased in the velocity condition, but in the acceleration condition lower kinematic load would lead to systematically worse performance. In summary, the results in [Experiment 4](#) imply that attentional allocation can flexibly modulate the precision of velocity information based on task demands.

General discussion and conclusion

In the current study, we asked if people can use motion information to track objects in scenarios where only velocity or acceleration can disambiguate targets from distractors. [Experiment 1](#) demonstrated that velocity direction was used for tracking when positional information was ambiguous, but average acceleration direction alone was not used when both position and instantaneous velocity were ambiguous. In [Experiment 2](#), we replicated the results in [Experiment 1](#) and further confirmed that position was more useful for solving the correspondence problem than velocity with our presentation parameters. In [Experiment 3](#), we showed that observers could detect changes to acceleration in exactly the conditions in which they failed to use average acceleration direction to track, indicating that the failure to use average acceleration direction could not be attributed to a sensory insensitivity to acceleration information. In [Experiment 4](#), we manipulated tracking load by varying whether kinematically challenging transitions coincided in time, and found that conditions with lower load yielded greater accuracy in the velocity condition, but lower accuracy in the acceleration condition. These results confirm that observers' used velocity direction but not average acceleration direction to disambiguate targets and distractors, and suggest that the effectiveness of using velocity was modulated by how many objects require precise velocity information simultaneously. Together, these results clarify the role of different kinematic information in object tracking.

Use of velocity for tracking

When one feature can unambiguously solve the correspondence problem, we cannot ascertain whether, let alone how, any other features are used for tracking. For instance, when objects have stable, unique colors, these are completely sufficient for

correspondence, and the tracking problem becomes trivial. Likewise, in real-world environments, objects tend to be reliably separated in space, making position information dominate solutions to the correspondence problem. [Experiment 2](#) confirms that when position is unambiguous other kinematic features (i.e., velocity and acceleration) have no discernable impact to tracking performance. Thus, only when position is rendered ambiguous is it possible to investigate how less diagnostic features like velocity are used to track objects.

In [Experiments 1](#) and [2](#), we explicitly tested whether people can use velocity to track multiple objects when position is completely ambiguous. We assigned targets and distractors intersecting trajectories. At the point of intersection, positions were totally ambiguous; therefore, ascertaining which object was the target and which was the distractor required using the local velocity and acceleration to track through the intersection. We found that, under these circumstances, people could reliably track objects by using velocity to solve the correspondence problem. That said, [Experiment 2](#) showed that tracking via velocity was less accurate than tracking via position alone, indicating that velocity simply did not provide as useful a cue as position, likely because velocity perception is relatively noisier compared with object position. Thus, our results indicate that people can use velocity direction to track objects when spatial position is completely ambiguous owing to the objects intersecting.

Failure to use acceleration for tracking

We showed that, when both instantaneous position and instantaneous velocity were ambiguous, people failed to use stable acceleration to disambiguate targets and distractors. One possible reason is that people do not believe acceleration to be an intrinsic property of objects, but rather believe that it is a property of the terrain; for instance, the acceleration profile of a ball rolled over a ridge is determined not by the ball, but by the topography of the ridge. It would be imprudent to solve the correspondence problem for two balls rolling over uneven terrain based on their acceleration profile, because that is determined by their location, rather than by a stable, intrinsic property of the objects.

However, there is another possible reason, hinted at by the fact that people are reliably below chance in our acceleration conditions. Being reliably below chance in our paradigm is consistent with using lagged velocity to solve the correspondence problem—while instantaneous velocity was perfectly ambiguous at the point of intersection, the velocity from the prior moments would effectively lead observers to systematically misassociate targets and distractors. This result is consistent with prior

work suggesting that reports of position are lagged (Howard et al., 2011; Howard & Holcombe, 2008) and extrapolated from a lagged position (Iordanescu et al., 2009). Thus, our below-chance results in the acceleration condition are consistent with the use of lagged velocity for correspondence. Because lagged velocity encourages misassociation, our conclusions about use of acceleration must be tempered: acceleration does not yield a strong enough correspondence cue to overcome a misleading lagged velocity signal.

Previous work showed that location trumps velocity (Franconeri, Pylyshyn, & Scholl, 2012), and we show that this does not indicate that velocity is not used; our experiment that does not pit prior position against the prediction of velocity shows that velocity may be used. Thus, similar logic may be applied to our finding of the failure to use acceleration: we effectively pit acceleration against lagged velocity and find that lagged velocity wins. It may be the case that an experiment that does not pit acceleration against velocity in this manner will find that acceleration may be used to solve the correspondence problem.

Implications of kinematic load

In addition to disagreements about the kinds of information that are used to track and select objects by visual attention, there remains an active debate whether attention and tracking are limited by a finite number of slots (Pylyshyn, 2001), or by a flexibly allocated resource (Alvarez & Franconeri, 2007; Holcombe & Chen, 2012). We showed in Experiment 3 that, when multiple objects underwent a transition that required velocity to disambiguate, tracking performance suffered. This result is consistent with the report that the precision of position estimates increases for targets adjacent to distracters—when greater precision is needed to track successfully (Iordanescu et al., 2009). Our result seems to indicate that when velocity is needed to track successfully, it may be used, but it cannot be used for free—the more objects require velocity simultaneously, the more performance suffers. This finding is consistent with velocity information, like position information, being modulated by a limited, albeit flexible, resource (Alvarez & Franconeri, 2007) the deployment of which might be either serial or parallel.

Mechanism of using velocity

Our results showed that people could use velocity to track objects when positional information was completely ambiguous. We now have three qualitatively different classes of experiments that yield seemingly conflicting results about velocity use in object tracking. In the target recovery paradigm (Fencsik et al., 2007;

Franconeri et al., 2012; Keane & Pylyshyn, 2006) a moving target disappears for some protracted period of time, and then reappears either at its original location, or at an extrapolated location. In this paradigm, people can better track objects when they reappear at their original, rather than the extrapolated, position, suggesting that people do not use velocity to track when tracking more than two targets. In the motion predictability paradigm (Howe & Holcombe, 2012; Luu & Howe, 2015), objects remain continuously visible (i.e., they only disappear for brief durations governed by the animation frame rate), but their velocity either changes smoothly or rapidly. When tracked objects undergo rapid changes in velocity direction, they are harder to track, suggesting that a stable velocity direction is somehow advantageous for tracking. In our paradigm, objects pass through the same position at the same time, but with opposite velocity directions. We find that people can accurately track objects despite position being completely ambiguous, indicating that they must somehow be using velocity to track. How can we reconcile the results where velocity seems to be used to track with the target recovery paradigm, where it seems decidedly not to be used?

To do so, we consider the correspondence problem inherent in the multiple object tracking paradigm task and, in particular, how velocity may be used to resolve correspondence. The correspondence problem arises whenever there are several latent causes (such as objects) of our observations, which may be changing over time or space, and we have to figure out which observations correspond to which objects. For instance, to determine the stereo disparity of a given object across our two eyes, we have to determine which features observed by the two retinas correspond to the same object in the world (e.g., Julesz, 1978; Marr & Poggio, 1976, 1979). In object tracking, the correspondence problem amounts to figuring out which of the unlabeled, and identical, spots in a given frame correspond to the objects we are trying to track (Vul et al., 2009). The computational core of the correspondence problem amounts to matching what we see to the observations expected from different latent states. How this matching is done determines what kinds of features are used for tracking.

Velocity may be used to resolve correspondence in at least two qualitatively different ways. First, velocity could be used solely to extrapolate position, in which case position is the only feature used for correspondence, but expected position is modulated by velocity. Two variants of this account are notable: Full extrapolation, in which extrapolation is done as though the last seen velocity continues unchanged for the full duration of disappearance, and partial extrapolation (Zhong et al., 2014), in which position is extrapolated by only a fraction of the amount expected under Full Extrapolation. There is a second, qualitatively different,

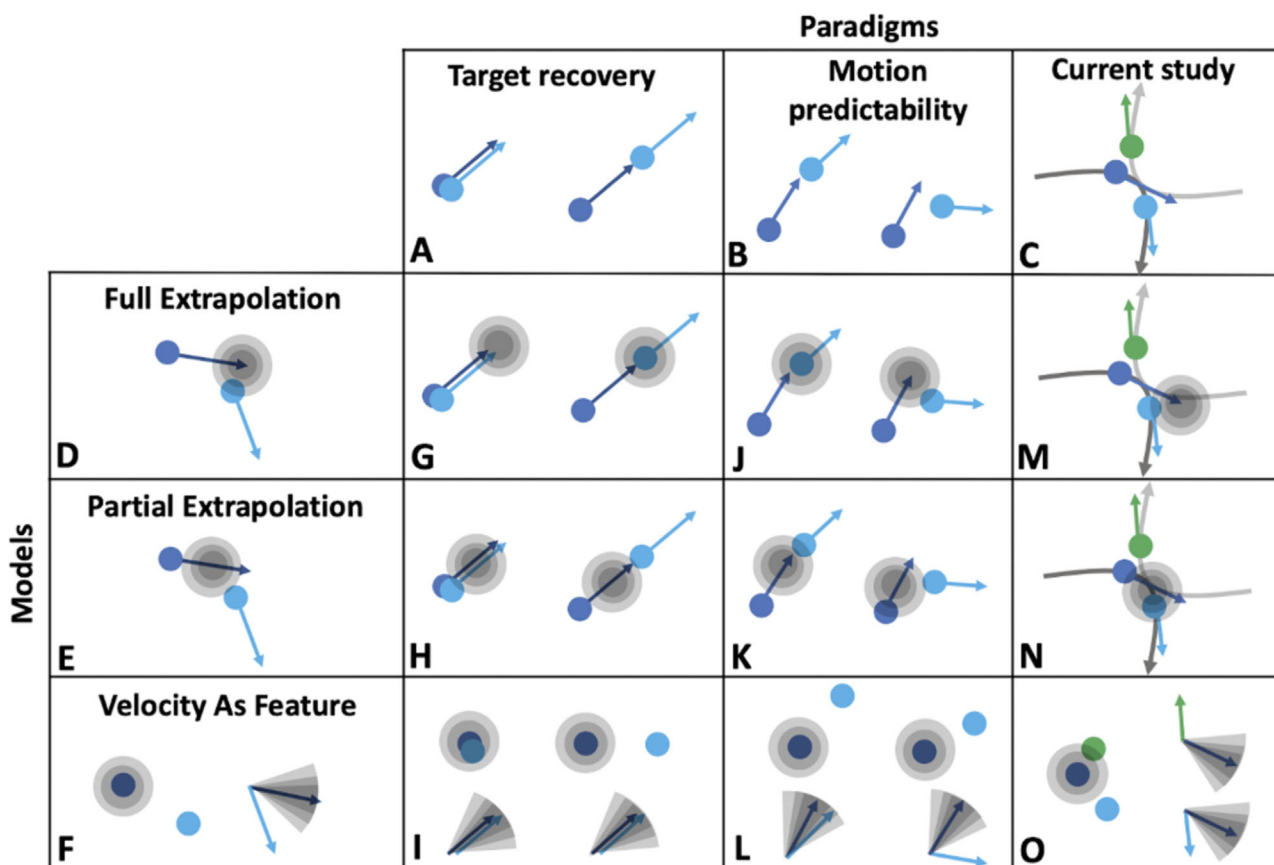


Figure 13. Different ways that velocity can be used in object tracking (rows) predict different patterns of behavior across three paradigms studying velocity use in tracking (columns). For all models, velocity is used to predict an observation (gray gradients) of an object at time t (light blue), from its behavior at time $t - 1$ (dark blue), and the extent to which the observation matches the prediction determines tracking accuracy. (A) Target recovery paradigm: The target (dark blue) disappears part way through tracking and reappears (light blue) either at its original location (left) or at the extrapolated position (right). People track better when it reappears in the original location. (B) Motion predictability paradigm: People can track targets better when they move in predictable paths, with velocity changing slowly (left) than in unpredictable paths, with velocity changing rapidly (right). (C) Current study: People can correctly track a target (dark gray path) through a transition where the target perfectly overlaps with a distracter (light gray path), rendering the position completely ambiguous for resolving the correspondence problem between the subsequent position of the target (light blue) and distracter (green). (D) Full extrapolation model: observations are matched only on position, but predicted position is fully extrapolated based on velocity. (E) Partial extrapolation model: again, correspondence is solved only by position, and position is extrapolated based on velocity, but it is extrapolated only partially. (F) Velocity as a feature: people match on both position (points) and velocity (arrows), but velocity is not used to extrapolate positions; it is treated as an independent and unrelated feature. The grey gradient fan represents predicted velocity. (G) Full extrapolation predicts better performance when an object reappears in the fully extrapolated position. (H) Partial extrapolation predicts worse performance in the extrapolation condition so long as the fraction of extrapolation is less than 50%. (I) Velocity as a feature predicts better performance when objects reappear in the original position, because that’s where position is a better match, and velocity is an equally good match in both conditions. (J–L) All three models predict that people would perform better in the predictable condition. (M–O) All three models predict an ability to track through the confusion point by exploiting the stability of velocity to solve the correspondence problem in favor of the target (red) rather than the distracter (green).

class of account—velocity as feature account—in which velocity may be used as an independent feature in solving correspondence, akin to color or spatial frequency. Under the feature account, velocity is not used to extrapolate position, but instantaneous velocity of observed objects is compared with the remembered velocity of tracked objects to resolve the correspondence problem. The prior literature has

mainly considered the full extrapolation account of how velocity may be used for tracking, and has come up with confusing, mixed results. We consider which reported results are consistent with which variants of these accounts.

Figure 13 illustrates how the different accounts predict performance in the three paradigms we are considering here. The full extrapolation account is

unique in predicting better tracking of objects that are completely extrapolated in the target recovery paradigm. However, people are unlikely to extrapolate fully the target positions in the target recovery paradigm. The reason may be that people have priors favoring slower speeds (Stocker & Simoncelli, 2006) and, when combined with noisy and uncertain velocity percepts, the estimated speed would be slowed and extrapolation would be partial and conservative (Zhong et al., 2014). Thus, partial extrapolation should be the default expectation and may lead to the expected location of objects being closer to where they were recently perceived than to where they are headed. For velocity as a feature, the target recovery result is anticipated because the original position, with the original velocity, will always be a better match than the extrapolated position. Thus, the partial extrapolation, and velocity as a feature accounts agree with the observation that people are worse when objects reappear in their fully extrapolated positions. Further disambiguating these accounts is a bit murky: tests of the key differential prediction—do people extrapolate object positions—have yielded mixed results. In contrast, Howard and his colleagues (2008, 2011) found that, instead of extrapolating while tracking, position estimates tend to lag behind the physical locations; in contrast, there is some evidence that people can explicitly extrapolate positions when tracking only a couple of objects (Iordanescu et al., 2009). It is possible that positions are sampled less frequently at higher load causing perceived location to lag behind the physical stimulus, and a limited capacity process can perform extrapolation from these lagged positions, compensating for the lag in low-load scenarios. Despite the remaining ambiguities, the existing results, including our own, are consistent with velocity being used during tracking in one of two ways: either it is used to partially extrapolate position or it is used solely as a feature.

Keywords: MOT, attention, correspondence, motion extrapolation, velocity, acceleration

Acknowledgments

Commercial relationships: none.
Corresponding author: Yang Wang.
Email: yaw001@ucsd.edu.
Address: Department of Psychology, University of California, San Diego, California, USA.

References

- Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, 7(13), 14, 1–10.
- Anstis, S., & Ramachandran, V. S. (1987). Visual inertia in apparent motion. *Vision Research*, 27(5), 755–764.
- Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, 408(6809), 196–199.
- Bettencourt, K., & Somers, D. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision*, 9(7), 1–11.
- Chen, W.-Y., Howe, P. D., & Holcombe, A. O. (2013). Resource demands of object tracking and differential allocation of the resource. *Attention, Perception & Psychophysics*, 75(4), 710–725.
- 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 and Psychophysics*, 69(4), 567–577.
- Fencsik, D. E., Urrea, J., Place, S. S., Wolfe, J. M., & Horowitz, T. S. (2006). Velocity cues improve visual search and multiple object tracking. *Visual Cognition*, 14, 92–95.
- Feria, C.S. (2013). Speed has an effect on multiple-object tracking independently of the number of close encounters between targets and distractors. *Attention, Perception and Psychophysics*, 75, 53–67.
- Franconeri, S. L., Lin, J. Y., Pylyshyn, Z. W., Fisher, B., & Enns, J. T. (2008). Evidence against a speed limit in multiple-object tracking. *Psychonomic Bulletin & Review*, 15, 802–808.
- Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science*, 21, 920–925.
- Franconeri, S. L., Pylyshyn, Z. W., & Scholl, B. J. (2012). A simple proximity heuristic allows tracking of multiple objects through occlusion. *Attention, Perception and Psychophysics*, 74(4), 691–702.
- Holcombe, A. O., & Chen, W. Y. (2012). Exhausting attentional tracking resources with a single fast-moving object. *Cognition*, 123(2), 218–228.
- Howard, C. J., & Holcombe, A. O. (2008). Tracking the changing features of multiple objects: Progressively poorer perceptual precision and progressively greater perceptual lag. *Vision Research*, 48(9), 1164–1180.
- Howard, C. J., Masom, D., & Holcombe, A. O. (2011). Position representations lag behind targets in multiple object tracking. *Vision Research*, 51(17), 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.
- 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.
- Huff, M., Papenmeier, F., Jahn, G., & Hesse, F. W. (2010). Eye movements across viewpoint changes in multiple object tracking. *Visual Cognition*, *18*, 1368–1391.
- 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), 1–12.
- Julesz, B. (1978). Global Stereopsis: Cooperative phenomena in stereoscopic depth perception. In: R. Held, H. Leibowitz, & H. Teuber (Eds.), *Handbook of Sensory Physiology, vol. 8, Perception*. Berlin, Heidelberg: Springer, pp. 215–256.
- 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*(4), 346–368.
- Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., . . . Enns, J. T. (2005). Multiple-object tracking is based on scene, not retinal, coordinates. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 235–247.
- Luu, T., & Howe, P. D. (2015). Extrapolation occurs in multiple object tracking when eye movements are controlled. *Attention, Perception & Psychophysics*, *77*, 1919–1929.
- Makovski, T., & Jiang, Y. V. (2009). The role of visual working memory in attentive tracking of unique objects. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1687–1697.
- Marr, D., & Poggio, T. (1976). Cooperative computation of disparity. *Science*, *194*, 283–287.
- Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *204*, 301–328.
- Meyerhoff, H. S., Papenmeier, F., & Huff, M. (2017). Studying visual attention using the multiple object tracking paradigm: A tutorial review. *Attention, Perception & Psychophysics*, *79*, 1255–1274.
- Meyerhoff, H. S., Papenmeier, F., Jahn, G., & Huff, M. (2016). Not FLEXible enough: Exploring the temporal dynamics of attentional reallocations with the multiple object tracking paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(6), 776–787.
- Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*(1), 65–97.
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 1–19.
- Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, *80*(1–2), 127–158.
- Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*. Cambridge, MA: MIT Press/Bradford Books.
- Ramachandran, V. S., & Anstis, S. M. (1983). Perceptual organization in moving patterns. *Nature*, *304*, 529–531.
- Shooner, C., Tripathy, S. P., Bedell, H. E., & Ogmen, H. (2010). High-capacity, transient retention of direction-of-motion information for multiple moving objects. *Journal of Vision*, *10*(6), 8, 1–20.
- Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), 578–85.
- Tombu, M., & Seiffert, A. E. (2011). Tracking planets and moons: Mechanisms of object tracking revealed with a new paradigm. *Attention, Perception, & Psychophysics*, *73*, 738–750.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- Vul, E., Frank, M. C., Tenenbaum, J. B., & Alvarez, G. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Advances in Neural Information Processing Systems*, *22*, 1–9.
- Watamaniuk, S. N. J., & McKee, S. P. (1995). Seeing motion behind occluders. *Nature*, *377*, 729–730.
- Watamaniuk, S. N. J., McKee, S. P., & Grzywacz, N. M. (1995). Detecting a trajectory embedded in random-direction motion noise. *Vision Research*, *35*(1), 65–77.
- Verghese, P., Watamaniuk, S. N., McKee, S. P., & Grzywacz, N. M. (1999). Local motion detectors cannot account for the detectability of an extended trajectory in noise. *Vision Research*, *39*(1), 19–30.
- 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.