

Studying visual attention using the multiple object tracking paradigm: A tutorial review

Hauke S. Meyerhoff¹ · Frank Papenmeier² · Markus Huff²

Published online: 5 June 2017
© The Psychonomic Society, Inc. 2017

Abstract Human observers are capable of tracking multiple objects among identical distractors based only on their spatio-temporal information. Since the first report of this ability in the seminal work of Pylyshyn and Storm (1988, *Spatial Vision*, 3, 179–197), multiple object tracking has attracted many researchers. A reason for this is that it is commonly argued that the attentional processes studied with the multiple object paradigm apparently match the attentional processing during real-world tasks such as driving or team sports. We argue that multiple object tracking provides a good mean to study the broader topic of continuous and dynamic visual attention. Indeed, several (partially contradicting) theories of attentive tracking have been proposed within the almost 30 years since its first report, and a large body of research has been conducted to test these theories. With regard to the richness and diversity of this literature, the aim of this tutorial review is to provide researchers who are new in the field of multiple object tracking with an overview over the multiple object tracking paradigm, its basic manipulations, as well as links to other paradigms investigating visual attention and working memory. Further, we aim at reviewing current theories of tracking as well as their empirical evidence. Finally, we review the state of the art in the most prominent research fields of multiple object tracking and how this research has helped to understand visual attention in dynamic settings.

Keywords Visual attention · Dynamic attention · Multiple object tracking · Pylyshyn

During naturalistic tasks, such as supervising children on a playground, watching sport events, or driving a vehicle, the human environment is dynamically changing over time. In many of these situations it is crucial to keep track of multiple independently moving objects over time. In laboratory experiments, this ability can be studied with the multiple object tracking paradigm (MOT; Pylyshyn & Storm, 1988). Within the almost three decades after the seminal work of Pylyshyn and Storm, there was an impressive increase in MOT research. Until today, there are more than 160 journal articles studying a broad variety of different aspects of MOT. Because MOT provides a good mean to study dynamic visual attention, the aim of this tutorial review is to provide researchers without prior experience in MOT with an introduction to MOT research (for more specific reviews, see Scholl, 2009; Scimeca & Franconeri, 2015). For this purpose, we have divided this manuscript into three sections. In the first section, we introduce the MOT paradigm as well as its link to visual attention. In the second section, we present and discuss the evidence of several theories of MOT. Finally, in the third section, we review the state of the art of recent research topics in multiple object tracking and summarize how the described research has fostered the understanding of visual attention. We conclude with a brief outlook into potential future directions of research on visual attention using MOT.

✉ Hauke S. Meyerhoff
h.meyerhoff@iwm-kmrc.de

¹ Leibniz-Institut für Wissensmedien, Schleichstr. 6,
72076 Tübingen, Germany

² Department of Psychology, University of Tübingen,
Tübingen, Germany

Multiple object tracking

The central features of the multiple object tracking task closely match several features of attentionally demanding

naturalistic tasks. Most importantly, the setup of the task is highly dynamic as the objects change their location over time. This transience of information requires a continuous deployment of visual attention in order to avoid confusions between the objects. In return, studying the MOT task thus might help us to understand the basic principles of dynamic visual attention as it operates in many real-world situations. Given the complexity of tracking, it is not surprising that MOT is not only related to other tasks that draw upon the efficiency of visual attention (Huang, Mo, & Li., 2012) but also to processes of attentional selection (Franconeri, Alvarez, & Enns, 2007) and working memory (e.g., Oksama & Hyönä, 2004).

A good way to get started with the MOT paradigm obviously is to experience the tracking task. For this purpose, we have uploaded several video demonstrations of the task (which increase in speed/difficulty; go to https://www.iwm-tuebingen.de/public/realistic_depictions_lab/mot_demo).

In this section, we will introduce the main paradigm of MOT research. Similar to the experiments of Pylyshyn and Storm (1988), most MOT studies have explored tracking performance within a 2-D frame of reference on common lab computers. Although some studies have embedded the MOT task in more immersive setups, such as virtual realities (Lochner & Trick, 2014; Thomas & Seiffert, 2010), large projection screens (Franconeri, Lin, Pylyshyn, Fisher, & Enns, 2008), or three-dimensional scenes (Pylyshyn, Haladjian, King, & Reilly, 2008), all MOT studies follow the same principals. Therefore, we will outline these principles first. We will then review the evidence illustrating that MOT draws upon attentional resources and how these attentional resources alter MOT performance.

The MOT paradigm

Several variations of the MOT paradigm have been published to study the mechanisms of tracking; however, the basic paradigm is similar across almost all MOT experiments. First, several visually indistinguishable objects appear onscreen (typically six to 10 objects). Thereafter, a subset of these objects is designated as target objects (typically three to five targets). The participants are instructed to track this subset of objects across an interval of object motion. The motion paths of the objects (e.g., Brownian motion vs. straight motion paths) as well as object speed (1 to 15 degrees of visual angle) usually vary between studies depending on the actual research question. After the interval of object motion, performance is typically measured either with a probe-one or a mark-all procedure. In a mark-all procedure, participants mark all of the targets that they were able to track and guess the remaining target objects. In contrast, in a probe-one procedure, one of the objects is probed (a target in one half of all trials), and participants decide for this specific object whether it is a target or not. Both procedures are depicted in Fig. 1.

Whether it is more sensible to apply the mark-all procedure or the probe-one procedure depends on the experimental manipulations of the concrete experiment. Due to the higher number of measurements within the same trial, the mark-all procedure is generally more powerful. However, if the experimental manipulations involve tracking load (i.e., the number of targets), a probe-one procedure is preferable because this procedure maintains chance level across different tracking loads even when the total number of objects is constant. As the dependent variable, most studies report the number of correctly identified targets, the proportion of correctly identified targets, or tracking capacity (i.e., the number of actually tracked objects when corrected for guessing). Specific formulas for these calculations have been summarized by Hulleman (2005).

The difficulty level of the MOT task can be varied parametrically (e.g., Alvarez & Franconeri, 2007; Bettencourt & Somers, 2009) up to an almost complete exhaustion of attentional resources. This becomes visible in experiments showing that tracking of a single object could be so exhausting that it is impossible to track a second object (Holcombe & Chen, 2012). Further, tracking also has been demonstrated to interfere with the extraction of the gist of a natural scene in a dual task setup (Cohen, Alvarez, & Nakayama, 2011). Across the large body of research, several variables that influence tracking performance have been identified. Some of these variables are linked to the evaluation of the competing theories of tracking (see section titled Theories of MOT and their evidence). However, besides their theoretical importance, these variables might serve to adjust the difficulty of the MOT task to any required level of difficulty. Typically, MOT performance declines with an increasing number of targets (e.g., Alvarez & Franconeri, 2007; Drew, Horowitz, Wolfe, & Vogel, 2011; Pylyshyn & Storm, 1988), an increasing number of distractors (e.g., Bettencourt & Somers, 2009; Sears & Pylyshyn, 2000), an increasing trial duration (e.g., Oksama & Hyönä, 2004), an increasing proximity between the objects in the display (e.g., Bettencourt & Somers, 2009; Franconeri, Jonathan, & Scimeca, 2010), as well as an increasing object speed (e.g., Holcombe & Chen, 2012; Meyerhoff, Papenmeier, Jahn, & Huff, 2016; Tombu & Seiffert, 2011). Please note that this list represents the most common manipulations only and makes no claim to be complete.

Situating MOT in research on visual attention

Although the MOT paradigm was not designed as an attentional paradigm initially, there is overwhelming behavioral, electrophysiological, and neuroimaging evidence that MOT draws heavily on attentional resources. For instance, Tombu and Seiffert (2008) demonstrated the attentional demands of MOT in a dual task setup. In their study, intervals of increased difficulty such as a spontaneous attraction of the objects or a

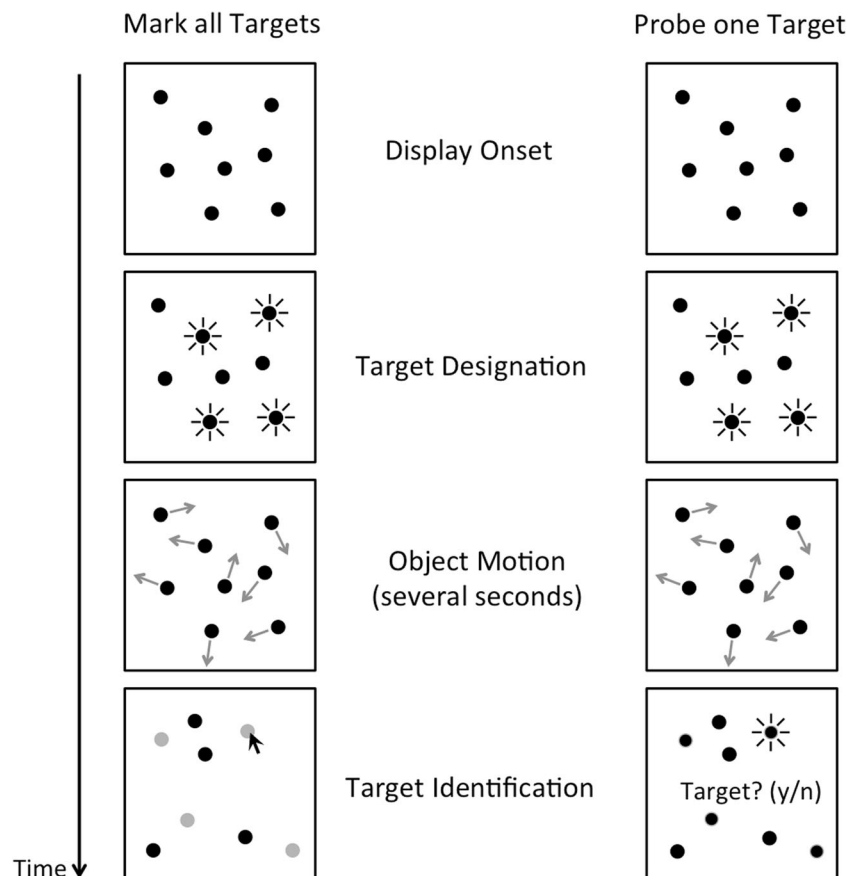


Fig. 1 Illustration of the procedure of multiple object tracking experiments. A subset of visually indistinguishable objects is designated as targets. Following target designation, the objects move for several seconds. Then the objects stop moving and participants

either mark all objects (mark all paradigm; *left column*) or indicate for one probed object whether this object was a target or a distractor (probe one paradigm; *right column*)

temporarily increase in object speed interfered with a tone discrimination task that was temporally aligned with these intervals (see Fig. 2). In a similar study, Kunar, Carter, Cohen, and Horowitz (2008) showed that MOT does not only interfere with other attentional demanding tasks but that tracking limitations also arise at more central stages of information processing. In their experiments, the generation of words during a telephone conversation (with the experimenter) impaired MOT performance. The close link between MOT and visual attention also becomes evident in a remarkable study by Huang et al. (2012), who investigated the interrelations between numerous attentional paradigms. In this study, MOT did not only correlate with all other attentional paradigms that draw upon attentional efficiency, but was also correlated with a general factor of visual attention that was extracted from the intercorrelations of the individual tasks.

Remarkably, however, participants are able to interrupt the tracking task for a few hundred milliseconds in order to perform a concurrent dual task. Such a rapid task switching was reported by Alvarez, Horowitz, Arsenio, DiMase, and Wolfe (2005) who observed better tracking performance with a concurrent search task than would be expected by shared

attentional resources. Even further, tracked objects do not need to be continuously visible during tracking as long as occlusion cues signal their disappearance (Scholl & Pylyshyn, 1999) or a global offset triggers memory processes (Horowitz, Birnkrant, Fencsik, Tran, & Wolfe, 2006). Further evidence for the role of task switching and memory processes during MOT arises from correlations between individual differences in MOT and task switching as well as working memory (Oksama & Hyönä, 2004). In general, previous research has established a strong connection between spatial attention and spatial working memory (Awh, Anillo-Vento, & Hillyard, 2000; Awh & Jonides, 2001; Smyth & Scholey, 1994). Therefore, a link between MOT and working memory is not too surprising (see Allen, McGeorge, Pearson, & Milne, 2006; Zhang, Xuan, Fu, & Pylyshyn, 2010). In reverse, a concurrent MOT task also disturbs processes of working memory such as feature binding (Fougnie & Marois, 2009). However, there is also clear evidence for a functional dissociation of tracking and memorizing (Carter et al., 2005; Fougnie & Marois, 2006).

In line with intuition, the MOT task requires visual selection as well as sustained visual attention to track the target objects (Ma & Flombaum, 2013). Although visual selection typically

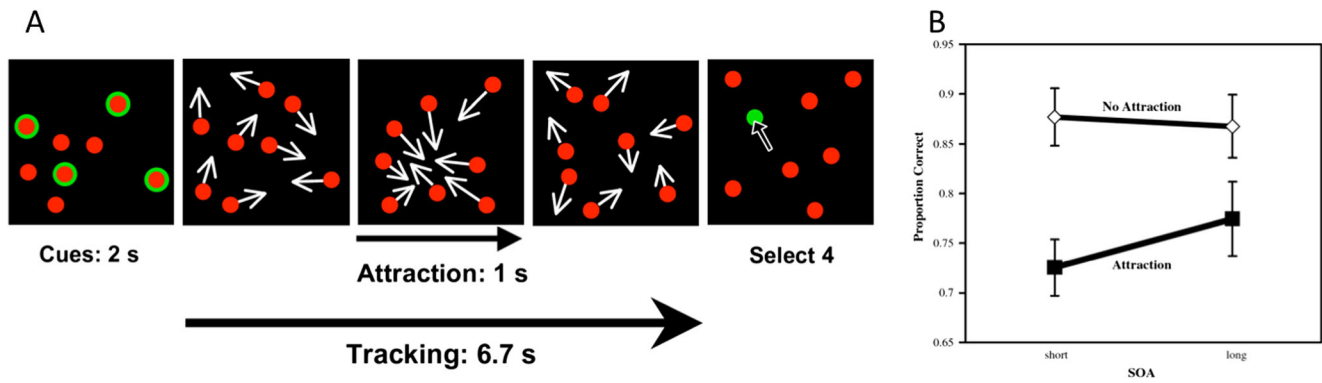


Fig. 2 Demonstration of attentional demands during multiple object tracking (Tombu & Seiffert, 2008). **a** The participants track four objects over time. During the trial the objects spontaneously attract each other (and/or increase in speed). Concurrently, the participants perform a tone pitch discrimination task. **b** Multiple object tracking performance as a

function of interdot attraction and the stimulus onset asynchrony (SOA) between the attraction interval and the dual task. Dual-task interference was more pronounced with short SOAs. Figures reproduced with permission from Elsevier

occurs at the beginning of the trials, these two processes do not mutually exclude each other. During tracking, observers are able to deselect previously tracked objects and select new target objects (multiple object juggling) without sacrificing accuracy (Wolfe, Place, & Horowitz, 2007; see also Ericson & Christensen, 2012; Pylyshyn & Annan, 2006). The distinction between selection and tracking has also been corroborated by electrophysiological and neuroimaging studies. For instance, in an event-related potential (ERP) approach, Drew and Vogel (2008) observed that increasing demands of the target selection process (i.e., prior to motion onset) were reflected in an increasing negativity roughly 200 ms after stimulus onset over the posterior electrodes (N2pc; see also Hopf et al., 2006; Woodman & Luck, 2003, for converging evidence from visual search). Increasing load during the tracking period was reflected in the contralateral delay activity (CDA; Drew et al., 2011). Because CDA amplitude is a better predictor for tracking performance than N2pc amplitude in trials of typical duration (Drew & Vogel, 2008), limitations in tracking performance arise from tracking itself rather than from visual selection. This matches the behavioral observation that the capacity limitation for visual selection (Franconeri et al., 2007) is higher than the limitation for tracking moderately fast-moving objects (e.g., Alvarez & Franconeri, 2007).

Importantly, Drew et al. (2011) were able to show that the load sensitivity of the CDA distinguishes attentive tracking from working memory tasks (see Vogel & Machizawa, 2004; Vogel, McCollough, & Machizawa, 2005). Although the CDA varied with the load manipulation in both types of tasks, it was much more pronounced in tracking trials than in memory trials. In fact, the CDA was even reduced when the tracked objects stopped moving. This observation agrees with the assumption that tracking moving objects requires sustained visual attention and cannot be reduced to a passive memorization of object locations. In further work, Drew, Horowitz, and Vogel (2013) took advantage of the load sensitivity of the CDA in order to disentangle different

kinds of tracking errors. Hereby, the CDA amplitude indicated that an increase in the number of distractors causes confusion between targets and distractors (i.e., swaps), whereas increasing object speed results in dropping objects that were to be tracked.

In general, the electrophysiological evidence matches neuroimaging studies that have identified several cerebral areas that contribute to MOT, such as the intraparietal sulcus (IPS), the superior parietal lobule (SPL), the frontal cortex such as the frontal eye fields (FEF), the precentral sulcus (PreCS), and the motion sensitive areas in the MT+ complex (Culham et al., 1998). Two of the neuroimaging studies aimed at distinguishing areas that are sensitive to tracking load from areas that respond to the tracking task in general (Culham, Cavanagh, & Kanwisher, 2001; Jovicich et al., 2001). Across both studies, IPS as well as PreCS were sensitive to the manipulation of attentional load (see also Howe, Horowitz, Morocz, Wolfe, & Livingstone, 2009). This agrees with previous research that has highlighted the role of the IPS for spatial attention (e.g., Coull & Frith, 1998). Surprisingly, evidence for the contribution of early visual areas to MOT performance is remarkably absent in the neuroimaging studies. However, electroencephalography (Störmer, Winther, Li, & Andersen, 2013) as well as eye tracking (i.e., pupil dilation; Alnaes et al., 2014) have provided evidence that early visual processing also predicts MOT performance.

How attention modulates MOT

Whereas the abovementioned studies clearly show that MOT requires attention, the question how attentional resources contribute to tracking is more controversial. The first studies that provide insights into the contributions of attention to tracking typically studied dual task setups in which participants performed a probe detection task besides tracking (e.g., Pylyshyn, 2006). The principal idea of these experiments is that the allocation of visual attention toward target objects should enhance probe detection performance on targets

relative to distractors. Indeed, this prediction was confirmed across a wide range of experiments (Huff, Papenmeier, & Zacks, 2012; Pylyshyn, 2006; Pylyshyn et al., 2008; Sears & Pylyshyn, 2000). A remarkable observation, however, was that probes that appeared on the empty background elicited higher detection rates than probes that appeared on distractors, suggesting that distractors are suppressed during tracking (Pylyshyn, 2006). The extent of this suppression depends on the similarity between targets and distractors in terms of motion and form (Feria, 2012) as well as depth plane (Pylyshyn et al., 2008; see also Rehman, Kihara, Matsumoto, Ohtsuka, 2015; Viswanathan & Mignolla, 2002).

A central challenge for the interpretation of these dual task experiments is that the probe detection task might have affected the allocation of visual attention in the MOT task. In order to avoid this fallacy, Drew, McCollough, Horowitz, and Vogel (2009; see also Stemshein & Sekuler, 2011) recorded ERPs elicited by task-irrelevant probes that appeared on targets, distractors, stationary distractors, or the empty background. They observed more pronounced amplitudes of the N1 and the P1 (see Luck, 1995; Luck et al., 1994) for probes that appeared on targets relative to moving distractors or empty space but no evidence for distractor suppression. Even further, the magnitude of ERPs signaling target enhancement predicted tracking performance on the behavioral level. In other words, good and poor trackers could be distinguished based on their neural response (see Fig. 3).

It is noteworthy that Doran and Hoffman (2010) observed evidence for distractor suppression in the N1 amplitude. There were several differences in the stimulus design that might be responsible for the contrasting patterns of results. However, this study points out that target enhancement and distractor suppression might occur in parallel during tracking. Because the study by Drew et al. (2009) indicates that target enhancement is capable of arising without distractor suppression, both mechanisms might reflect functionally independent processes. On the behavioral level, target enhancement and distractor suppression seems to warp the perceived spacing between the moving objects (Liverence & Scholl, 2011). Nevertheless, suppression does not blank out distractors completely, as demonstrated by studies showing that distractor locations are represented above chance level (Alvarez & Oliva, 2008), that repeating motion paths of targets as well as distractors enhances tracking performance (Ogawa, Watanabe, & Yagi, 2009), and that displacing distractors impairs tracking even when the displacements maintain the spacing between target and distractors (Meyerhoff, Papenmeier, Jahn, & Huff, 2015).

Theories of MOT and their evidence

Several theoretical frameworks have been proposed in order to explain limitations in MOT. A central difference in these

theories is whether they propose fixed architectural constraints, such as slots, or a limited attentional resource that is allocated among objects being tracked. The theories also differ in the role that is attributed to visual attention. Although recent tendencies seem to favor models without architectural constraints, the empirical evidence is still conflicting. Thus, the evaluation of the MOT theories is still a subject of change. In this section, we will now review the theories in the order of their publication dates.

Visual index theory (FINST theory)

The basic assumption of the visual index theory (also called FINST theory; Pylyshyn, 1989, 2001, 2007) is the existence of a visual index mechanism in early vision that provides a connection between objects in the distal world and the visual representation in the mind (see Fig. 4a). This connection is achieved by preconceptual visual indexes (also called FINSTs, derived from FINGers of INSTantiation) that point and stick to feature clusters on the retina (Pylyshyn, 1989); that is, they provide a reference to objects in the scene. They are characterized as preconceptual because the visual indexes provide a reference to objects (like the referential “this” or “that” in speech) without encoding information about their identities. While this property was called “preattentive” in earlier work (Pylyshyn, 1989), this term was replaced by the term “preconceptual” in later work (Pylyshyn, 2001) to avoid misunderstandings and make clear that focused attention can also play a role during tracking (Pylyshyn, 2001). The visual indexes stick to the indexed objects across motion or eye movements and thereby allow for an automatic tracking of multiple objects in parallel, even if they look identical. Because visual indexes are a part of a mental architecture, their number is limited, presumably at about four or five (Pylyshyn, 2001). The visual index theory is not a theory of MOT in particular but a theory of vision in general, with findings beyond MOT, such as visual search or subitizing (Pylyshyn, 1994; Pylyshyn et al., 1994) supporting its assumptions.

Originally, the MOT paradigm was designed to test the prediction of the visual index theory that observers can track multiple objects at a higher accuracy than predicted by a serial spotlight metaphor (Pylyshyn & Storm, 1988). It is important to note, however, that participants’ response times for the detection of flashes on targets increased with the number of targets in the display. That is, participants could track multiple objects in parallel but required covert attention to serially scan the indexed objects for feature changes because the visual indexes are merely pointers that do not provide feature information by themselves. Furthermore, Sears and Pylyshyn (2000) showed that a zoom-lens model of attention could also not account for the distribution of attention during MOT, because response latencies for form changes on distractors in a

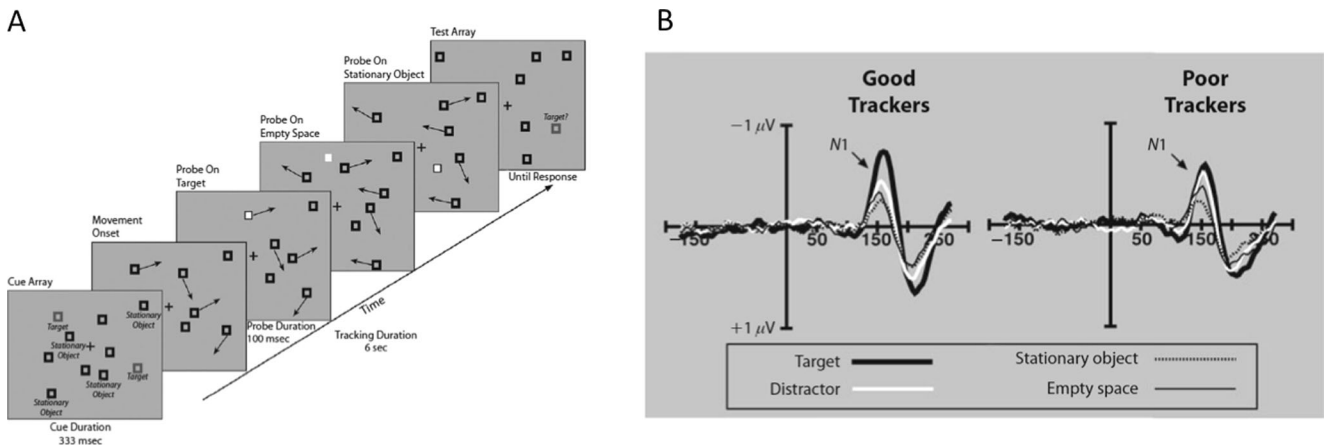


Fig. 3 Illustration of a study conducted by Drew, McCollough, Horowitz, and Vogel (2009) addressing the role of attentional enhancement during multiple object tracking. **a** During the tracking task, task-irrelevant probes appear on target objects, distractor objects, stationary

objects, or the empty background. **b** *Good trackers* show a stronger electrophysiological response to task-irrelevant probes on targets than *poor trackers*, signaling attentional enhancement. Figures reprinted with permission from Springer

secondary task did not differ between distractors within or outside the convex hull formed by the target objects.

Regarding object selection, the visual index theory states that objects grab the visual indexes in a data-driven, parallel, and automatic manner, such as by new objects appearing in the visual field or by having objects that blink (Pylyshyn, 2001). In contrast, a voluntary allocation of visual indexes to objects requires focused attention such that local features can then pop out and attract a visual index (Pylyshyn, 2001). Results found by Pylyshyn and Annan (2006) support this assumption by showing that participants could track voluntarily selected objects (e.g., marked by digits) but that voluntary selection was sensitive to the duration of the marking phase and the number of target objects present in the display while automatic selection (blinking) was not.

A number of findings have challenged the visual index theory. Those studies were mainly concerned with the role of attention during MOT. For example, they showed that attention is utilized during MOT (Tomblu & Seiffert, 2008) or that separate tracking resources are available for the left and right visual hemispheres (Alvarez & Cavanagh, 2005; see also section titled Multifocal Attention Theory). Furthermore, it was argued that tracking is achieved by a flexible resource rather than by a fixed architecture allowing participants to track up to eight objects when these objects are slow enough (Alvarez & Franconeri, 2007; see also the section titled FLEX Model). Increases in object speed can also result in tracking capacities far below the assumed four or five visual indexes (Alvarez & Franconeri, 2007; Holcombe & Chen, 2012). While those findings speak against the strong view that tracking is achieved solely by the visual index mechanism, the question whether or not mechanisms of visual indexing and attention might co-occur during MOT (Pylyshyn, 2001) needs to be resolved.

Perceptual grouping model

Yantis (1992) proposed that the visual system groups the individual target objects into a higher order visual representation (i.e., three targets into a triangle or four targets into a quadrangle; see Fig. 4b). Evidence for this proposition comes from experiments that investigated grouping processes during either group formation or group maintenance. Although tracking performance benefits from manipulations that foster the formation of perceptual groups such as canonical configurations or explicit grouping instructions, Yantis (1992) has shown that these effects are of short duration. He therefore suggested that group formation is automatic and preattentive. In contrast, group maintenance during tracking was shown to be effortful. For instance, Yantis (1992) reported more accurate tracking when the higher order object (i.e., the polygon formed by the individual objects) remained intact as compared to when it collapsed during the trial.

Grouping processes during MOT are also affected by identity information (Erlikhman, Keane, Mettler, Horowitz, & Kellman, 2013; see also Zhao et al., 2014). Compared to a condition in which all targets shared the same feature, tracking performance was impaired when identity information divided the objects into two groups, each containing two targets and two distractors. Evidence for the importance of abstracted higher order representations in MOT comes from studies that emphasized the importance of the centroid of the target objects. If multiple targets are integrated into a higher order object such as the polygon connecting these targets, the centroid seems to reflect a plausible instance of this higher order object in terms of a summary statistic of the locations of the individual targets. In fact, Alvarez and Oliva (2008) showed that the location of the centroid is represented above chance level during tracking.

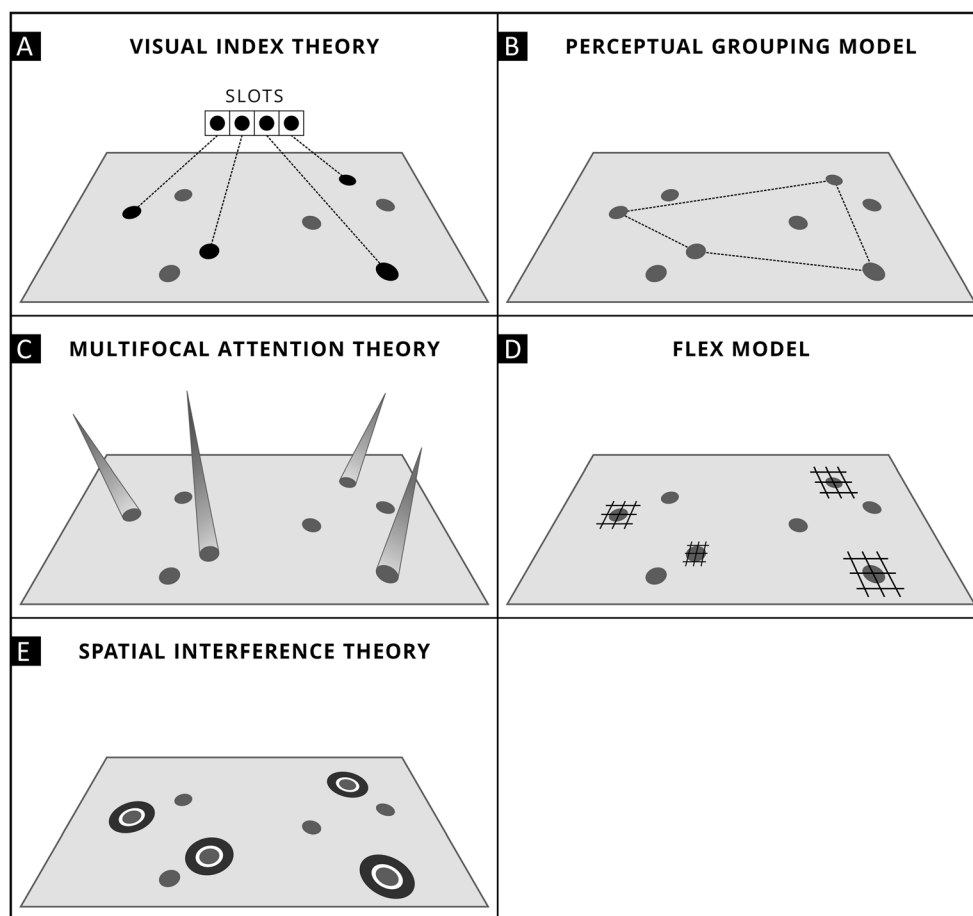


Fig. 4 MOT theories. **a** According to the visual index theory, four or five indices that provide a pointer toward the actual object locations remain “sticking” onto moving objects. **b** The perceptual grouping model states that observers track the higher order object (i.e., the *polygon*) formed by the individual objects. **c** In the multifocal attention theory, independent attentional spotlights track the target objects. **d** The FLEX model

proposes a demand based and dynamically changing allocation of visual attention toward target objects (e.g., based on interobject spacing). **e** According to the spatial interference theory, distractors that break through an inhibitory zone around targets increase the probability of tracking errors. Also, the inhibitory surround of one individual target is considered to interfere with the enhancement of other nearby targets

This observation matches findings from studies that monitored eye movements during MOT. These studies revealed that observers tend to fixate the (invisible) centroid rather than the individual objects during tracking (Fehd & Seiffert, 2008). Within observers, oculomotor processes are remarkably stable across repetitions of trials (Lukavský, 2013; see also Lukavský & Děchtěrenko, 2016), however, properties of the moving objects themselves have been demonstrated to alter fixation behavior. For instance, the tendency to fixate the centroid increases with increasing object speeds (Huff, Papanmeier, Jahn, & Hesse, 2010) but decreases with increasing tracking load (Zelinsky & Neider, 2008) or reduced target-distractor spacing (rescue saccades toward the individual targets that are in danger of getting lost; Zelinsky & Todor, 2010; see also Colas, Flacher, Tanner, Bessiere, & Girard, 2009). This overall pattern of results matches the idea of an automatic (and thus stimulus driven) formation of perceptual groups during tracking. In line with the observation of Yantis (1992) that perceptual grouping fosters MOT performance,

centroid looking behavior is predictive for successful tracking (Fehd & Seiffert, 2010).

Multifocal attention theory

Both the FINST as well as the grouping theory involve a single focus of visual attention. As an alternative, Cavanagh and Alvarez (2005) proposed the multifocal attention theory of MOT. This model suggests that multiple foci of attention follow the objects being tracked across the tracking trial (see Fig. 4c). In this sense, the multiple foci of attention serve a similar function as the visual indices within the FINST theory. The critical difference to the FINST model, however, is that the multiple foci of attention allow for continuous attentional access to all objects being tracked. Indeed, there is convincing behavioral (Awh & Pashler, 2000; Castiello & Umiltà, 1992; Kramer & Hahn, 1995) as well as neuroscientific evidence (McMains & Somers, 2004; Müller, Malinowski, Gruber, &

Hillyard, 2003) that the attentional focus can be split, thus enhancing the processing of stimuli at independent locations.

There are at least two empirical findings that support the multifocal theory of MOT. The first line of evidence stems from hemispherical independence during tracking. For instance, Alvarez and Cavanagh (2005) observed that participants were able to track twice as many objects when they were equally distributed across the left and right visual hemifields (see also Battelli et al., 2001; Chen, Howe, & Holcombe, 2013; Hudson, Howe, & Little, 2012; Störmer, Alvarez, & Cavanagh, 2014). In other words, they observed that the capacity limitation of MOT is two objects per hemifield rather than four objects for the entire visual field. Indeed, the observation of hemifield independence rules out a simple switching model with a single focus of attention (see also Delvenne, 2005). The second finding supporting the multifocal theory is that tracking performance is more accurate when all objects move simultaneously than when they move sequentially (Howe, Cohen, Pinto, & Horowitz, 2010). Because a single focus of attention that switches between the target objects would predict the opposite pattern of results, this finding also supports the assumption of multiple foci of visual attention.

Despite the evidence in favor of parallel tracking as suggested by multiple attentional spotlights, there also is evidence highlighting serial components during tracking. For instance, Howard, Masom, and Holcombe (2011; see also Howard & Holcombe, 2008) observed a lag between actual and remembered object locations that increased with tracking load. This observation is in line with a serial updating process that returns to an individual object less often at higher tracking load. Furthermore, the idea of serial updating also fits in with other studies that demonstrate decreasing temporal resolution of visual attention with increasing tracking load (d'Avossa, Shulman, Snyder, & Corbetta, 2006; Holcombe & Chen, 2013). In fact, Holcombe and Chen (2013) showed that the temporal resolution of tracking dropped from 7 Hz when tracking a single target to 4 Hz when tracking two objects to 2.6 Hz when tracking three objects. Because this decline closely matches the computational predictions of a single focus of attention switching between targets, this finding supports serial models of object tracking rather than the multifocal attention account.

It is, however, currently unclear whether the load-dependent decrease in the temporal resolution spreads across the distinct hemifields or whether each hemifield has a distinct temporal resolution at command. Such a study would be useful to distinguish between accounts that suggest parallel tracking between, but serial tracking within, distinct hemispheres and accounts that predict a serial limitation across the entire visual field (see also Chen et al., 2013). From the current work, it seems plausible that tracking occurs in parallel between the different hemifields but serially within each hemifield.

FLEX model

The previously introduced theories of MOT describe limitations in the ability to track objects as a consequence of architectural constraints such as the number of visual indices or attentional foci. Critically, such fixed architecture models predict a fixed precision of tracking as well as accuracy as long as the number of targets does not exceed the architectural constraints. In a study by Alvarez and Franconeri (2007), however, observers were able to track up to eight objects when they were moving at sufficiently slow speed. Because the detrimental effects of object speed also were more pronounced at reduced spacing between the objects, the overall pattern of results indicates that it is the speed of objects or their spatial interference that limits tracking rather than tracking load in terms of architectural constraints. Based upon this observation, Alvarez and Franconeri introduced the FLEX model of MOT, which proposes a flexible allocation of an attentional resource between the objects being tracked. According to this model, tracking errors arise when the attentional resource is insufficient to cover the demands of all targets. Critically, the actual demand of a target being tracked varies with stimulus properties such as spatial proximity or object speed. Because spatial proximity between the objects varies across a tracking trial, the demand-based allocation of visual attention also needs to change continuously across the tracking interval (see Fig. 4d).

Supporting evidence for the demand-based allocation of visual attention comes from a study by Iordanescu, Grabowecky, and Suzuki (2009), who asked their participants to localize target discs that had disappeared after an interval of object tracking. When Iordanescu et al. analyzed the localization errors as a function of the distance between the corresponding targets and their closest distractor, they observed that targets with close distractors were localized more precisely than those without close distractors. This pattern of results indicates that indeed more attentional resources are devoted to targets with close distractors. Horowitz and Cohen (2010) investigated whether the process underlying MOT is limited by fixed architectural constraints or a flexible resource by applying the mixture model approach previously used within the slot versus resource debate in visual working memory (Zhang & Luck, 2008). By doing so, they observed that the precision of participants in reporting the motion direction of multiple moving targets decreased with an increasing tracking load. In fact, this decline matched the predictions of a model assuming an attentional resource being shared among all items being tracked. Additionally, Holcombe and Chen (2012) demonstrated that a single target (moving on a circular path) is capable of consuming the full tracking resources. If this target moves fast enough, adding a second target results in a tracking performance that matches the expected performance that appears as if the observer were able to track only

one of the two targets. This observation also cannot be reconciled with fixed architecture models.

An advantage (and a drawback at the same time) of the FLEX model is that it can explain a broad variety of findings across a broad spectrum of the tracking literature such as a flexible switching between tasks (Alvarez et al., 2005) or flexible switching between location and identity tracking (e.g., Cohen, Pinto, Howe, & Horowitz, 2011). However, a very strong argument against the FLEX model is that it does not fulfill the criteria of a good scientific theory. There is hardly any pattern of results that cannot be resolved within the FLEX framework. The major reason for this deficit is that the FLEX model is rather vaguely specified. Similar to the multifocal attention approach, the existence of hemifield independence suggests that there are two distinct resources rather than one. This is most evident in a set of experiments of Chen et al. (2013). Although these authors observed evidence in favor of a flexible allocation of an attentional resource in general, this was only true when the corresponding objects were within the same visual hemifield.

Spatial interference theory

As a testable alternative to the FLEX model, Franconeri et al. (2010) proposed the spatial interference theory of MOT. According to this model, objects being tracked receive attentional enhancement that is accompanied by an inhibitory surround (see Hopf et al., 2006; Müller, Mollenhauer, Rösler, & Kleinschmidt, 2005). Tracking errors arise when distractor objects break through the inhibitory surrounds of targets or when the inhibitory surround of one target interferes with the attentional enhancement of another target (see Fig. 4e). Because these events occur only at reduced interobject spacing, spatial interference from close objects is supposed to be the only limiting factor of multiple object tracking performance in this approach (see also Intriligator & Cavanagh, 2001). Indeed, the experiments of Franconeri et al. (2010; see also Franconeri et al., 2008) have shown that it is the distance that objects travel in close proximity to other objects rather than trial duration that constrains tracking performance. Importantly, most of the previous parameters that have been identified to alter tracking performance such as number of targets (e.g., Pylyshyn & Storm, 1988), the number of distractors (e.g., Bettencourt & Somers, 2009), or object speed at a constant tracking interval (e.g., Alvarez & Franconeri, 2007) can be retraced to modulations of the spatial proximity between the moving objects.

Despite its parsimony, the spatial interference account captures a large portion of the variance in tracking performance and the outstanding influence of spatial interference on tracking is corroborated by several studies. For instance, tracking performance increased

markedly when Bae and Flombaum (2012) briefly colored distractors during events of spatial interference (see Fig. 5). Further, Shim, Alvarez, and Jiang (2008) demonstrated that not only proximity between targets and distractors but also proximity among targets impairs tracking performance—a finding that is hard to reconcile with the other theories on MOT. The spatial interference theory also receives support from recent computational models that have conceptualized tracking errors as a result of probabilistic processes and their summation across the duration of a trial (Zhong, Ma, Wilson, Liu, & Flombaum, 2014, see also Vul, Frank, Alvarez, & Tenenbaum, 2009).

The clear predictions of the spatial interference account have inspired much research on the relationship between object speed and spatial proximity during tracking. Whereas the spatial interference account states that object speed per se has no influence on tracking performance beyond the modulation of spatial interference, a couple of recent studies have revealed data that are incompatible with this view. For instance, in Tombu and Seiffert (2011), participants tracked objects that were rotating around each other on an orbital path. Although the orbital rotation left spatial interference unaffected, tracking performance declined with orbital speed (see also Feria, 2013, for similar results). Further, Holcombe, Chen, and Howe (2014) observed that also objects that are well beyond the range of inhibitory surrounds are capable of interfering with the tracking task. Finally, in one of our own studies, we asked participants to track objects that dynamically changed their speed of motion following a sine wave pattern (Meyerhoff et al., 2016). In the control condition, the objects moved at a constant speed that matched the average of the dynamically changing condition. Thus, traveled distance as well as spatial interference was controlled for. The only difference between the conditions was that the events of spatial interference were less predictable and their duration was more variable in the condition with dynamic speed changes than in the condition with constant object speed (see Fig. 6). Tracking performance was worse with dynamically changing object speed, indicating that speed is capable of impairing MOT. One possible solution to resolve the immediate effects of speed within the spatial interference theory is to assume that the inhibitory surround of targets needs to unfold over time. In return, fast moving distractors might interfere with the objects being tracked before the inhibition is fully established. Given the liveliness of the debate about the effects of speed and spatial interference, we expect to see further theoretical progress with regard to the spatial interference theory within the near future.

Research topics in MOT

Besides the research that explicitly aimed at evaluating the theories of MOT, other lines of experiments have explored

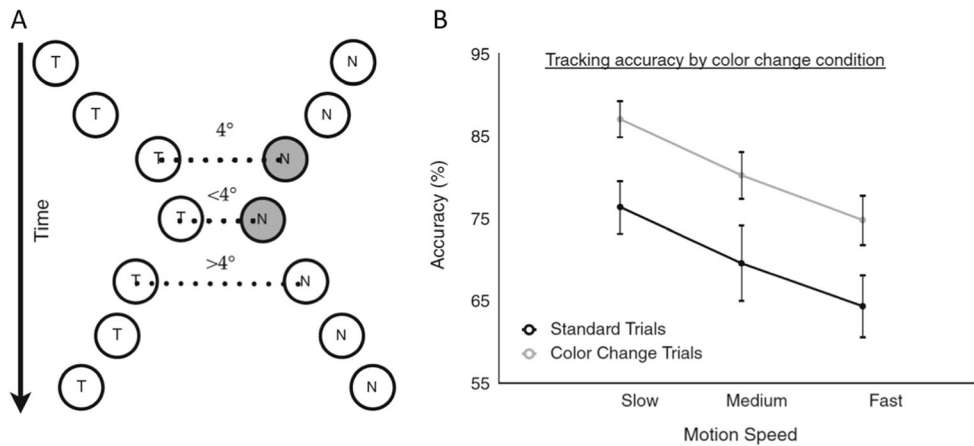


Fig. 5 Demonstration of the detrimental effects of spatial proximity on multiple object tracking performance (by Bae & Flombaum, 2012). **a** In the experimental trials, distractor objects that are close to targets change in color. **b** Coloring distractor objects that are close to targets enhances

tracking performance as these distractors are typically involved in target-distractor confusions. Figures reprinted with permission from Springer

topics with a broader scope. For the purpose of this tutorial review, we have identified four such fields of research that used the MOT paradigm in order to study broader properties of visual attention. These fields encompass questions about the basic unit of dynamic visual attention, the reference frame of dynamic attention, whether attentional processes use motion information to anticipate prospective object locations, and how distinct identity information affects attentive tracking. In addition, we briefly summarize the growing literature on tracking in special groups such as children, experts, and clinical samples.

MOT reveals the object-based nature of dynamic visual attention

The allocation of attentional resources is either space based (Posner, 1980; Posner, Snyder, & Davidson, 1980) or object based (Egley, Driver, & Rafal, 1994; Scholl, 2001). This paragraph reviews studies on MOT that provide evidence that tracking and thus the allocation of dynamic visual attention toward moving objects is mostly object based.

Scholl, Pylyshyn, and Feldman (2001; see Fig. 7) demonstrated the object-based nature of MOT with a target-distractor

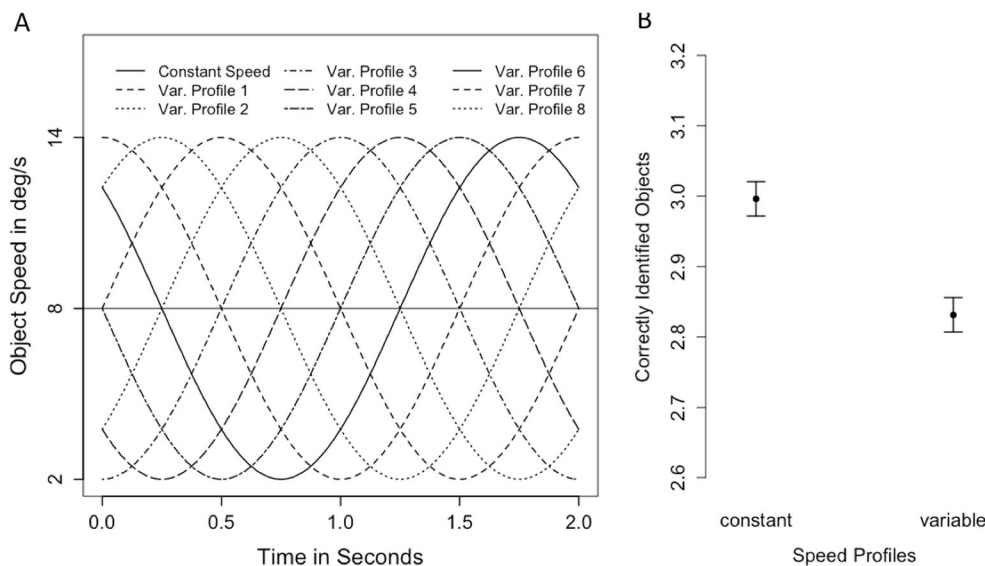


Fig. 6 Demonstration of the influence of objects speed beyond effects of spatial proximity (by Meyerhoff, Papenmeier, Jahn, & Huff, 2016). **a** The participants tracked four out of eight objects that moved at a constant or variable object speed. Importantly, the traveled distance and average speed as well as spatial proximity was identical between the conditions.

b Variable object speed impaired tracking performance indicating that object speed affects tracking beyond the modulation of effects of spatial interference. Figures reproduced with permission from the American Psychological Association

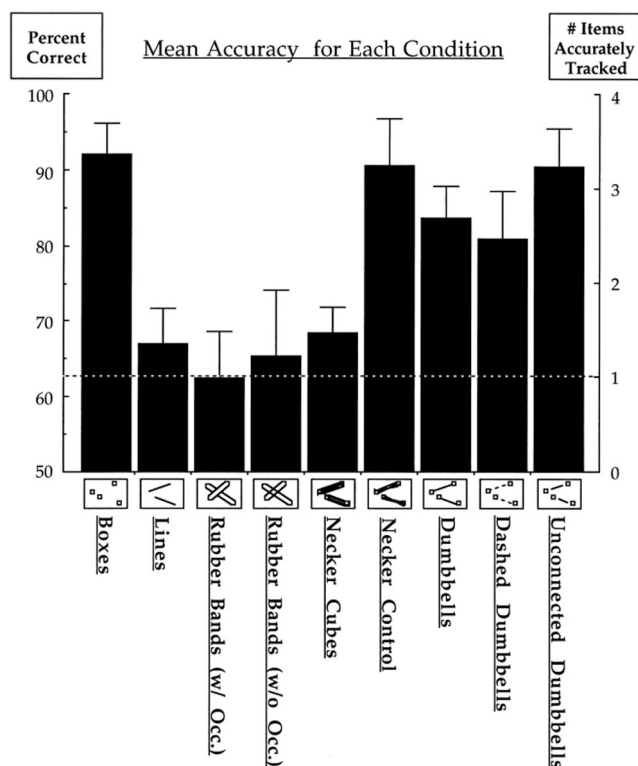


Fig. 7 Demonstration of the object-based nature of multiple object tracking (by Scholl, Feldman, & Pylyshyn, 2001). The participants were able to track the objects accurately only when they appeared as distinct objects. Tracking the endpoint(s) of larger objects reduced tracking capacities to one objects. Figure reprinted with permission from Elsevier

merging technique. In their study, targets and distractors were either distinct boxes or each target was visually merged to a distractor. To be more concrete, targets and distractors either were the endpoints of a common object such as a line (object-based merging) or they were connected so that they appeared as dumbbells (connection-based merging). As predicted by an object-based account of attention, object-based merging severely impaired tracking performance. Connection-based merging impairments were smaller and occurred only if the connection lines touched the targets and distractors. This shows that participants had difficulties in directing their attention to just one endpoint of connected objects. Merging effects occurred also when the size of the merged objects remained constant across the tracking trials (Howe, Incledon, & Little, 2012) instead of expanding and shrinking across a trial which in itself impairs tracking performance (van Marle & Scholl, 2003). Remarkably, a physical connection between target and distractor is not necessary to induce detrimental effects on tracking because illusory contours have been demonstrated to impair tracking performance similarly (Keane, Mettler, Tsoi, & Kellman, 2011). While the effect of connection-based merging was less pronounced in children with autism (Evers et al., 2014), indicating a relatively higher amount of

local processing in this population, object-based merging affected this population equally (Van der Hallen et al., 2015).

Beyond merging, further studies demonstrated the object-based nature of MOT. In line with an account that objects are defined by topological invariants such as the number of holes in the object shape, participants tracking performance was impaired when objects frequently changed their shape between two instances varying in their number of holes such as filled disc versus disc with hole (Zhou, Luo, Zhou, Zhuo, & Chen, 2010). However, nontopological changes such as shape changes between filled discs and S-shapes or constant filled discs with changing color did not impair tracking performance. Note that it is not the holes per se that impair tracking because participants can track shapes with holes (Zhou et al., 2010) and holes (Horowitz & Kuzmova, 2011) as efficient as solid discs. Instead, the change in topological invariants likely caused the formation of a new object thus disturbing tracking. While abrupt transition in object shape between small squares and long rectangles did not affect tracking (Zhou et al., 2010), continuous transitions did (Howe, Holcombe, Lapierre, & Cropper, 2013; van Marle & Scholl, 2003). However, in this case it was not the formation of new objects that impaired tracking but the reduced ability to locate objects that expand and contract, particularly along the axis of elongation or contraction (Howe et al., 2013).

By asking participants to track lines instead of discs or squares, a number of studies investigated the allocation of attention across tracked objects (Alvarez & Scholl, 2005; Doran, Hoffman, & Scholl, 2009; Feria, 2008). These studies explored detection performance for probes appearing either in the center or near the endpoints of the tracked lines. Because probe detection was more efficient in the center than at the endpoints, these studies suggest an attentional bias toward the center of the objects being tracked (Alvarez & Scholl, 2005; Doran et al., 2009; Feria, 2008). This center benefit was not the result of overt attention shifts being biased toward object centers (Vishwanath & Kowler, 2003) because it was also observed when eye movements were controlled for (Doran et al., 2009). Although the center bias is strong and increases with line length, it is not completely automatic because it is sensitive to the distribution of probe probabilities between the center and endpoints (Feria, 2008, 2010).

The reference frame of visual attention

What is the reference frame of (dynamic) visual attention? Does attentional tracking operate within a retinotopic coordinate system (i.e., relative to the corresponding locations on the retina) or within an allocentric coordinate system (i.e., scene-based coordinates)? These questions are of interest because the visual cortex is mostly organized in retinotopic maps (DeYoe et al., 1996; Engel, Glover, & Wandell, 1997; Lennie, 1998; Van Essen et al., 2001); however, mental

representations of scenes appear to be in allocentric coordinates rather than retinotopic coordinates (e.g., Li & Warren, 2000; Wang & Simons, 1999).

In a first attempt to disentangle these two alternatives for MOT, G. Liu et al. (2005) asked participants to track a set of moving targets within a virtual three-dimensional box that itself underwent translations, rotations, and zooms. When G. Liu et al. manipulated object speed independently of scene speed, they observed that increasing object speed but not increasing scene speed impaired tracking. This finding suggests that tracking is carried out in allocentric coordinates.

Although the results of research by G. Liu et al. (2005) were straightforward in favor of an allocentric reference system, subsequent studies with different methodologies have revealed more mixed results. In two studies, Howe and his colleagues (Howe, Drew, Pinto, & Horowitz, 2011; Howe, Pinto, Horowitz, 2010) explored the stability of tracking performance across saccades and smooth pursuits. In these studies, disrupting the allocentric frame of reference impaired tracking performance during saccades as well as smooth pursuits. This matches the results found by G. Liu et al. (2005) as well as related research providing evidence that the programming of saccades (Deubel, Bridgeman, & Schneider, 1998) as well as the execution of smooth pursuits (Raymond, Shapiro, & Rose, 1984) also is based on an allocentric representation of the scene. However, in contrast to the results found by G. Liu et al. (2005), tracking performance was also sensitive to manipulations of the retinotopic coherence of the scene during smooth pursuits, indicating that tracking partially is based on retinotopic coordinates (Howe et al., 2010). The observation that allocentric as well as retinocentric coordinate systems contribute to tracking matches with related research from our lab showing that continuous scene information is necessary to maintain objects in an allocentric frame of reference (Huff, Meyerhoff, Papenmeier, & Jahn, 2010; see also Jahn, Wendt, Lotze, Papenmeier, & Huff, 2012, for neuroimaging evidence) whereas abrupt changes in the frame of reference seem to draw onto retinocentric processes (Huff, Jahn, & Schwan, 2009; Jahn, Papenmeier, Meyerhoff, & Huff, 2012).

In a further study, we showed that observers automatically track objects within an allocentric frame of reference as long as continuous motion cues were available (Meyerhoff, Huff, Papenmeier, Jahn, & Schwan, 2011). In this study, our participants tracked objects while either the floor plane, the set of objects, neither, or both underwent continuous rotations of 30 degrees during brief intervals of object invisibility. Thus, we disentangled the spatial orientation of the objects from the spatial orientation of the floor. We observed that participants were unable to suppress continuous visual information about the rotations of the floor plane even under experimental conditions under which this updating process actually disturbed tracking performance (see Fig. 8). Despite the automatic usage

of continuous motion information from the frame of reference, the perceived space also affects tracking performance. For instance, tracking accuracy is impaired when the tracking space is turned upside down (Papenmeier, Meyerhoff, Brockhoff, Jahn, & Huff, *in press*). Further, tracking performance declines when participants have to track objects relative to external points of reference. Such an effect was demonstrated in a study by Thomas and Seiffert (2010). In the critical conditions of this study, the participants' viewpoint on a tracking display (projected via a HMD) changed as a consequence of self-motion. When contrasted to conditions with the same viewpoint change without proprioceptive motion cues, self-motion impaired tracking performance because participants had to track their own location in space in addition to the target objects (see also Thomas & Seiffert, 2011). In fact, it is the self-motion during tracking that draws upon the tracking resource rather than the execution of motor actions themselves (Thornton, Bühlhoff, Horowitz, Rynning, & Lee, 2014; Thornton & Horowitz, 2015; Thomas & Seiffert, 2010; but see also Trick, Guindon, & Vallis, 2006).

With regard to the broader picture of visual attention, the findings from the MOT studies indicate that most attentional processing occurs within an allocentric frame of reference. In other words, the objects are addressed as being at a relative position to other objects or frames of reference (including the own location) rather than in the corresponding retinal coordinates.

Extrapolation

How predictive are visual processes? Does dynamic visual attention anticipate prospective object locations based on their actual trajectories? These questions have been studied extensively with variants of the MOT paradigm. Most of this research has operationalized this questions by testing whether MOT is restricted to pure location tracking (i.e., not predictive) or whether spatiotemporal object information such as motion direction or heading are used to extrapolate object locations and thus predict prospective demands.

The first studies that explored such extrapolatory processes during MOT asked participants to track objects that were rendered invisible for several 100 ms during the trial. Because performance was best when the objects remained stationary during their invisibility, Keane and Pylyshyn (2006) concluded that prospective object locations are not predicted based on previous direction information (see Fig. 9 for a replication of the results). These results were confirmed in a similar study conducted by Fencsik, Klieger, and Horowitz (2007), who showed that direction information might aid tracking but only under specific preview conditions and for a maximum of two targets. With respect to tracking continuously visible objects, however, these results should be interpreted with caution because subsequent work has demonstrated that tracking

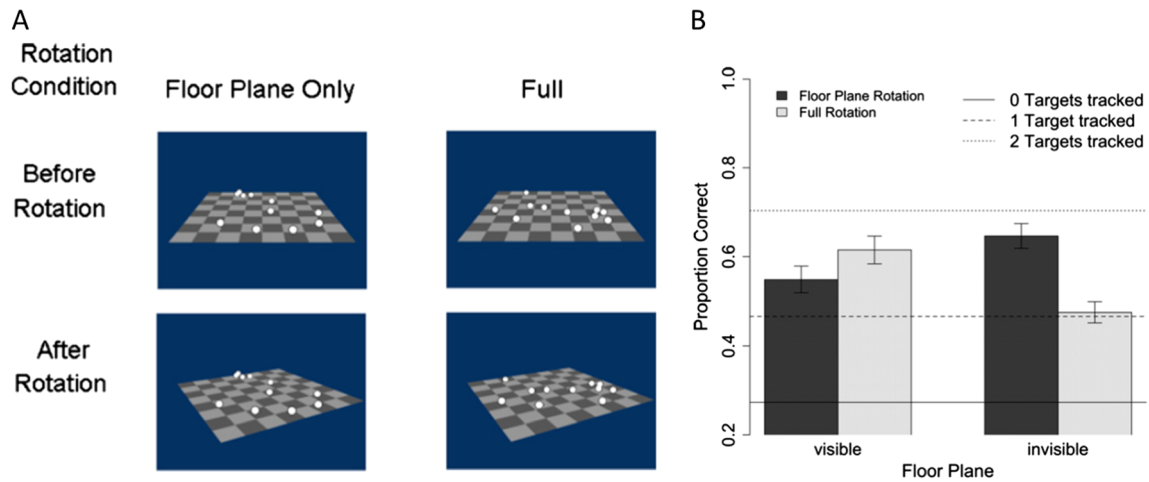


Fig. 8 Experiment conducted by Meyerhoff, Huff, Papenmeier, Jahn, and Schwan (2011). **a** Participants track objects that briefly become invisible during a rotation of the floor plane. After the rotation, the objects reappear at their previous screen location (*floor plane only*) or at the location where they would have been if they rotated with the floor plane (*full*). **b** When the floor plane was invisible during the rotation, tracking accuracy was more accurate when the objects reappeared at the

same screen locations than when they rotated with the floor plane. When the floor plane was visible during the rotation, the results were inverted, namely, tracking was more accurate when the objects rotated with the floor plane than when they reappeared at their previous screen location. This pattern of results shows that participants use the continuous visual information of the floor plane in order to maintain the tracked objects in scene based coordinates. Figures reprinted with permission from Elsevier

invisible objects increases the attentional demands of tracking (Flombaum, Scholl, & Pylyshyn, 2008; see also Ilg, 2008). Also, the mental representation of multiple moving objects has been demonstrated to lag behind the actual object locations. This lag increases with tracking load and object speed (Howard et al., 2011). As a consequence of these restrictions in the interpretation of experiments that require the recovery of invisible objects, later studies have explored variants of the MOT paradigm that involved continuously visible objects.

Holcombe, 2012), even when eye movements were controlled for (Luu & Howe, 2015). This effect depended on tracking load and object speed. Extrapolation occurred less in conditions with higher tracking loads (i.e., more than two target objects) and lower object speed. In line with the idea that motion information is evaluated during tracking, even a single direction change of a target is sufficient to impair tracking performance (Meyerhoff, Papenmeier, Jahn, & Huff, 2013).

A first set of studies that investigated whether observers extrapolate motion paths during tracking continuously visible objects manipulated the predictability of the motion paths of the objects. In general, MOT performance was more accurate with predictable than unpredictable motion paths (Howe &

A second set of studies approached the question of location extrapolation by studying the effect of conflicting motion information on tracking performance. In these studies, the texture of the moving objects signaled motion either in the same or in the opposing direction as the actual direction of the object itself. In the first study of this kind, St. Clair, Huff, and Seiffert

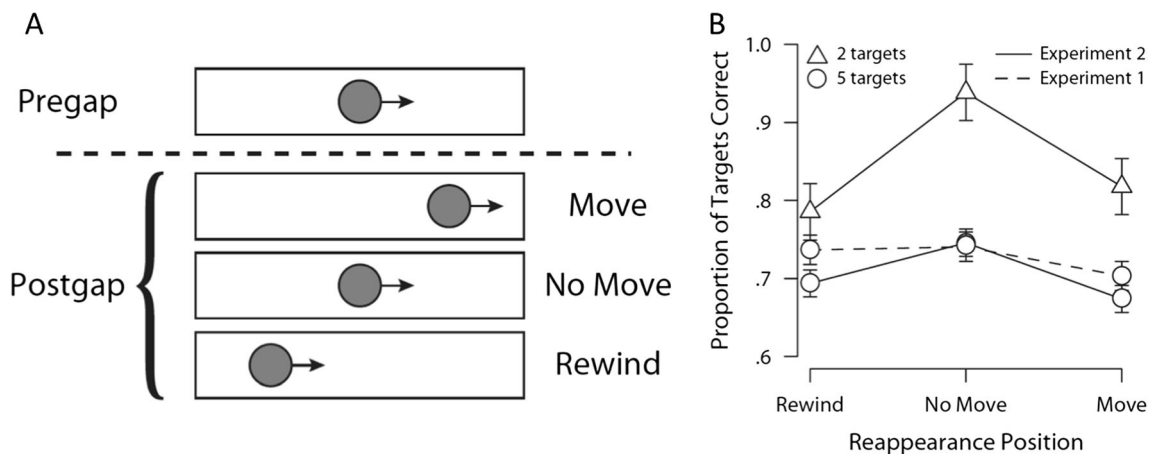


Fig. 9 Illustration of Experiment 1 and 2 conducted by Fencsik, Klieger, and Horowitz (2007) which replicates and extends the finding of Keane and Pylyshyn (2006). **a** During the tracking trials the object briefly disappears. The location of their reappearance is behind, at, or in front of

their last location. **b** Tracking accuracy as a function of tracking load and the location of object reappearance. Participants seem to use only the last object location in order to relocate the tracked targets. Figures reprinted with permission from Springer

(2010) showed that conflicting motion information impaired tracking relative to a neutral baseline. Because there was no benefit in tracking performance when the texture moved along with the object, St. Clair et al. as well as Huff and Papenmeier (2013) have attributed this effect to shifts in the perception of the object location rather than the extrapolation of object locations. This detrimental effect of conflicting motion signals, however, does stem from object-based processing during MOT because tracking impairments arise selectively on objects that actually exhibit conflicting motion signals (Meyerhoff, Papenmeier, & Huff, 2013).

In sum, it is still controversial whether motion information is used during tracking in order to extrapolate object locations. It seems fair to conclude that if extrapolation occurs during tracking, it might occur only to a limited extent. This conclusion also would be in agreement with a recent computational model showing that tracking benefits due to extrapolation would be marginal at best (Zhong et al., 2014). For the broader question regarding predictive attentional processing, this pattern of results of course does not rule out predictive processing in general; however, in the spatiotemporal domain they seem not to arise consistently—at least not with the demands of a MOT task.

Tracking objects with identities

Is the attentional processing of the spatiotemporal information of objects affected by unique object identities? Although many MOT researchers provide motivation for their research with real-world examples such as tracking cars, tracking children on a playground, or tracking players in sports, most research reviewed above was dedicated to the tracking of indistinguishable objects. In real-world scenarios, however, observers track objects with distinct identities. For the purpose of this review, we focus on studies investigating the influence of identities on location tracking performance. This review, however, does not include studies concerned with task-relevant identity information that require maintaining location identity bindings during tracking to give identity-related responses at the end of the trial, sometimes called multiple identity tracking (MIT; Botterill, Allen, & McGeorge, 2011; Cohen et al., 2011; Horowitz et al., 2007; Howard & Holcombe, 2008; Li, Oksama, & Hyönä, 2016; Oksama & Hyönä, 2004, 2008, 2016; Pinto, Howe, Cohen, & Horowitz, 2010; Pylyshyn, 2004; Ren, Chen, Liu, & Fu, 2009).

When observers track unique objects (e.g., colored discs or cartoon animals) instead of indistinguishable objects, location tracking performance increases (Horowitz et al., 2007; Makovski & Jiang, 2009a, 2009b). One reason for this benefit is that adding identities to objects helps in separating targets from distractors (Bae & Flombaum, 2012; Feria, 2012). Therefore, adding identity information to targets does not improve location tracking performance per se but only when

targets and distractors do not share the same features (Horowitz et al., 2007; Howe & Holcombe, 2012; Jardine & Seiffert, 2011; Makovski & Jiang, 2009a, 2009b). The beneficial effect of identity information on location tracking has been attributed to effortful deliberate processing as well as automatic processes. For instance, Makovski and Jiang (2009b) provided evidence for the suggestion that effects of identity information stem from deliberate processing. In their study, they showed that observers effortfully encoded target identities into memory and recovered lost targets based on this memory representation. In their experiments, Makovski and Jiang were able to prevent such a memory-based target recovery by a concurrent color memory task that frequently changed the color of the objects in the display. However, there is also evidence supporting the view that automatic processing of identity information enhances location tracking performance. For instance, when participants were asked to track a set of faces, they were better at tracking attractive faces than unattractive faces, even though face identities were task irrelevant (C. H. Liu & Chen, 2012). Furthermore, identity information influenced participants' tracking performance even when relying on object identities was always harmful to MOT performance (Erlikhman et al., 2013; Papenmeier, Meyerhoff, Jahn, & Huff, 2014).

The occurrence of identity effects on location tracking is also determined by the reliability of spatiotemporal information; that is, effects of identity information arise when spatiotemporal information becomes less reliable. For example, adding identity information to objects influenced tracking performance particularly at reduced interobject spacing (Bae & Flombaum, 2012; Makovski & Jiang, 2009b) or with an increasing number of distractors in the display (Drew et al., 2013), although identity effects can also occur across large distances with low object speeds (Störmer, Li, Heekeren, & Lindenberger, 2011). In one of our studies (Papenmeier et al., 2014), we manipulated spatiotemporal reliability. Whereas swapping object colors between targets and distractors left tracking performance unaffected by continuous spatiotemporal information, we observed identity effects when we introduced spatiotemporal discontinuities such as abrupt scene rotations, abrupt zooms, or reduced presentation frame rates. These findings indicate that tracking itself does not only rely on location information but also identity information of objects. The exact mechanisms by which spatiotemporal and identity information work together still needs to be resolved. Identity information could either be encoded during tracking and used to establish object correspondence (Papenmeier et al., 2014; see also Jardine & Seiffert, 2011) or identity information might just be another source of input that is directly utilized by the tracking mechanism (Oksama & Hyönä, 2016), such as the visual indices (FINSTs) that stick to feature clusters on the retina without recognizing their identities (Pylyshyn, 1989).

Development and expertise

How does development and experience shape attentional processing? The ability to track multiple objects in parallel arises relatively early in normally developing children. Even children as young as 6 months are able to keep track of objects moving synchronously on circular trajectories (Richardson & Kirkham, 2004). At the age of 6.5 years, children are able to track up to four moving objects (O’Hearn, Hoffman, & Landau, 2010; O’Hearn, Landau, & Hoffman, 2005; see also Brockhoff et al., 2016). Remarkably, even children suffering from autism spectrum disorders show a qualitatively similar pattern of development of object tracking as healthy children, although they reach the same quantitative level of tracking at a later biological age (Koldewyn, Weight, Kanwisher, & Jiang, 2013; see also Griffith, Pennington, Wehner, & Rogers, 1999; Poirier, Martin, Gaigg, & Bowler, 2011). To date, science is only beginning to understand how visual processing is affected by disorders. As an example, the development of MOT in children suffering from Williams syndrome (a genetic disorder involving impairments of visuospatial processing; O’Hearn et al., 2010) seem to differ qualitatively from their peers.

At older ages (e.g., above 60 years), MOT performance declines relatively to younger adults (20–35 years; Sekuler, McLaughlin, Yotsumoto, 2008; Störmer et al., 2011; Trick, Perl, & Sethi, 2005). However, even at older ages, performance remained well above one target (Trick et al., 2005). Some studies also have suggested that certain types of expertise come along with an increased ability to track multiple objects. For instance, radar operators (Allen, McGeorge, Pearson, & Milne, 2004) as well as regular video game players (Dye & Bavelier, 2010; Green & Bavelier, 2006; Sekuler et al., 2008) outperform their corresponding control groups. However, such effects of expertise seem to be restricted to closely matching tasks that involve object tracking. Other plausible activities such as team sports were uncorrelated with the capability to track multiple objects even after more than 10 years of extensive practice (Memmert, Simons, & Grimme, 2009).

Future directions

Whereas MOT has been studied mostly in laboratory settings, the conclusions of these studies often encompass everyday life activities, such as driving cars or supervising children on a playground. Therefore, a central challenge for future research on MOT is to determine and to strengthen the ecological validity of MOT by considering relevant factors, such as the environmental complexity and its constraints. This could be accomplished by embedding the tracking task into more naturalistic scenarios (e.g., Ericson, Parr, Beck, Wolshon, 2017; Lochner & Trick, 2014) or by investigating the impact of the

ability to track indistinguishable objects on performance in naturalistic tasks such as team sports (Memmert et al., 2009). Research on the ecological validity of MOT has just begun and much more work is needed to provide a complete and compelling demonstration of the relevance of MOT for everyday life activities.

On a theoretical level, much progress has been made over recent years. Nevertheless, there are still several open questions that cannot be answered sufficiently based on the existing work yet. Related to the question of the ecological validity, it is still an open question to what extent MOT reflects a singular process or whether it consists of several sub-routines (including attentional selection and working memory processes) that interact with each other based on current task demands (e.g., Drew et al., 2012; Oksama & Hyöna, 2016). Further, more work is necessary to distinguish between models based on fixed architectural constraints or models based on the idea of a flexible attentional resource. Although the pendulum currently seems to tend toward the more flexible attentional resource models, until today no decisive evidence in favor of any one of these theories has been obtained. One potential way to deepen the understanding of the mechanism of tracking and to further develop more sophisticated models of tracking would be a more systematic investigation of concrete instances of tracking errors. An interesting approach to work toward this goal comes from computational modeling that can be helpful to specify and quantify the cognitive operations determining both successful tracking and tracking errors during MOT.

References

- Allen, R., McGeorge, P., Pearson, D. G., & Milne, A. B. (2004). Attention and expertise in multiple target tracking. *Applied Cognitive Psychology, 18*(3), 337–347. doi:10.1002/acp.975
- Allen, R., McGeorge, P., Pearson, D. G., & Milne, A. B. (2006). Multiple-target tracking: A role for working memory? *The Quarterly Journal of Experimental Psychology, 59*, 1101–1116. doi:10.1080/02724980543000097
- Alnaes, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision, 14*(4), 1–20. doi:10.1167/14.4.1
- Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science, 16*, 637–643. doi:10.1111/j.1467-9280.2005.01587.x
- 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), 1–10. doi:10.1167/7.13.14
- Alvarez, G. A., Horowitz, T. S., Arsenio, H. C., DiMase, J. S., & Wolfe, J. M. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology: Human Perception and Performance, 31*, 643–667. doi:10.1037/0096-1523.31.4.643

- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*, 392–398. doi:10.1111/j.1467-9280.2008.02098.x
- Alvarez, G. A., & Scholl, B. J. (2005). How does attention select and track spatially extended objects? New effects of attentional concentration and amplification. *Journal of Experimental Psychology: General*, *134*, 461–476. doi:10.1037/0096-3445.134.4.461
- Awh, E., Anillo-Vento, L., & Hillyard, S. A. (2000). The role of spatial selective attention in working memory for locations: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *12*, 840–847. doi:10.1162/089892900562444
- Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*, 119–126. doi:10.1016/S1364-6613(00)01593-X
- Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 834–846. doi:10.1037/0096-1523.26.2.834
- Bae, G. Y., & Flombaum, J. I. (2012). Close encounters of the distracting kind: Identifying the cause of visual tracking errors. *Attention, Perception & Psychophysics*, *74*, 703–715. doi:10.3758/s13414-011-0260-1
- Battelli, L., Cavanagh, P., Intriligator, J., Tramo, M. J., Hénaff, M. A., Michèl, F., & Barton, J. J. (2001). Unilateral right parietal damage leads to bilateral deficit for high-level motion. *Neuron*, *32*, 985–995. doi:10.1016/S0896-6273(01)00536-0
- Bettencourt, K. C., & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision*, *9*(7), 1–11. doi:10.1167/9.7.9
- Botterill, K., Allen, R., & McGeorge, P. (2011). Multiple-object tracking: The binding of spatial location and featural identity. *Experimental Psychology*, *58*, 196–200. doi:10.1027/1618-3169/a000085
- Brockhoff, A., Papenmeier, F., Wolf, K., Pfeiffer, T., Jahn, G., & Huff, M. (2016). Viewpoint matters: Exploring the involvement of reference frames in multiple object tracking from a developmental perspective. *Cognitive Development*, *37*, 1–8. doi:10.1016/j.cogdev.2015.10.004
- Carter, O. L., Burr, D. C., Pettigrew, J. D., Wallis, G. M., Hasler, F., & Vollenweider, F. X. (2005). Using psilocybin to investigate the relationship between attention, working memory, and the serotonin 1A and 2A receptors. *Journal of Cognitive Neuroscience*, *17*, 1497–1508. doi:10.1162/089892905774597191
- Castiello, U., & Umiltà, C. (1992). Splitting focal attention. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 837–848. doi:10.1037/0096-1523.18.3.837
- Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*, 349–354. doi:10.1016/j.tics.2005.05.009
- 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*, 710–725. doi:10.3758/s13414-013-0425-1
- Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological Science*. doi:10.1177/0956797611419168
- Cohen, M. A., Pinto, Y., Howe, P. D., & Horowitz, T. S. (2011). The what–where trade-off in multiple-identity tracking. *Attention, Perception & Psychophysics*, *73*, 1422–1434. doi:10.3758/s13414-011-0089-7
- Colas, F., Flacher, F., Tanner, T., Bessiere, P., & Girard, B. (2009). Bayesian models of eye movement selection with retinotopic maps. *Biological Cybernetics*, *100*, 203–214. doi:10.1007/s00422-009-0292-y
- Coull, J. T., & Frith, C. D. (1998). Differential activation of right superior parietal cortex and intraparietal sulcus by spatial and nonspatial attention. *NeuroImage*, *8*, 176–187. doi:10.1006/nimg.1998.0354
- Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, *80*, 2657–2670.
- Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001). Attention response functions: Characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron*, *32*, 737–745. doi:10.1016/S0896-6273(01)00499-8
- d’Avossa, G., Shulman, G. L., Snyder, A. Z., & Corbetta, M. (2006). Attentional selection of moving objects by a serial process. *Vision Research*, *46*, 3403–3412. doi:10.1016/j.visres.2006.04.018
- Delvenne, J. F. (2005). The capacity of visual short-term memory within and between hemifields. *Cognition*, *96*, B79–B88. doi:10.1016/j.cognition.2004.12.007
- Deubel, H., Bridgeman, B., & Schneider, W. X. (1998). Immediate post-saccadic information mediates space constancy. *Vision Research*, *38*, 3147–3159. doi:10.1016/S0042-6989(98)00048-0
- DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., & Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 2382–2386. doi:10.1073/pnas.93.6.2382
- Doran, M. M., & Hoffman, J. E. (2010). The role of visual attention in multiple object tracking: Evidence from ERPs. *Attention, Perception & Psychophysics*, *72*, 33–52. doi:10.3758/APP.72.1.33
- Doran, M. M., Hoffman, J. E., & Scholl, B. (2009). The role of eye fixations in concentration and amplification effects during multiple object tracking. *Visual Cognition*, *17*, 574–597. doi:10.1080/13506280802117010
- Drew, T., Horowitz, T. S., Wolfe, J. M., & Vogel, E. K. (2011). Delineating the neural signatures of tracking spatial position and working memory during attentive tracking. *The Journal of Neuroscience*, *31*, 659–668. doi:10.1523/JNEUROSCI.1339-10.2011
- Drew, T., Horowitz, T. S., & Vogel, E. K. (2013). Swapping or dropping? Electrophysiological measures of difficulty during multiple object tracking. *Cognition*, *126*, 213–223. doi:10.1016/j.cognition.2012.10.003
- Drew, T., McCollough, A. W., Horowitz, T. S., & Vogel, E. K. (2009). Attentional enhancement during multiple-object tracking. *Psychonomic Bulletin & Review*, *16*, 411–417. doi:10.3758/PBR.16.2.411
- Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *The Journal of Neuroscience*, *28*, 4183–4191. doi:10.1523/JNEUROSCI.0556-08.2008
- Dye, M. W., & Bavelier, D. (2010). Differential development of visual attention skills in school-age children. *Vision Research*, *50*, 452–459. doi:10.1016/j.visres.2009.10.010
- Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*, 161. doi:10.1037/0096-3445.123.2.161
- Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, *7*, 181–192. doi:10.1093/cercor/7.2.181
- Ericson, J. M., & Christensen, J. C. (2012). Reallocating attention during multiple object tracking. *Attention, Perception & Psychophysics*, *74*, 831–840. doi:10.3758/s13414-012-0294-z
- Ericson, J. M., Parr, S. A., Beck, M. R., & Wolshon, B. (2017). Compensating for failed attention while driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *45*, 65–74. doi:10.1016/j.trf.2016.11.015
- Erlikhman, G., Keane, B. P., Mettler, E., Horowitz, T. S., & Kellman, P. J. (2013). Automatic feature-based grouping during multiple object

- tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 1625–1637. doi:10.1037/a0031750
- Evers, K., de-Wit, L., Van der Hallen, R., Haesen, B., Steyaert, J., Noens, I., & Wagemans, J. (2014). Reduced grouping interference in children with ASD: Evidence from a multiple object tracking task. *Journal of Autism and Developmental Disorders*, 44, 1779–1787. doi:10.1007/s10803-013-2031-4
- Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition*, 108, 201–209. doi:10.1016/j.cognition.2007.11.008
- Fehd, H. M., & Seiffert, A. E. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision*, 10, 1–13. doi:10.1167/10.4.19
- Fencsik, D. E., Klieger, S. B., & Horowitz, T. S. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception & Psychophysics*, 69, 567–77. doi:10.3758/BF03193914
- Feria, C. S. (2008). The distribution of attention within objects in multiple-object scenes: Prioritization by spatial probabilities and a center bias. *Perception & Psychophysics*, 70, 1185–1196. doi:10.3758/PP.70.7.1185
- Feria, C. S. (2010). Attentional prioritizations based on spatial probabilities can be maintained on multiple moving objects. *Attention, Perception & Psychophysics*, 72, 926–938. doi:10.3758/APP.72.4.926
- Feria, C. S. (2012). The effects of distractors in multiple object tracking are modulated by the similarity of distractor and target features. *Perception*, 41, 287–304. doi:10.1068/p7053
- 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 & Psychophysics*, 75, 53–67. doi:10.3758/s13414-012-0369-x
- Flombaum, J. I., Scholl, B. J., & Pylyshyn, Z. W. (2008). Attentional resources in visual tracking through occlusion: The high-beams effect. *Cognition*, 107, 904–931. doi:10.1016/j.cognition.2007.12.015
- Fougnie, D., & Marois, R. (2006). Distinct capacity limits for attention and working memory evidence from attentive tracking and visual working memory paradigms. *Psychological Science*, 17, 526–534. doi:10.1111/j.1467-9280.2006.01739.x
- Fougnie, D., & Marois, R. (2009). Attentive tracking disrupts feature binding in visual working memory. *Visual Cognition*, 17, 48–66. doi:10.1080/13506280802281337
- Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected at once? *Journal of Experimental Psychology: Human Perception and Performance*, 33, 1003–1012. doi:10.1037/0096-1523.33.5.1003
- 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. doi:10.1177/0956797610373935
- 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. doi:10.3758/PBR.15.4.802
- Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, 101, 217–245. doi:10.1016/j.cognition.2005.10.004
- Griffith, E. M., Pennington, B. F., Wehner, E. A., & Rogers, S. J. (1999). Executive functions in young children with autism. *Child Development*, 70, 817–832. doi:10.1111/1467-8624.00059
- Holcombe, A. O., & Chen, W. Y. (2012). Exhausting attentional tracking resources with a single fast-moving object. *Cognition*, 123, 218–228. doi:10.1016/j.cognition.2011.10.003
- Holcombe, A. O., & Chen, W. Y. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to <3 Hz when tracking three. *Journal of Vision*, 13, 12–12. doi:10.1167/13.1.12
- Holcombe, A. O., Chen, W. Y., & Howe, P. D. (2014). Object tracking: Absence of long-range spatial interference supports resource theories. *Journal of Vision*, 14(6), 1–24. doi:10.1167/14.6.1
- Hopf, J.-M., Boehler, C. N., Luck, S. J., Tsotsos, J. K., Heinze, H.-J., & Schoenfeld, M. A. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proceedings of the National Academy of Sciences*, 103, 1053–1058. doi:10.1073/pnas.0507746103
- Horowitz, T. S., Birnkrant, R. S., Fencsik, D. E., Tran, L., & Wolfe, J. M. (2006). How do we track invisible objects? *Psychonomic Bulletin & Review*, 13, 516–523. doi:10.3758/BF03193879
- Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception & Psychophysics*, 72, 1765–1775. doi:10.3758/APP.72.7.1765
- Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics*, 69, 172–184. doi:10.3758/BF03193740
- Horowitz, T. S., & Kuzmova, Y. (2011). Can we track holes? *Vision Research*, 51, 1013–1021. doi:10.1016/j.visres.2011.02.009
- 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, 1164–1180. doi:10.1016/j.visres.2008.01.023
- Howard, C. J., Masom, D., & Holcombe, A. O. (2011). Position representations lag behind targets in multiple object tracking. *Vision Research*, 51, 1907–19. doi:10.1016/j.visres.2011.07.001
- Howe, P. D., Cohen, M. A., Pinto, Y., & Horowitz, T. S. (2010). Distinguishing between parallel and serial accounts of multiple object tracking. *Journal of Vision*, 10(8), 1–13. doi:10.1167/10.8.11
- Howe, P. D., Drew, T., Pinto, Y., & Horowitz, T. S. (2011). Remapping attention in multiple object tracking. *Vision Research*, 51, 489–495. doi:10.1016/j.visres.2011.01.001
- 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), 1–10. doi:10.1167/12.13.10
- Howe, P. D., Holcombe, A. O., Lapierre, M. D., & Cropper, S. J. (2013). Visually tracking and localizing expanding and contracting objects. *Perception*, 42, 1281–1300. doi:10.1068/p7635
- Howe, P. D., Horowitz, T. S., Morocz, I. A., Wolfe, J., & Livingstone, M. S. (2009). Using fMRI to distinguish components of the multiple object tracking task. *Journal of Vision*, 9(4), 1–11. doi:10.1167/9.4.10
- Howe, P. D., Incedon, N. C., & Little, D. R. (2012). Can attention be confined to just part of a moving object? Revisiting target-distractor merging in multiple object tracking. *PLoS ONE*, 7, e41491. doi:10.1371/journal.pone.0041491
- Howe, P. D., Pinto, Y., & Horowitz, T. S. (2010). The coordinate systems used in visual tracking. *Vision Research*, 50, 2375–2380. doi:10.1016/j.visres.2010.09.026
- Huang, L., Mo, L., & Li, Y. (2012). Measuring the interrelations among multiple paradigms of visual attention: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, 38, 414–428. doi:10.1037/a0026314
- Hudson, C., Howe, P. D., & Little, D. R. (2012). Hemifield effects in multiple identity tracking. *PLoS ONE*, 7, e43796. doi:10.1371/journal.pone.0043796
- Huff, M., Jahn, G., & Schwan, S. (2009). Tracking multiple objects across abrupt viewpoint changes. *Visual Cognition*, 17, 297–306. doi:10.1080/13506280802061838
- Huff, M., Meyerhoff, H. S., Papenmeier, F., & Jahn, G. (2010). Spatial updating of dynamic scenes: Tracking multiple invisible objects across viewpoint changes. *Attention, Perception & Psychophysics*, 72, 628–636. doi:10.3758/APP.72.3.628
- Huff, M., & Papenmeier, F. (2013). It is time to integrate: The temporal dynamics of object motion and texture motion integration in

- multiple object tracking. *Vision Research*, 76, 25–30. doi:10.1016/j.visres.2012.10.001
- Huff, M., Papenmeier, F., Jahn, G., & Hesse, F. (2010). Eye movements across viewpoint changes in multiple object tracking. *Visual Cognition*, 18, 1368–1391. doi:10.1080/13506285.2010.495878
- Huff, M., Papenmeier, F., & Zacks, J. M. (2012). Visual target detection is impaired at event boundaries. *Visual Cognition*, 20, 848–864. doi:10.1080/13506285.2012.705359
- Hulleman, J. (2005). The mathematics of multiple object tracking: From proportions correct to number of objects tracked. *Vision Research*, 45, 2298–2309. doi:10.1016/j.visres.2005.02.016
- Ilg, U. J. (2008). The role of areas MT and MST in coding of visual motion underlying the execution of smooth pursuit. *Vision Research*, 48, 2062–2069. doi:10.1016/j.visres.2008.04.015
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, 43, 171–216. doi:10.1006/cogp.2001.0755
- 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. doi:10.1167/9.4.1
- Jahn, G., Papenmeier, F., Meyerhoff, H. S., & Huff, M. (2012). Spatial reference in multiple object tracking. *Experimental Psychology*, 59, 163–173. doi:10.1027/1618-3169/a000139
- Jahn, G., Wendt, J., Lotze, M., Papenmeier, F., & Huff, M. (2012). Brain activation during spatial updating and attentive tracking of moving targets. *Brain and Cognition*, 78, 105–113. doi:10.1016/j.bandc.2011.12.001
- Jardine, N. L., & Seiffert, A. E. (2011). Tracking objects that move where they are headed. *Attention, Perception & Psychophysics*, 73, 2168–2179. doi:10.3758/s13414-011-0169-8
- Jovicich, J., Peters, R. J., Koch, C., Braun, J., Chang, L., & Ernst, T. (2001). Brain areas specific for attentional load in a motion-tracking task. *Journal of Cognitive Neuroscience*, 13, 1048–1058. doi:10.1162/089892901753294347
- Keane, B. P., Mettler, E., Tsoi, V., & Kellman, P. J. (2011). Attentional signatures of perception: Multiple object tracking reveals the automaticity of contour interpolation. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 685. doi:10.1037/a0020674
- Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cognitive Psychology*, 52, 346–368. doi:10.1016/j.cogpsych.2005.12.001
- Koldewyn, K., Weigelt, S., Kanwisher, N., & Jiang, Y. (2013). Multiple object tracking in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 43, 1394–1405. doi:10.1007/s10803-012-1694-6
- Kramer, A. F., & Hahn, S. (1995). Splitting the beam: Distribution of attention over noncontiguous regions of the visual field. *Psychological Science*, 381–386. doi:10.1111/j.1467-9280.1995.tb00530.x
- Kunar, M. A., Carter, R., Cohen, M., & Horowitz, T. S. (2008). Telephone conversation impairs sustained visual attention via a central bottleneck. *Psychonomic Bulletin & Review*, 15, 1135–1140. doi:10.3758/PBR.15.6.1135
- Lennie, P. (1998). Single units and visual cortical organization. *Perception*, 27, 889–935. doi:10.1068/p270889
- Li, J., Oksama, L., & Hyönä, J. (2016). How facial attractiveness affects sustained attention. *Scandinavian Journal of Psychology*, 57(5), 383–392. doi:10.1111/sjop.12304
- Li, L., & Warren, W. H. (2000). Perception of heading during rotation: Sufficiency of dense motion parallax and reference objects. *Vision Research*, 40, 3873–3894. doi:10.1016/S0042-6989(00)00196-6
- 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. doi:10.1037/0096-1523.31.2.235
- Liu, C. H., & Chen, W. (2012). Beauty is better pursued: Effects of attractiveness in multiple-face tracking. *The Quarterly Journal of Experimental Psychology*, 65, 553–564. doi:10.1080/17470218.2011.624186
- Liverence, B. M., & Scholl, B. J. (2011). Selective attention warps spatial representation parallel but opposing effects on attended versus inhibited objects. *Psychological Science*, 22, 1600–1608. doi:10.1177/0956797611422543
- Lochner, M. J., & Trick, L. M. (2014). Multiple-object tracking while driving: The multiple-vehicle tracking task. *Attention, Perception, & Psychophysics*. doi:10.3758/s13414-014-0694-3
- Luck, S. J. (1995). Multiple mechanisms of visual-spatial attention: Recent evidence from human electrophysiology. *Behavioural Brain Research*, 71, 113–123. doi:10.1016/0166-4328(95)00041-0
- Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: Psychophysical and electrophysiological evidence for early selection. *Journal of Experimental Psychology: Human Perception & Performance*, 20, 887–904. doi:10.1037/0096-1523.20.4.887
- Lukavský, J. (2013). Eye movements in repeated multiple object tracking. *Journal of Vision*, 13(7), 1–16. doi:10.1167/13.7.9
- Lukavský, J., & Děchtěrenko, F. (2016). Gaze position lagging behind scene content in multiple object tracking: Evidence from forward and backward presentations. *Attention, Perception & Psychophysics*, 78(8), 2456–2468. doi:10.3758/s13414-016-1178-4
- Luu, T., & Howe, P. D. (2015). Extrapolation occurs in multiple object tracking when eye movements are controlled. *Attention, Perception & Psychophysics*, 77, 1919–1929. doi:10.3758/s13414-015-0891-8
- Ma, Z., & Flombaum, J. I. (2013). Off to a bad start: Uncertainty about the number of targets at the onset of multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 1421–1432. doi:10.1037/a0031353
- Makovski, T., & Jiang, Y. V. (2009a). Feature binding in attentive tracking of distinct objects. *Visual Cognition*, 17, 180–194. doi:10.1080/13506280802211334
- Makovski, T., & Jiang, Y. V. (2009b). The role of visual working memory in attentive tracking of unique objects. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1687–1697. doi:10.1037/a0016453
- McMains, S. A., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron*, 42, 677–686. doi:10.1016/S0896-6273(04)00263-6
- Memmert, D., Simons, D., & Grimme, T. (2009). The relationship between visual attention and expertise in sports. *Psychology of Sport and Exercise*, 10, 146–151. doi:10.1016/j.psychsport.2008.06.002
- Meyerhoff, H. S., Huff, M., Papenmeier, F., Jahn, G., & Schwan, S. (2011). Continuous visual cues trigger automatic spatial target updating in dynamic scenes. *Cognition*, 121, 73–82. doi:10.1016/j.cognition.2011.06.001
- Meyerhoff, H. S., Papenmeier, F., Jahn, G., & Huff, M. (2013). A single unexpected change in target-but not distractor motion impairs multiple object tracking. *i-Perception*, 4, 81–83. doi:10.1068/i0567sas
- Meyerhoff, H. S., Papenmeier, F., Jahn, G., & Huff, M. (2015). Distractor locations influence multiple object tracking beyond interobject spacing: Evidence from equidistant distractor displacements. *Experimental Psychology*, 62, 170–180. doi:10.1027/1618-3169/a000283
- 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, 776–787. doi:10.1037/xhp0000187

- Meyerhoff, H. S., Papenmeier, F., & Huff, M. (2013). Object-based integration of motion information during attentive tracking. *Perception*, 42, 119–121. doi:10.1068/p7273
- Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature*, 424, 309–312. doi:10.1038/nature01812
- Müller, N. G., Mollenhauer, M., Rösler, A., & Kleinschmidt, A. (2005). The attentional field has a Mexican hat distribution. *Vision Research*, 45, 1129–1137. doi:10.1016/j.visres.2004.11.003
- O’Hearn, K., Hoffman, J. E., & Landau, B. (2010). Developmental profiles for multiple object tracking and spatial memory: Typically developing preschoolers and people with Williams syndrome. *Developmental Science*, 13, 430–440. doi:10.1111/j.1467-7687.2009.00893.x
- O’Hearn, K., Landau, B., & Hoffman, J. E. (2005). Multiple object tracking in people with Williams syndrome and in normally developing children. *Psychological Science*, 16, 905–912. doi:10.1111/j.1467-9280.2005.01635.x
- Ogawa, H., Watanabe, K., & Yagi, A. (2009). Contextual cueing in multiple object tracking. *Visual Cognition*, 17, 1244–1258. doi:10.1080/13506280802457176
- Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higherorder cognition? An individual difference approach. *Visual Cognition*, 11, 631–671. doi:10.1080/13506280344000473
- Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology*, 56, 237–283. doi:10.1016/j.cogpsych.2007.03.001
- Oksama, L., & Hyönä, J. (2016). Position tracking and identity tracking are separate systems: Evidence from eye movements. *Cognition*, 146, 393–409. doi:10.1016/j.cognition.2015.10.016
- Papenmeier, F., Meyerhoff, H.S., Brockhoff, A., Jahn, G., & Huff, M. (in press). Upside-down: Perceived space affects object-based attention. *Journal of Experimental Psychology: Human Perception and Performance*. doi:10.1037/xhp0000421
- Papenmeier, F., Meyerhoff, H. S., Jahn, G., & Huff, M. (2014). Tracking by location and features: Object correspondence across spatiotemporal discontinuities during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 159–171. doi:10.1037/a0033117
- Pinto, Y., Howe, P. D., Cohen, M. A., & Horowitz, T. S. (2010). The more often you see an object, the easier it becomes to track it. *Journal of Vision*, 10, 1–15. doi:10.1167/10.10.4
- Poirier, M., Martin, J. S., Gaigg, S. B., & Bowler, D. M. (2011). Short-term memory in autism spectrum disorder. *Journal of Abnormal Psychology*, 120, 247–252. doi:10.1037/a0022298
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25. doi:10.1080/0033558008248231
- Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109, 160. doi:10.1037/0096-3445.109.2.160
- Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32, 65–97. doi:10.1016/0010-0277(89)90014-0
- Pylyshyn, Z. W. (1994). Some primitive mechanisms of spatial attention. *Cognition*, 50, 363–384. doi:10.1016/0010-0277(94)90036-1
- Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80, 127–158. doi:10.1016/S0010-0277(00)00156-6
- Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, 11, 801–822. doi:10.1080/13506280344000518
- Pylyshyn, Z. W. (2006). Some puzzling findings in multiple object tracking (MOT): II. Inhibition of moving nontargets. *Visual Cognition*, 14, 175–198. doi:10.1080/13506280544000200
- Pylyshyn, Z. W. (2007). *Things and places: How the mind connects with the world*. Cambridge: MIT Press.
- Pylyshyn, Z. W., & Annan, V. (2006). Dynamics of target selection in multiple object tracking (MOT). *Spatial Vision*, 19, 485–504. doi:10.1163/156856806779194017
- Pylyshyn, Z. W., Burkell, J., Fisher, B., Sears, C., Schmidt, W., & Trick, L. (1994). Multiple parallel access in visual attention. *Canadian Journal of Experimental Psychology*, 48, 260–283. doi:10.1037/1196-1961.48.2.260
- Pylyshyn, Z. W., Haladjian, H. H., King, C. E., & Reilly, J. E. (2008). Selective nontarget inhibition in multiple object tracking. *Visual Cognition*, 16, 1011–1021. doi:10.1080/13506280802247486
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3, 179–197. doi:10.1163/156856888X00122
- Raymond, J. E., Shapiro, K. L., & Rose, D. J. (1984). Optokinetic backgrounds affect perceived velocity during ocular tracking. *Perception & Psychophysics*, 36, 221–224. doi:10.3758/BF03206362
- Rehman, A. U., Kihara, K., Matsumoto, A., & Ohtsuka, S. (2015). Attentive tracking of moving objects in real 3D space. *Vision Research*, 109, 1–10. doi:10.1016/j.visres.2015.02.004
- Ren, D., Chen, W., Liu, C. H., & Fu, X. (2009). Identity processing in multiple-face tracking. *Journal of Vision*, 9(5), 1–15. doi:10.1167/9.5.18
- Richardson, D. C., & Kirkham, N. Z. (2004). Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General*, 133, 46–62. doi:10.1037/0096-3445.133.1.46
- Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, 80, 1–46. doi:10.1016/S0010-0277(00)00152-9
- Scholl, B. J. (2009). What have we learned about attention from multiple object tracking (and vice versa)? In D. Dedrick & L. Trick (Eds.), *Computation, cognition, and Pylyshyn* (pp. 49–78). Cambridge: MIT Press.
- Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, 38, 259–290. doi:10.1006/cogp.1998.0698
- Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, 80, 159–177. doi:10.1016/S0010-0277(00)00157-8
- Scimeca, J. M., & Franconeri, S. L. (2015). Selecting and tracking multiple objects. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6, 109–118. doi:10.1002/wcs.1328
- Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology*, 54, 1–14. doi:10.1037/h0087326
- Sekuler, R., McLaughlin, C., & Yotsumoto, Y. (2008). Age-related changes in attentional tracking of multiple moving objects. *Perception*, 37, 867–876. doi:10.1068/p5923
- Shim, W. M., Alvarez, G. A., & Jiang, Y. V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychonomic Bulletin & Review*, 15, 390–397. doi:10.3758/PBR.15.2.390
- Smyth, M. M., & Scholey, K. A. (1994). Interference in immediate spatial memory. *Memory & Cognition*, 22, 1–13. doi:10.3758/BF03202756
- St. Clair, R., Huff, M., & Seiffert, A. E. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision*, 10, 1–13. doi:10.1167/10.4.18
- Sternsheim, H., Agam, Y., & Sekuler, R. (2011). EEG correlates of attentional load during multiple object tracking. *PLoS ONE*, 6(7), e22660.
- Störmer, V. S., Alvarez, G. A., & Cavanagh, P. (2014). Within-hemifield competition in early visual areas limits the ability to track multiple

- objects with attention. *The Journal of Neuroscience*, *34*, 11526–11533. doi:10.1523/JNEUROSCI.0980-14.2014
- Störmer, V. S., Li, S.-C., Heekeren, H. R., & Lindenberger, U. (2011). Feature-based interference from unattended visual field during attentional tracking in younger and older adults. *Journal of Vision*, *11*, 1–12. doi:10.1167/11.2.1
- Störmer, V. S., Winther, G. N., Li, S. C., & Andersen, S. K. (2013). Sustained multifocal attentional enhancement of stimulus processing in early visual areas predicts tracking performance. *The Journal of Neuroscience*, *33*, 5346–5351. doi:10.1523/JNEUROSCI.4015-12.2013
- Thomas, L. E., & Seiffert, A. E. (2010). Self-motion impairs multiple-object tracking. *Cognition*, *117*, 80–86. doi:10.1016/j.cognition.2010.07.002
- Thomas, L. E., & Seiffert, A. E. (2011). How many objects are you worth? Quantification of the self-motion load on multiple object tracking. *Frontiers in Psychology*, *2*(245), 1–5. doi:10.3389/fpsyg.2011.00245
- Thornton, I. M., Bühlhoff, H. H., Horowitz, T. S., Rynning, A., & Lee, S. W. (2014). Interactive multiple object tracking (iMOT). *PLoS ONE*, *9*, e86974. doi:10.1371/journal.pone.0086974
- Thornton, I. M., & Horowitz, T. S. (2015). Does action disrupt multiple object tracking (MOT)? *Psichologija*, *48*, 289–301. doi:10.2298/PSI1503289T
- Tombu, M., & Seiffert, A. E. (2008). Attentional costs in multiple-object tracking. *Cognition*, *108*, 1–25. doi:10.1016/j.cognition.2007.12.014
- 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. doi:10.3758/s13414-010-0060-z
- Trick, L. M., Guindon, J., & Vallis, L. A. (2006). Sequential tapping interferes selectively with multiple-object tracking: Do finger-tapping and tracking share a common resource? *The Quarterly Journal of Experimental Psychology*, *59*, 1188–1195. doi:10.1080/17470210600673990
- Trick, L. M., Perl, T., & Sethi, N. (2005). Age-related differences in multiple-object tracking. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*, 102–P105. doi:10.1093/geronb/60.2.P102
- Van der Hallen, R., Evers, K., de-Wit, L., Steyaert, J., Noens, I., & Wagemans, J. (2015). Multiple object tracking reveals object-based grouping interference in children with ASD. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-015-2463-0
- Van Essen, D. C., Lewis, J. W., Drury, H. A., Hadjikhani, N., Tootell, R. B., Bakircioglu, M., & Miller, M. I. (2001). Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Research*, *41*, 1359–1378. doi:10.1016/S0042-6989(01)00045-1
- van Marle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science*, *14*, 498–504. doi:10.1111/1467-9280.03451
- Vishwanath, D., & Kowler, E. (2003). Localization of shapes: Eye movements and perception compared. *Vision Research*, *43*, 1637–1653. doi:10.1016/S0042-6989(03)00168-8
- Viswanathan, L., & Mingolla, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking. *Perception*, *31*, 1415–1437. doi:10.1068/p3432
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751. doi:10.1038/nature02447
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500–503. doi:10.1038/nature04171
- Vul, E., Frank, M., Alvarez, G. A., & Tenenbaum, J. B. (2009, February 26–March 3). *Statistical decision theory and the allocation of cognitive resources in multiple object tracking*. Presentation at the Computational and Systems Neuroscience Meeting, Salt Lake City, UT.
- Wang, R. F., & Simons, D. J. (1999). Active and passive scene recognition across views. *Cognition*, *70*, 191–210. doi:10.1016/S0010-0277(99)00012-8
- Wolfe, J. M., Place, S. S., & Horowitz, T. S. (2007). Multiple object juggling: Changing what is tracked during extended multiple object tracking. *Psychonomic Bulletin & Review*, *14*, 344–349. doi:10.3758/BF03194075
- Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 121–138. doi:10.1037/0096-1523.29.1.121
- Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, *24*, 295–340. doi:10.1016/0010-0285(92)90010-y
- Zelinsky, G. J., & Neider, M. B. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition*, *16*, 553–566. doi:10.1080/13506280802000752
- Zelinsky, G. J., & Todor, A. (2010). The role of “rescue saccades” in tracking objects through occlusions. *Journal of Vision*, *10*. doi:10.1167/10.14.29
- Zhang, W., & Luck, S. J. (2008). Discrete, fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235. doi:10.1038/nature06860
- Zhang, H., Xuan, Y., Fu, X., & Pylyshyn, Z. W. (2010). Do objects in working memory compete with objects in perception? *Visual Cognition*, *18*, 617–640. doi:10.1080/13506280903211142
- Zhao, L., Gao, Q., Ye, Y., Zhou, J., Shui, R., & Shen, M. (2014). The role of spatial configuration in multiple identity tracking. *PLoS ONE*, *9*, e93835. doi:10.1371/journal.pone.0093835
- 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. doi:10.1167/14.12.12
- Zhou, K., Luo, H., Zhou, T., Zhuo, Y., & Chen, L. (2010). Topological change disturbs object continuity in attentive tracking. *Proceedings of the National Academy of Sciences*, *107*, 21920–21924. doi:10.1073/pnas.1010919108