Selecting and tracking multiple objects

Jason M. Scimeca[1]* and Steven L. Franconeri[2]

When interacting with the world, people can dynamically split attention across multiple objects in the environment, both when the objects are stationary and when the objects are moving. This type of visual processing is commonly studied in lab settings using either static selection tasks or moving tracking tasks. We describe performance limits that are common to both tasks, including limits on capacity, crowding, visual hemifield arrangement, and speed. Because these shared limits on performance suggest common underlying mechanisms, we examine a set of models that might account for limits across both. We also review cognitive neuroscience data relevant to these limits, which can provide constraints on the set of models. Finally, we examine performance limits that are unique to tracking tasks, such as trajectory encoding, and identity encoding. We argue that a complete model of multiple object tracking must account for both those limits shared between static selection and dynamic tracking, as well as limits unique to tracking. It must also provide neurally plausible mechanisms for the underlying processing resources. © 2014 John Wiley & Sons, Ltd.

INTRODUCTION

When navigating through the world, we frequently feel that we split our attention across multiple objects in the environment. In static scenes or displays, we might decide which of two apples is larger, or compare separate parts of a diagram. In dynamic scenes, we feel that we watch multiple cars when we cross the street, or that we can simultaneously monitor multiple children at the playground. We also track multiple objects in constructed displays, such as dynamic depictions of patterns in data,[1,2] or air-traffic control systems.[3] How many objects can we attend to at once? Is the number fixed or does it vary for different situations? When objects move, is it harder to maintain our attention on them, especially as they move faster?

There are many experimental tasks used to address these questions in the lab. In this review, we will consider parallels between two classes of tasks that have previously been considered largely independently. One class incorporates tasks that use static displays, which we will refer to as *multiple object selection* (MOS) tasks. The second class has explored divided attention abilities in dynamic displays filled with moving objects, which we will refer to as *multiple object tracking* (MOT) tasks. These tasks have served as well-controlled 'toy worlds' that allow the isolated study of a fundamental mechanism within the spatial attention processing system, and results from each of these tasks have helped to provide answers to many of our questions related to splitting attention.

For both types of task, various manipulations of display presentation, stimulus dimensions, and task demands impact an observer's performance. Understanding how these factors constrain performance is critical to a complete model of MOT. In the first half of this article, we highlight performance limits that are shared between MOS and MOT because these similarities imply shared underlying mechanisms. We then discuss a set of models that might account for these shared constraints. In the second half of this article, we review cognitive neuroscience models of the brain networks that support performance in static selection and dynamic tracking. Finally, we review two performance

*Correspondence to: jason_scimeca@brown.edu

[1]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA

[2]Department of Psychology, Northwestern University, Evanston, IL, USA

constraints imposed by the unique dynamic nature of tracking tasks: limited encoding of trajectory information and limited encoding of property information.

PERFORMANCE LIMITS: MOS TASKS

A large literature shows that an attentional 'spotlight' can select specific aspects of the visual world, including specific locations. This selection process amplifies information from the attended area, relative to unattended area (see Ref 4 for review). Although early work assumed a single spotlight of attention (e.g., Ref 5), it is now assumed that spatial attention can be split to select multiple noncontiguous locations.[6–9] MOS tasks typically demand that participants select and attend to spatial locations and report on some aspect of stimuli that are presented in the monitored locations (Figure 1, first row). For example, when participants are asked to report the identity of objects flashed in the visual field, they are better able to identify objects that appear in two attended locations compared with objects that appear in the intervening space between the locations.[6]

Figure 2 depicts many of the performance limitations on the splitting of attention across noncontiguous locations. For example, some studies attempt to measure the maximum number of independent locations that can be selected at once. Observers are cued to a subset of locations within a display, before presenting a search display consisting of a full set (e.g., a grid) of locations. While one such study suggested that observers could select subsets of up to five items,[10] subset sizes larger than five items were not tested. More recent work has shown that people can select up to 7–8 locations at once.[11] Such high performance levels were only possible when locations were spaced out across the visual field, allowing for coarser specification of each individual location. When attended locations were crowded more closely to each other, requiring precise specification of each location, the

capacity of selection was about 2–3 locations. This effect of close spacing impairs performance both when targets are crowded near each other as well as when they are crowded near distractor items.[12,13]

The capacity of MOS is also constrained by the locations of attended objects relative to the vertical visual hemifield boundary. When attended locations fall within a single hemifield, capacity is lower than when the locations straddle the hemifield boundary (Figure 2). This effect has been seen across a variety of experimental paradigms such as letter identification and orientation discrimination tasks.[6,14–16]

PERFORMANCE LIMITS: MOT TASKS

MOT tasks are similar to multiple selection tasks, except that the selected objects move. Observers are less likely to be asked to encode featural information about the objects, which are typically depicted as a homogeneous set of circles or boxes. Instead, observers are typically challenged to remember *which* objects were originally designated as targets, by tracking their locations dynamically over time. The task is similar to the game played by street magicians who hide a 'target' ball under one of three cups, quickly shuffle the cups around, and then ask the observer to indicate the cup that contains the target object. The difference is that laboratory MOT tasks require tracking of multiple targets—as many as eight.

Targets are typically designated by briefly changing color before the trial, leaving the other items designated as 'distractors' (Figure 1, second row). The items move about the screen for several seconds before coming to a stop, and then performance is measured by having participants indicate which target items they were tracking. Because the distractor items typically move through the space intervening the target items, accurate performance in this task requires the observer to select only the target objects and not the intervening space or objects.

The performance limits shown in Figure 2 are as relevant for MOT as they were for static MOS. For capacity, the first studies of MOT suggested that participants could track four or five items at once (e.g., Refs 17–19). This limit was historically consistent with other results suggesting a 'magical number 4' in visual cognition,[20,21] but see also Ref 22 and other cognitive systems.[23]

But again, more recent studies have shown that capacity in object tracking tasks is not necessarily fixed at four or five objects but instead can reach up to seven or eight objects at once.[24] Like in static selection tasks, dynamic tracking performance is also strongly affected by crowding—how closely the objects in the display
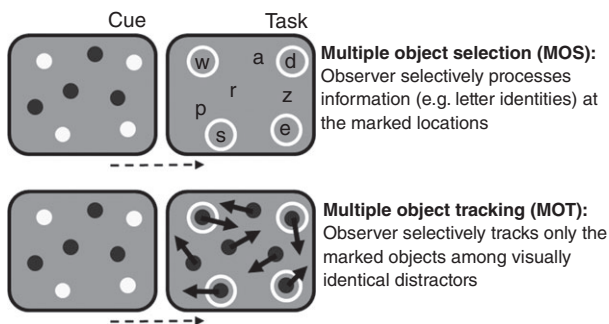

	Cue	Task

FIGURE 1 | Multiple object selection (MOS) and multiple object tracking (MOT).

Multiple object selection (MOS): Observer selectively processes information (e.g. letter identities) at the marked locations

Multiple object tracking (MOT): Observer selectively tracks only the marked objects among visually identical distractors
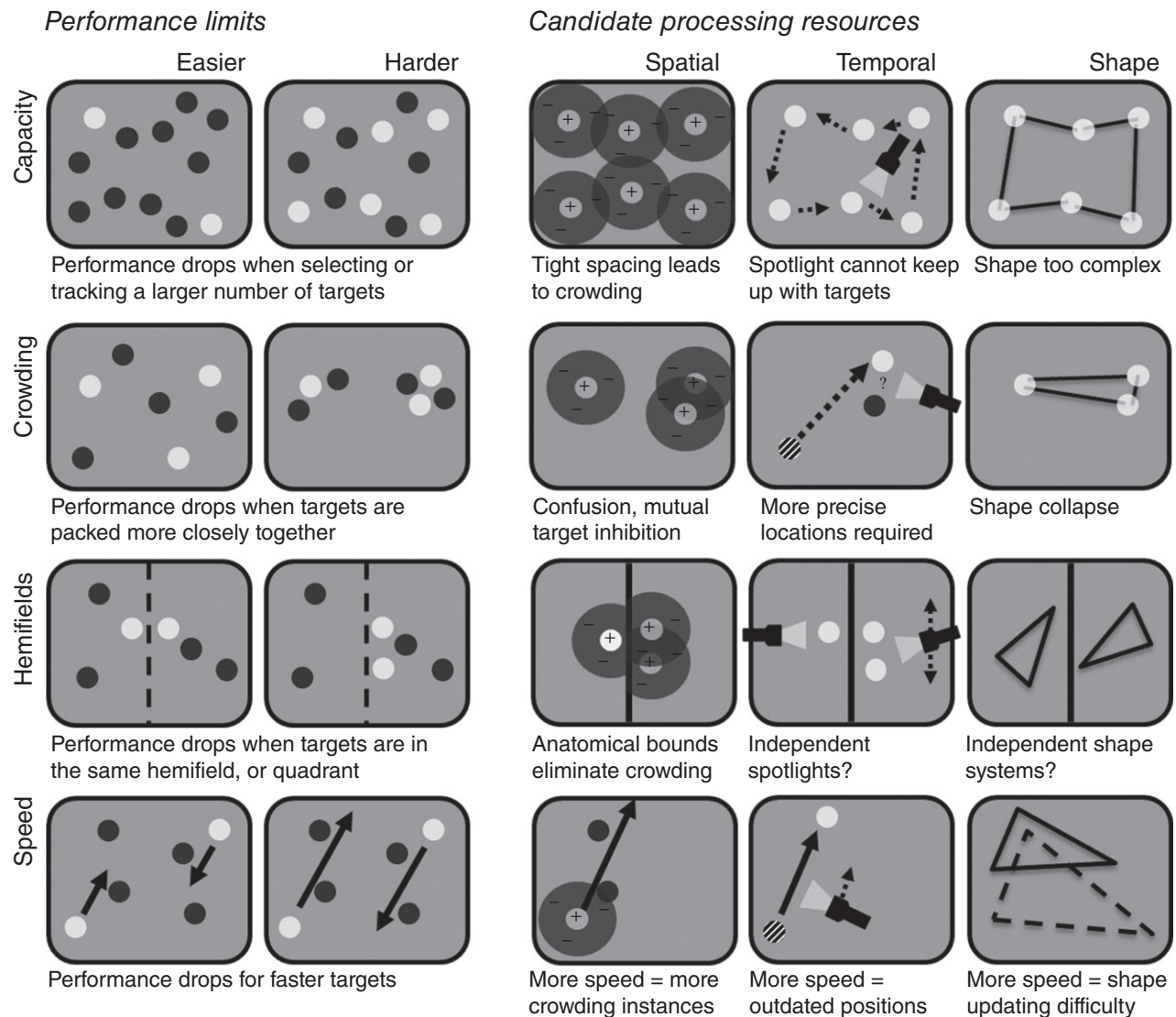
FIGURE 2 | Performance limits and possible underlying processing resources.

are spaced. Moving the objects closer together leads to lower tracking capacity, relative to when objects are more loosely spaced.[17,25–27] Again in parallel to static selection tasks, dynamic tracking capacity is also constrained by the locations of objects relative to the visual hemifield boundary. Tracking objects within the same visual hemisphere leads to worse performance relative to tracking objects in different hemispheres.[28] This performance benefit occurs across both the vertical and horizontal midlines.[29]

Dynamic displays also bring limits that are not measureable in selection tasks that use static displays. Most saliently, tracking capacity is affected by the speed at which the objects travel. There are very fast speeds where observers can track only a single object, and extremely slow speeds where observers can

track up to eight objects.[24] There are additional limits on performance that are relevant only to dynamic tracking tasks, including trajectory extrapolation and identity encoding, and these unique limitations are discussed in a separate section below. First, we will outline three models of the roots of these processing resource limits, and discuss how these models might account for the limits related to crowding, capacity, hemifield placement, and object speed, for both static and dynamic tasks.

WHERE DO THESE PERFORMANCE LIMITATIONS COME FROM?

The common limits on performance in these tasks suggest a common underlying mechanism. Here we

review three classes of limited processing resources that might underlie many of these performance limits seen in selecting and tracking multiple objects. The first class is a set of *spatial* limitations within the cortical maps that represent selection regions within the visual field. The second class is a set of *temporal* limitations on some processing resource that shifts over time among tracked objects. The third class assumes that tracking relies on the *shape recognition* system, treating a set of static or moving objects as vertices in a polygon—and that limitations in this system underlie the performance limits discussed so far.

Spatial Limits

The first paper to introduce the MOT paradigm[18] also suggested that the mechanism underlying tracking was a parallel set of object location pointers (called 'FINSTs') that automatically shifted to follow objects when they moved. These pointers reference only the object's location, allowing the visual system to mark, attend, track, or count the objects within the selected set. We take this mechanism to be a higher-level description of what is now known as an attentional priority map,[30] a cortical representation that marks a set of locations in the visual field for enhanced processing.

We have recently proposed that a set of known properties of such maps can explain all of the limitations on MOS and MOT discussed so far—capacity, crowding, hemifield, and speed limits in MOT tasks. The spatial interference account[31] suggests that because regions of spatial selection tend to merge with[17,32] or inhibit[13–15,33–36] other regions as they are pushed closer together, this damages a person's ability to isolate target locations from distractor locations.

The third column of Figure 2 depicts how these constraints might account for the performance limits shared between static selection and dynamic tracking. The capacity limit comes from squeezing more selection regions into a visual field of fixed size. The crowding limit is similar, and stems from squeezing the same number of selection regions into a smaller region of the visual field, and squeezing distractor items dangerously close to selection markers for targets. The hemifield benefit stems from the ability of the anatomical hemifield boundary to block the competitive effects of close spacing.[37] This account gives a surprising explanation for why more objects can be tracked when they move more slowly—and it has little to do with tracking impairment from speed increases, per se. Instead, when objects move more quickly there are more instances where objects might interfere with each other due to changes in their spacing. When these interference instances were equated across different

speeds, speed no longer showed evidence of drawing on limited tracking resources.[25,31]

This account is not without controversy, and some argue that the tradeoffs between capacity and speed are better explained by more abstract processing resource limitations.[38–40] Future work is needed to test the predictions of this model and the underlying mechanisms. The magnitude of this spatial interference is fairly well studied in static displays, but less understood in dynamic displays. Quickly moving objects may rely on larger receptive fields, which may lead to less precise position representations.[24,41] In order to explain all results from tracking tasks, this account may also have to specify that the interference reach through entire visual hemifields. For example, some models of the principles underlying visual attention networks specify that adding more 'targets' (more activation bumps in the cortical map) leads to a weakening of all bumps, and thus a more noisy representation and lower performance ([42]; see also Ref 43 for a similar model of visual working memory).

Temporal Limits

Although there is evidence for parallel operation across multiple objects in both MOS[6–9] and MOT,[44,45] an alternative model is that tracking involves an attentional spotlight that engages in serial shifts between objects.[46,47] In this model, the position of a tracked object can only be updated when some resource-limited process (anthropomorphized in the fourth column of Figure 2 as a physical spotlight) operates on that particular object.

Studies using change detection and similar paradigms have reported a limit on the speed at which the attentional system can select locations. In a change detection task where observers monitored two locations, one in each hemifield, for a briefly flashed target stimulus, performance at each location fluctuated with a rhythm of 4 Hz, and the two rhythms were in antiphase.[48] This pattern is consistent with a sampling rate of 8 Hz. A similar limit is observed even when attention is focused on a single target: when observers monitor a single location for a target, psychometric modeling suggested a 7 Hz sampling rate.[49] Similarly, Holcombe and Chen[46] argued that splitting attention between multiple moving objects reduces the temporal resolution of attention. Observers tracked objects that rotated in concentric rings; on each ring a single target object was intermixed between visually identical distractor objects. By varying the rotational speed of the rings and the number of distractor objects on each ring, they estimated that the temporal resolution of tracking was about 7 Hz when tracking one target,

but fell to 4 Hz when tracking two targets, and 2.6 Hz when tracking three targets.

In the serial switching model, the 'spotlight' must cycle through the target objects, and the limit on the speed of this cycle could underlie many of the processing limits of MOS and MOT. The fourth column of Figure 2 shows how this switching limit relates to different performance limits. With an increased number of targets, the attentional spotlight must cycle between more objects, allowing more time to pass between indexing the location of any one target and returning to that location. When the attentional spotlight returns to this location, it must find the nearest object and update the location index. Increasing the speed or the number of targets will increase the uncertainty and error in this updating process, and have a negative impact on performance. Increasing both the speed and number of target objects should result in an over-additive impairment on performance. The hemifield effect is attributed to attentional spotlights that are independent for each hemifield. Because the crowding limits are spatial in their essence, they are most likely captured by the known spatial limitations discussed in the previous section. Nevertheless, Figure 2 depicts a temporal framing of this limit, focused on the increasing need for spatial precision when the temporal resource is strained.

Shapes

One intriguing alternative processing resource is the visual system's shape recognition system. Tracking multiple objects might require encoding the complex polygon created by the locations of the tracked targets.[19] As the targets move, the shape of the polygon morphs. The fifth column of Figure 2 depicts how this shape system might account for performance limits in MOS and MOT. Limits on tracking capacity might stem from the maximum number of vertices that could be robustly stored as a shape (performance is high at least up to 4–5 objects). Although this is a plausible and falsifiable model, surprisingly little research has tested it. This account could even explain other tracking effects with minor extensions. Crowding objects together might make shapes smaller and harder to distinguish, and may cause vertex representations to 'slip' from a target position to a distractor position. Various aspects of shape encoding or maintenance that operate more efficiently across hemifields than within hemifields could explain the hemifield effect. The tradeoff between capacity and speed could stem from more complex shapes (needed when asking a viewer to track more objects), requiring more time to update or consolidate when objects move. While we

consider this account a viable potential mechanism for tracking, the predictions of this model remain unexplored. For example, when an observer's shape memory is loaded to capacity, performance on both static selection and dynamic tracking tasks should suffer.

Which Mechanism Best Explains Performance Limits?

Each of the potential processing limits outlined above give plausible explanations for the performance limits that are shared by static selection and dynamic tracking tasks. Many of the model extensions we have proposed above have yet to be explicitly tested, and we hope that future behavioral work will provide empirical tests of these extensions. We suggest that any model of MOT will benefit from considering those performance limitations shared with static selection. An advantage of the spatial and shape models is that the parallel processing inherent in these models more parsimoniously captures the limits shared between static selection and dynamic tracking. Indeed, the spatial interference model was motivated by the similarity in performance limits observed in MOS and MOT.

The limit due to crowding is likely due to spatial factors, which are well studied in the domain of static selection and are easy to extrapolate to the domain of dynamic tracking. Explaining the other performance limits remains is more challenging, and it will take time to reach a consensus. Under the spatial model, other performance limits are extensions of the underlying interference seen in crowding effects. Competing models introduce new factors: The temporal model suggests that the capacity limit arises because spotlights must visit more locations per unit of time, and the shape model suggests that capacity limits are reached when the underlying shape becomes too complex for the shape recognition system.

The models also differ in their explanation of the hemifield effect: tracking is known to be largely independent in the left and right sides of a display, when the observer strictly fixates the center. The spatial model accounts for this effect with a reminder that the hemifield boundary is known to minimize the effects of crowding. The temporal and shape models must posit new resources (either spotlights or shape systems) that are independent for each hemifield. The models differ most starkly in their explanation of limits due to speed. The spatial model argues that this limit derives simply because faster moving objects results in more instances of crowding and spatial interference. In contrast, the temporal and shape models argue that faster moving objects tax a higher-level processing resource: either the rate at

which the spotlight can accurately alternate between target locations, or the rate at which the shape recognition system can update shape representations.

COGNITIVE NEUROSCIENCE OF MOT

Any model of these tasks should consider how their underlying mechanisms might be implemented in the brain. We turn our attention to cognitive neuroscience research on the neural systems linking selection and tracking—what can the brain tell us about the relationship between MOS and MOT? Functional neuroimaging points to a network of involved brain regions, some shared between these two tasks and some unique to MOT. In fMRI studies, MOT tasks consistently activate a broad dorsal network including areas in posterior parietal cortex and areas in frontal cortex.[50–52] One study contrasted a condition where observers *tracked* moving targets versus a condition where they simply *selected* static targets.[53] Their analysis revealed the posterior intraparietal sulcus was implicated in both selection and tracking. This region has previously been linked to visual working memory.[21,54] Furthermore, they found that object *tracking* was associated with increased activity in area MT+, parietal areas (posterior and anterior intraparietal sulcus and superior parietal lobule), and the frontal eye fields in frontal cortex.

These regions appear to be dissociable in their response to manipulations of various MOT task components. For example, activation in posterior parietal cortex correlates with increases in the number of targets but not increases in the speed at which the objects move, whereas activation in frontal eye fields correlates with increases in both target number and rotational speed.[55] Although this work has identified regions that contribute to tracking, further work should seek to reveal the process specificity and representational content of these areas. The recruitment of this fronto-parietal network is consistent with possible roles for these regions suggested in the broader literature on attention: posterior parietal cortex may be involved in allocating and maintaining the spatial indices or spotlights; while frontal eye fields may be involved in suppressing eye movements or in maintaining a certain level of attentional precision.[53,55]

Research using electrophysiology and event-related potentials (ERP) provides further constraints on models of MOT. Evidence from ERPs suggest that a similar neural substrate may underlie visual working memory and MOT: across ERP studies, the contralateral delay activity (CDA) component has been shown to track visual working memory load[56,57]

and MOT load.[58–60] A recent study using an electrophysiological measure of visual processing in early visual areas found that enhanced processing in these areas was correlated with tracking performance.[45] The relative magnitude of this enhancement was similar across a manipulation of set size, suggesting that attention was equally allocated to all of the tracked targets regardless of set size. The magnitude of the effect was also similar to effects reported in selection tasks, consistent with a common selection mechanism for static and moving displays.

LIMITS ON PROPERTY ENCODING UNIQUE TO TRACKING TASKS: TRAJECTORIES AND VISUAL FEATURES

Trajectories

Motion is the defining property that makes MOT distinct from MOS tasks. In the real world, moving objects often travel in predictable ways—when a bird flies behind a tree, we expect it to emerge from behind the other side. Do we make use of such trajectory information when tracking multiple objects?

Past work has tested this idea by testing for motion extrapolation during disruptions in tracking tasks. One example is the blanking paradigm, where of the objects disappear and then reappear after a short delay, either in the location extrapolated from their initial motion or at the same location that they disappeared, either via sudden disappearance (Figure 3, second row), or gradual occlusion and disocclusion (Figure 3, third row). In both types of study, tracking impairment is worse when objects reappear at locations further away from their original locations.[61–63]

Although these experiments did not find evidence that trajectory encoding was useful for tracking through occlusion, there is evidence that observers may use trajectory information in certain conditions. When asked, observers can explicitly report the direction of motion of objects,[64,65] but only when tracking one or two objects. When objects are given an internal texture that moves in a conflicting direction with the object's motion (Figure 3, fourth row), tracking is impaired.[66] There appears to be a capacity limit on trajectory encoding: there is evidence that tracking through occlusion may benefit from trajectory information when the tracking load is sufficiently low.[67,68] In one study, objects either moved in a straight line and only changed direction when they reached the edge of the screen, or changed directions randomly and unpredictably. There was no difference in tracking performance between the two conditions when observers
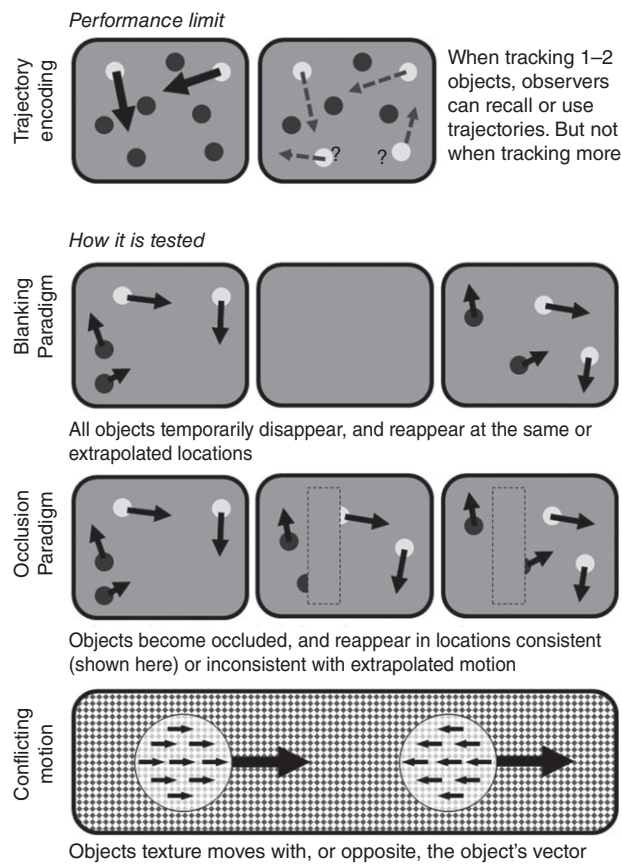
FIGURE 3 | Does tracking use trajectory information?

tracked four objects, but performance was lower in the unpredictable condition when observers tracked only two objects, again suggesting the possibility of extrapolation, but only for 1–2 objects.[69]

Visual Features

Experiments in the 'visual short-term memory' literature show that observers can detect identity changes to a display, with typical capacity estimates of around 3–4 objects (e.g., Ref 20; cf. Ref 22). But when objects change positions, matching identities to objects can be a challenge in such tasks.[70] The challenge is no smaller in MOT tasks, where motion has devastating consequences for an observer who attempts to remember who is who.

When objects are briefly given unique identities at the start of a tracking task (e.g., via unique numbers or pictures; Figure 4, second row), and then these identities are masked during the motion phase of the task, observers can tell which objects were targets but have little idea about which identity corresponded to which object.[71–73] The number of items that can be tracked in this manner is rather small: one study used a similar design with uniquely colored objects and
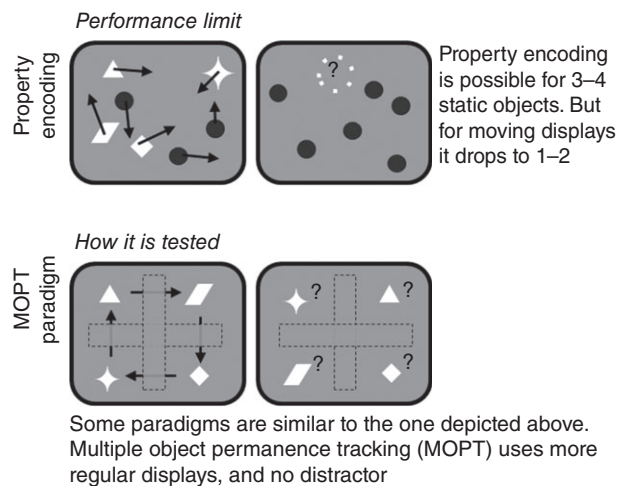

FIGURE 4 | Does tracking maintain property information?

estimated a capacity of only two identities that could be tracked at once, and another found a range between 1.4 and 2.6.[72] Intriguingly, this performance limit also seems to be modulated by visual hemifield, with one study reporting that capacity increases slightly when targets are split across visual hemifields.[74]

In summary, there appears to be some ability to match identities with moving objects, but this ability is as capacity limited as the ability to extrapolate the motion path of an object—a maximum of two objects at a time. Why is the limit on identity encoding more severe than the capacity limit on tracking? One possibility is that binding is required between separate maps that represent identity ('what') and location ('where').[72,75–77] If such bindings require the unitary focus of the attentional spotlight, as predicted by some general models of attention[77] as well as a model of tracking that focuses on this issue,[76] then it makes sense that this ability should be limited to one object. Or perhaps up to two when the objects stay in separate visual hemifields, which may contain separable maps of both 'what' and 'where'.[75]

CONCLUSIONS AND FUTURE DIRECTIONS

Various limits constrain people's ability to allocate attention to multiple objects in tasks involving static selection or dynamic tracking. We have outlined limits on capacity, crowding, hemifield modulations of these limits, object speed, and encoding of object trajectory and featural identities. Many of these limits are shared between MOS and MOT tasks, although models of MOT do not typically address this similarity explicitly. In this article, we have discussed several models that

might underlie the shared limits on these tasks—a parsimonious model must account for both shared performance limits and those unique to tracking tasks.

We hope that cognitive neuroscience data will provide important constraints, and novel predictions, for these models. It is critical that proposals for mechanisms underlying the performance limitations in these tasks go beyond circular declarations of 'resource limitations', to provide a biologically plausible mechanistic explanation. For example, the spatial interference account links performance limits in MOT and MOS to competition in cortical maps of external space.[75] A better understanding of the core cognitive processes and neural systems involved in both static selection and dynamic tracking will lead to a more complete model of MOT.

ACKNOWLEDGMENTS

REFERENCES

1. Chevalier F, Dragicevic P, Franconeri S. The not-so-staggering effect of staggered animations. *IEEE T Vis Comput Gr* 2014, 20:2241–2250.

2. Heer J, Robertson GG. Animated transitions in statistical data graphics. *IEEE T Vis Comput Gr* 2007, 13:1240–1247.

3. Landry SJ, Sheridan TB, Yufik YM. A methodology for studying cognitive groupings in a target-tracking task. *IEEE T Intell Transp* 2001, 2:92–100.

4. Franconeri SL. The nature and status of visual resources. In: Reisberg D, ed. *Oxford Handbook of Cognitive Psychology*, vol. 8481. New York, NY: Oxford University Press; 2013, 147–162.

5. Posner MI. Orienting of attention. *Q J Exp Psychol* 1980, 32:3–25.

6. Awh E, Pashler H. Evidence for split attentional foci. *J Exp Psychol Hum Percept Perform* 2000, 26:834–846.

7. Müller M, Malinowski P, Gruber T, Hillyard S. Sustained division of the attentional spotlight. *Nature* 2003, 424:309–312.

8. McMains SA, Somers DC. Multiple spotlights of attentional selection in human visual cortex. *Neuron* 2004, 42:677–686.

9. McMains SA, Somers DC. Processing efficiency of divided spatial attention mechanisms in human visual cortex. *J Neurosci* 2005, 25:9444–9448.

10. Burkell J, Pylyshyn Z. Searching through subsets: a test of the visual indexing hypothesis. *Spat Vis* 1997, 11:225–258.

11. Franconeri SL, Alvarez GA, Enns JT. How many locations can be selected at once? *J Exp Psychol Hum Percept Perform* 2007, 33:1003–1012.

12. McCarley JS, Mounts JRW, Kramer AF. Spatially mediated capacity limits in attentive visual perception. *Acta Psychol (Amst)* 2007, 126:98–119.

13. McCarley JS, Mounts JRW. On the relationship between flanker interference and localized attentional interference. *Acta Psychol (Amst)* 2008, 128:102–109.

14. Bahcall DO, Kowler E. Attentional interference at small spatial separations. *Vision Res* 1999, 39:71–86.

15. Mounts JRW, Gavett BE. The role of salience in localized attentional interference. *Vision Res* 2004, 44:1575–1588.

16. Torralbo A, Beck DM. Perceptual-load-induced selection as a result of local competitive interactions in visual cortex. *Psychol Sci* 2008, 19:1045–1050.

17. Intriligator J, Cavanagh P. The spatial resolution of visual attention. *Cogn Psychol* 2001, 43:171–216.

18. Pylyshyn ZW, Storm RW. Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat Vis* 1988, 3:179–197.

19. Yantis S. Multielement visual tracking: attention and perceptual organization. *Cogn Psychol* 1992, 24:295–340.

20. Luck SJ, Vogel EK. The capacity of visual working memory for features and conjunctions. *Nature* 1997, 390:279–281.

21. Todd JJ, Marois R. Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 2004, 428:751–754.

22. Brady T, Konkle T, Alvarez G. A review of visual memory capacity: beyond individual items and toward structured representations. *J Vis* 2011, 11:1–34.

23. Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* 2001, 24:87–114.

24. Alvarez GA, Franconeri SL. How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *J Vis* 2007, 7:1–10.

25. Franconeri SL, Lin JY, Enns JT, Pylyshyn ZW, Fisher B. Evidence against a speed limit in multiple-object tracking. *Psychon Bull Rev* 2008, 15:802–808.

26. Shim WM, Alvarez GA, Jiang YV. Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychon Bull Rev* 2008, 15:390–397.

27. Tombu M, Seiffert AE. Attentional costs in multiple-object tracking. *Cognition* 2008, 108:1–25.

28. Alvarez GA, Cavanagh P. Independent resources for attentional tracking in the left and right visual hemifields. *Psychol Sci* 2005, 16:637–643.

29. Carlson TA, Alvarez GA, Cavanagh P. Quadrantic deficit reveals anatomical constraints on selection. *Proc Natl Acad Sci USA* 2007, 104:13496–13500.

30. Serences JT, Yantis S. Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. *Cereb Cortex* 2007, 17:284–293.

31. Franconeri SL, Jonathan SV, Scimeca JM. Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychol Sci* 2010, 21:920–925.

32. Pelli DG, Majaj NJ. Crowding is unlike ordinary masking: distinguishing feature integration from detection. *J Vis* 2004, 4:1136–1169.

33. Tsotsos JK, Culhane SM, Wai WYK, Lai Y, Davis N, Nuflo F. Artificial Intelligence Modeling visual attention. *Artif Intell* 1995, 78:507–545.

34. Cutzu F, Tsotsos JK. The selective tuning model of attention: psychophysical evidence for a suppressive annulus around an attended item. *Vision Res* 2003, 43:205–219.

35. Hopf J, Boehler CN, Luck SJ, Tsotsos JK, Heinze H, Schoenfeld MA. Direct neurophysiological evidence for spatial suppression surrounding the focus. *Proc Natl Acad Sci USA* 2006, 103:1053–1058.

36. Scalf PE, Beck DM. Competition in visual cortex impedes attention to multiple items. *J Neurosci* 2010, 30:161–169.

37. Chelazzi L, Duncan J, Miller EK, Desimone R. Responses of neurons in inferior temporal cortex during memory-guided visual search. *J Neurophysiol* 1998, 80:2918–2940.

38. Holcombe AO, Chen W-Y. Exhausting attentional tracking resources with a single fast-moving object. *Cognition* 2012, 123:218–228.

39. Tombu M, Seiffert AE. Tracking planets and moons: mechanisms of object tracking revealed with a new paradigm. *Atten Percept Psychophys* 2010, 73: 738–750.

40. Feria CS. Speed has an effect on multiple-object tracking independently of the number of close encounters between targets and distractors. *Atten Percept Psychophys* 2013, 75:53–67.

41. Vul E, Frank MC, Tenenbaum JB, Alvarez G. Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Adv Neural Inf Process Syst* 2010, 22:1–9.

42. Reynolds JH, Heeger DJ. The normalization model of attention. *Neuron* 2009, 61:168–185.

43. Wei Z, Wang X-J, Wang D-H. From distributed resources to limited slots in multiple-item working memory: a spiking network model with normalization. *J Neurosci* 2012, 32:11228–11240.

44. Howe PDL, Cohen MA, Pinto Y, Horowitz TS. Distinguishing between parallel and serial accounts of multiple object tracking. *J Vis* 2010, 10:1–13.

45. Stormer VS, Winther GN, Li S-C, Andersen SK. Sustained multifocal attentional enhancement of stimulus processing in early visual areas predicts tracking performance. *J Neurosci* 2013, 33:5346–5351.

46. Holcombe AO, Chen W. Splitting attention reduces temporal resolution from 7 Hz for tracking one object to, 3 Hz when tracking three. *J Vis* 2013, 13:1–19.

47. Jans B, Peters JC, De Weerd P. Visual spatial attention to multiple locations at once: the jury is still out. *Psychol Rev* 2010, 117:637–684.

48. Landau AN, Fries P. Attention samples stimuli rhythmically. *Curr Biol* 2012, 22:1000–1004.

49. VanRullen R, Carlson T, Cavanagh P. The blinking spotlight of attention. *Proc Natl Acad Sci USA* 2007, 104:19204–19209.

50. Culham JC, Brandt SA, Cavanagh P, Kanwisher NG, Dale AM, Tootell RBH, Jody C. Cortical fMRI activation produced by attentive tracking of moving targets. *J Neurophysiol* 1998, 80:2657–2670.

51. Culham J, Cavanagh P, Kanwisher N. Attention response functions: characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron* 2001, 32:737–745.

52. Jovicich J, Peters RJ, Koch C, Braun J, Chang L, Ernst T. Brain areas specific for attentional load in a motion-tracking task. *J Cogn Neurosci* 2001, 13: 1048–1058.

53. Howe PD, Horowitz TS, Morocz IA, Wolfe J, Livingstone MS. Using fMRI to distinguish components of the multiple object tracking task. *J Vis* 2009, 9:1–11.

54. Xu Y, Chun MM. Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 2006, 440:91–95.

55. Shim WM, Alvarez GA, Vickery TJ, Jiang YV. The number of attentional foci and their precision are dissociated in the posterior parietal cortex. *Cereb Cortex* 2010, 20:1341–1349.

56. Mccollough AW, Machizawa MG, Vogel EK. Electrophysiological measures of maintaining representations in visual working memory. *Cortex* 2007, 43:77–94.

57. Vogel EK, McCollough AW, Machizawa MG. Neural measures reveal individual differences in controlling access to working memory. *Nature* 2005, 438:500–503.

58. Drew T, Horowitz TS, Vogel EK. Swapping or dropping? Electrophysiological measures of difficulty during multiple object tracking. *Cognition* 2013, 126: 213–223.

59. Drew T, Horowitz TS, Wolfe JM, Vogel EK. Neural measures of dynamic changes in attentive tracking load. *J Cogn Neurosci* 2012, 24:440–450.

60. Drew T, Vogel EK. Neural measures of individual differences in selecting and tracking multiple moving objects. *J Neurosci* 2008, 28:4183–4191.

61. Franconeri SL, Pylyshyn ZW, Scholl BJ. A simple proximity heuristic allows tracking of multiple objects through occlusion. *Atten Percept Psychophys* 2012, 74: 691–702.

62. Keane BP, Pylyshyn ZW. Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cogn Psychol* 2006, 52:346–368.

63. Horowitz TS, Birnkrant RS, Fencsik DE, Tran L, Wolfe JM. How do we track invisible objects? *Psychon Bull Rev* 2006, 13:516–523.

64. Horowitz T, Cohen M. Direction information in multiple object tracking is limited by a graded resource. *Attent Percept Psychophys* 2010, 72: 1765–1775.

65. Shooner C, Tripathy S. High-capacity, transient retention of direction-of-motion information for multiple moving objects. *J Vis* 2010, 10:1–20.

66. Clair RS, Huff M, Seiffert AE. Conflicting motion information impairs multiple object tracking. *J Vis* 2010, 10:1–13.

67. Iordanescu L, Grabowecky M, Suzuki S. Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *J Vis* 2009, 9:1–12.

68. Fencsik DE, Klieger SB, Horowitz TS. The role of location and motion information in the. *Percept Psychophys* 2007, 69:567–577.

69. Howe PDL, Holcombe AO. Motion information is sometimes used as an aid to the visual tracking of objects. *J Vis* 2012, 12:1–10.

70. Treisman AM, Zhang W. Location and binding in visual working memory. *Mem Cognit* 2006, 34:1704–1719.

71. Pylyshyn ZW. Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Vis Cogn* 2004, 11:801–822.

72. Horowitz TS, Klieger SB, Fencsik DE, Yang KK, Alvarez GA, Wolfe JM. Tracking unique objects. *Percept Psychophys* 2007, 69:172–184.

73. Saiki J. Multiple-object permanence tracking: limitation in maintenance and transformation of perceptual objects. *Prog Brain Res* 2002, 140:133–148.

74. Hudson C, Howe PDL, Little DR. Hemifield effects in multiple identity tracking. *PLoS One* 2012, 7:e43796.

75. Franconeri SL, Alvarez GA, Cavanagh P. Flexible cognitive resources: competitive content maps for attention and memory. *Trends Cogn Sci* 2013, 17:134–141.

76. Oksama L, Hyönä J. Dynamic binding of identity and location information: a serial model of multiple identity tracking. *Cogn Psychol* 2008, 56:237–283.

77. Treisman AM, Gelade G. A feature-integration theory of attention. *Cogn Psychol* 1980, 12:97–136.