COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 39 (2015) 1622–1645

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12204

The Identity-Location Binding Problem

Piers D. L. Howe, Adam Ferguson

Psychological Sciences, University of Melbourne

Received 30 June 2013; received in revised form 14 August 2014; accepted 2 September 2014

Abstract

The binding problem is fundamental to visual perception. It is the problem of associating an object's visual properties with itself and not with some other object. The problem is made particular difficult because different properties of an object, such as its color, shape, size, and motion, are often processed independently, sometimes in different cortical areas. The results of these separate analyses have to be combined before the object can be seen as a single coherent entity as opposed to a collection of unconnected features. Visual bindings are typically initiated and updated in a serial fashion, one object at a time. Here, we show that one type of binding, location-identity bindings, can be updated in parallel. We do this by using two complementary techniques, the simultaneous-sequential paradigm and systems factorial technology. These techniques make different assumptions and rely on different behavioral measures, yet both came to the same conclusion.

Keywords: Binding problem; Attention; Visual perception; Feature integration theory; Tracking; Visual awareness

1. Introduction

The effortlessness of vision belies the challenges faced by the visual system. To accurately perceive a single object, we need to distinguish its properties from those of other objects in the visual scene and then associate its properties with itself, an issue known as the binding problem (von der Malsburg, 1981; von der Malsburg & Schneider, 1986; Milner, 1974; Treisman, 1996). For example, when looking at Fig. 1, you are probably aware that there are a number of crosses and that the colors red, blue, and green are present. However, you are probably not aware whether a particular cross has both a green vertical bar and a red horizontal bar without first attending to it and binding (i.e., associating) its features together. Without binding you can be aware of the individual features present in

---

Correspondence should be sent to Piers D. L. Howe, Psychological Sciences, University of Melbourne, Parkville, VIC 3010, Australia. E-mail: pdhowe@unimelb.edu.au

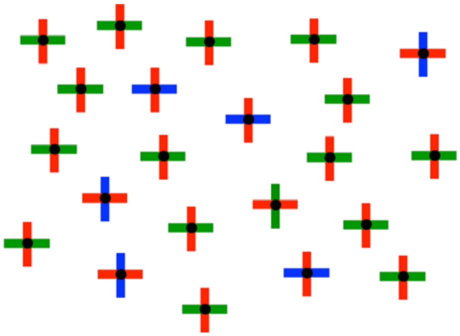

Fig. 1. In the first glance, you are probably aware that the scene contains a number of crosses and the colors red, green, and blue. However, to see the colors that make up a particular cross, you need to bind those features together. This process requires attention (Howe, Evans, et al., 2009; Treisman, 1996; Wolfe, 2012).

the scene (Treisman & Gelade, 1980; Wolfe & Bennett, 1997), but binding is required for you to be able to perceive whether or not an given object has a particular combination of these features (von der Malsburg, 1981; von der Malsburg & Schneider, 1986; Milner, 1974; Treisman, 1996). The binding problem is ubiquitous to visual perception and strongly constrains visual awareness (Crick & Koch, 1990; Howe, Evans, et al., 2009; Wolfe, 2012).

It has been suggested that the binding problem is solved by the serial application of attention (Treisman, 1996; Treisman & Gelade, 1980; Wolfe, 2012). In Fig. 1, you may have found yourself inspecting the items one item at a time. According to feature integration theory, object features such as color, size, and shape are processed preattentively and registered in parallel in separate feature maps. However, to perceive a complete object, one must first bind together, across the separate feature maps, those features that belong to the same object. Feature integration theory posits that this binding is achieved by attending to the objects sequentially, one at a time. Because the visual system analyzes only the features of the object to which it attends, this avoids the possibility of inadvertently combining features of different objects (Treisman & Gelade, 1980). Feature integration theory further predicts that if more than one object is simultaneously attended, then properties of different objects may be perceptually combined, thereby forming illusory conjunctions (Treisman & Schmidt, 1982). Consistent with this prediction, such illusory conjunctions are perceived to occur in situations where it is difficult for observers to confine their attention to just one object (Cohen & Ivry, 1989; Hazeltine, Prinzmetal, & Elliott, 1997; Treisman & Schmidt, 1982).

The binding problem comes in a number of different forms (Di Lollo, 2012; Treisman, 1996). The property binding problem describes those situations where an observer needs to associate different properties, or features, of the same object with each other (Treisman, 1996). For example, perceiving a red vertical bar requires the association of red with vertical. Conversely, the location-identity binding problem describes those situations where an observer associates a particular object identity with a particular location

(Treisman, 1996). For example, having found a red vertical, green horizontal cross, you need to bind its identity with its location to be able to point to it. In this article, we will concern ourselves with only the latter type of binding problem.

Here, we present evidence that shows that in some instances location-identity bindings can be updated in parallel. We do this by using the multiple identity tracking (MIT) paradigm (Horowitz et al., 2007; Oksama & Hyönä, 2004). This paradigm is closely related to the more common multiple object tracking paradigm (Alvarez & Cavanagh, 2005; Pylyshyn & Storm, 1988; Scholl, 2001; Yantis, 1992) but differs in that all the targets are distinct and have to be tracked individually, thereby necessitating that their location-identity bindings are continuously updated. In MIT the observer is shown a number of distinct objects, in our case different colored disks (Fig. 2). At the start of the trial four of them are briefly ringed to indicate that these are targets to be tracked. The objects then move around a computer monitor, and at the end of the trial the objects are briefly masked, all become identical, and the observer is asked to locate each target in turn. For example, the observer might first be asked to locate the target that was previously red. Because the observer did not know when the trial was going to end, he can only locate each target if he continuously associated each target with its ever-changing location. For this reason, this task requires the observer to continuously solve the location-identity binding problem for each target.

When tracking multiple objects, observers will often make eye movements. Often the observers will fixate the centroid of the tracked objects but, depending on the number of targets and the separation between the objects in the display, observers will sometimes also track the targets by fixating each one in turn (Fehd & Seiffert, 2008; Fehd & Seifert, 2010; Huff et al., 2010; Zelinksy & Neider, 2008). The latter strategy necessarily introduces a serial component into tracking. Because we were interesting in determining whether location-identity bindings can ever be updated in parallel, we needed to prevent

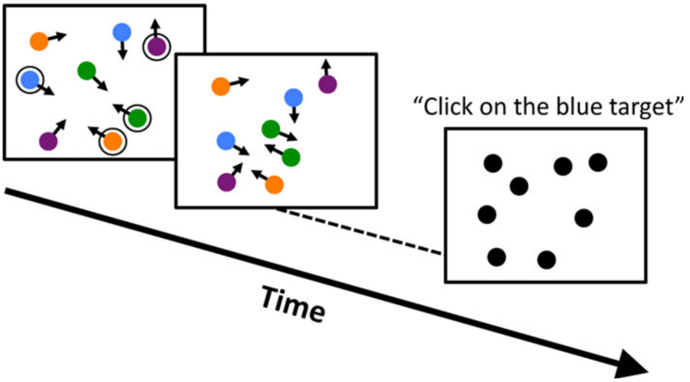

Fig. 2. The multiple identity tracking (MIT) paradigm. In MIT, a subset of the objects is initially ringed to indicate that these are the targets to be tracked. The objects then all move around. At the end of the trial, they are briefly masked by multicolored masks and then all become identical. The observer is then asked to locate a particular target (e.g., the blue one).

observers from fixating each target in turn. Consequently, we studied MIT in a situation where observers were required to maintain fixation on a fixation cross, so could not fixate each target in turn.

Currently, there is only one model of MIT, the model of multiple identity tracking (MOMIT; Oksama & Hyönä, 2008). According to this model, the identity-location bindings are refreshed and maintained by a serial mechanism and this is how observers are able to keep track of which target is where. MOMIT assumes that the observer can obtain the location information of all objects continuously and in parallel. However, this information is not indexed, so this cannot be used to determine which target is where. Instead, identity-location bindings are maintained and updated in visual short-term memory (VSTM) by the serial application of attention. To refresh a particular identity-location binding, MOMIT first inspects whichever object is closest to the previously remembered location for the target of the identity-location binding to be refreshed. If the inspected object is in fact the target of that particular identity-location binding, then the binding is refreshed with the new location information. Otherwise, MOMIT repeatedly attends to the other objects in the scene until the target is found, at which point its identity-location information is refreshed. While this procedure allows for a target to be recovered, the cost is that it then takes longer to refresh the other identity-location bindings, making it more likely that these targets will also be lost, so will also need to be recovered. This is why MOMIT only has time to track a finite number of targets. In an extended study, Oksama and Hyönä (2008) demonstrated that MOMIT can provide a good account of a wide range of behavioral data on MIT.

The work by Oksama and Hyönä (2008) would therefore seem to strongly argue that location-identity bindings are updated in series. However, related work in the multiple object tracking (MOT) paradigm argues against this conclusion. MOT is very similar to MIT except that in the former all the targets are identical to each other. At the start of an MOT trial the observer is shown a number of objects and a subset are briefly highlighted to indicate that these are the targets to be tracked. All the objects are physically the same, for example, they might all be gray disks. The objects then all move about for a few seconds after which the observer is asked to indicate the locations of the targets. Crucially, the observer does not need to distinguish the targets from each other as he would in the MIT paradigm.

There is considerable evidence that MOT occurs in parallel. For example, both Pylyshyn and Storm (1988) and Yantis (1992) attempted to construct serial models of MOT. Like the MOMIT model, both sets of these models assumed that the targets are attended to one at a time. When it was time to re-attend to a given target, the models assumed that whichever object was closest to that target's previous remembered position was the target. If the objects were moving too fast or the time taken to re-attend a target was too long, then this assumption would not be valid and a distractor would be mistaken for a target. Both studies concluded that serial models could not account for MOT performance without assuming that attention could be switched between targets at an unrealistically fast rate. Consequently, both studies rejected serial models in favor of a parallel account of MOT.

This conclusion was supported by two further studies. Alvarez and Cavanagh (2005) demonstrated that tracking occurs essentially independently in the left and right hemifields, a result that is inconsistent with a serial model that assumes a single focus of attention. To explain this result a serial model would at a minimum have to assume two independent attention foci, one for each hemifield (Cavanagh & Alvarez, 2005). Using the simultaneous-sequential paradigm, discussed below, Howe, Cohen, Pinto, and Horowitz (2010) provided direct evidence even against this type of serial model, concluding that in MOT tracking most likely occurs in parallel.

There is therefore an apparent contradiction in the literature. While there is strong evidence that MOT occurs in parallel, there is also evidence that MIT occurs in series. Given that the paradigms are so similar, it is unlikely that they would be achieved by different mechanisms. However, to date, no one has attempted to directly investigate whether MIT is achieved by a serial or by a parallel process. That is the aim of this study. By answering this question we will then be able to determine whether identity-location bindings can be maintained in parallel as MIT necessitates the continuous updating of identity-locations bindings.

1.1. The simultaneous-sequential paradigm

We have adapted the simultaneous-sequential paradigm (Eriksen & Spencer, 1969; Shiffrin & Gardner, 1972) to MIT, having previously applied it to MOT (Howe et al., 2010). Unlike the conventional MIT paradigm, in the simultaneous-sequential paradigm the disks do not move continuously. Instead, they repeatedly pause (Fig. 3). In the *simultaneous* condition, all the disks move and pause simultaneously. Conversely, in the *sequential* condition, the disks are divided into two groups, each group containing half the targets, and only one group moves at any one time. Crucially, in each condition, each disk moves and pauses the same number of times, does so for the same duration and travels the same total distance. In these respects the motion discontinuities are the same in the two conditions.

To illustrate the utility of this paradigm, let us consider what predictions both a parallel model with independent channels and a serial model with independent stages make. These models are often referred to as the standard parallel model and the standard serial model, respectively (Townsend & Asby, 1983). A standard parallel model assumes that all objects are continuously and independently tracked regardless of their state of movement (Alvarez & Franconeri, 2007; Pylyshyn & Storm, 1988). Thus, it would treat both conditions the same and consequently predict that the observers' tracking performance would be the same in both conditions.

Conversely, a standard serial model would predict tracking performance that should be greater in the sequential condition. For example, according to MOMIT, the identity-location bindings of the targets are updated in order of increasing location uncertainty, which in the case of our experiment would mean that the moving targets would be updated preferentially as their location uncertainty is usually the greatest. Thus, in the sequential condition, MOMIT would preferentially attend to the two moving disks. So its effective

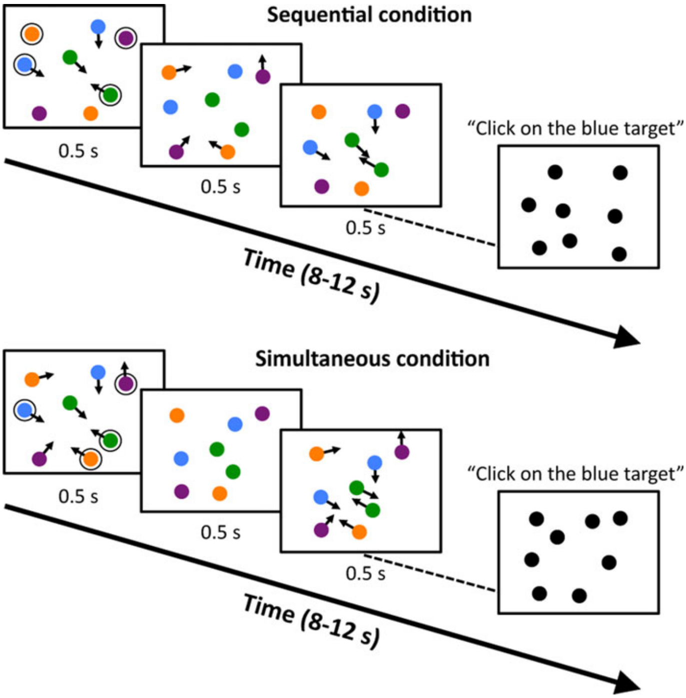

Fig. 3. Stimulus display used in the sequential and simultaneous conditions of Experiment 1. Four targets were presented in four different colors (orange, green, blue, and purple), each with a color-matched distractor. At the start of the trial, targets were briefly identified with black rings. All objects moved and paused alternately. The duration of each movement/pause phase was 0.5 s. In the sequential condition, only half of the objects moved at one time. In the simultaneous condition, all objects moved and paused in synchrony. Trial length randomly varied between 8 and 16 s. At the end of each trial, objects were masked and observers identified each target in turn in a random, unpredictable order.

tracking load would be approximately two disks. Conversely, in the simultaneous condition, it would need to attend to all four disks, thereby approximately doubling its effective tracking load. Because halving the effective tracking load allows MOMIT to track disks at approximately double the speed than it could otherwise, MOMIT predicts that an observer should be able to track the disks at approximately twice the speed in the sequential condition as it can in the simultaneous condition.

In the above analysis, we have focused on MOMIT as an example of a serial model of MIT. However, a similar analysis would hold for any serial model that assumes that attention is directed preferentially to the moving targets. Such an assumption is reasonable for three reasons. First, it is known that motion onsets preferentially direct attention to moving targets (Abrams & Christ, 2003). Second, there is both EEG (Drew, Horowitz, Wolfe, & Vogel, 2011) and fMRI (Howe, Horowitz, Morocz, Wolfe, & Livingstone, 2009) evidence that during tracking attention is indeed preferentially directed to moving targets as opposed to stationary ones. Third, a previous MIT study has shown that observers preferentially attend to those targets that are in danger of being lost (Iordanescu, Grabowecky, & Suzuki, 2009), which in this case would typically be the moving targets. Thus, while our analysis is expressed in terms of MOMIT, it is likely to apply to any serial account of MIT.

1.2. Systems factorial technology

Although the simultaneous-sequential paradigm has been commonly used to determine whether various psychological processes are serial or parallel (Eriksen & Spencer, 1969; Howe et al., 2010; Huang, Treisman, & Pashler, 2007; Shiffrin & Gardner, 1972), it is not definitive (Townsend & Asby, 1983). As discussed above, it relies on the assumption that attention will be directed preferentially to moving objects. For this reason we sought to obtain converging evidence by using a complementary approach, systems factorial technology (Townsend, 1992; Townsend & Nozawa, 1995). Because these two approaches make different assumptions and are based on different behavioral measures (accuracy vs. reaction times), if both arrive at the same conclusion then that would constitute very strong evidence for that conclusion.

Systems factorial technology is an extension of the Sternberg additive factors methodology (Sternberg, 1969) and is a common technique for distinguishing between parallel and serial processing (Fific & Townsend, 2010; Fific, Townsend, & Eidels, 2008; Houpt & Townsend, 2010; Nozawa, Reuter-Lorenz, & Hughes, 1994; Sung, 2008; Townsend & Fific, 2004; Townsend & Nozawa, 1995; Yamani, McCarley, Mounts, & Kramer, 2012). Consequently, we will give only a brief outline of how it works and refer the reader to the original sources for mathematic proofs (Townsend & Nozawa, 1995; Townsend & Wenger, 2004).

The observers perform an MIT experiment similar to before. Partway through the movement phase, the two targets might darken either by a small (S) amount, by a large (L) amount, or not at all (N). If either the targets darkened, the observer is instructed to push a response button as quickly as possible. In the case that both targets did darken, they may both do so by a large amount (the LL condition), one might darken by a large amount and the other by a small amount (a mixed trial), or both might darken by a small amount (the SS condition)[1]. The more salient the change, the faster it is likely to be processed, and the faster the reaction time is likely to be. By measuring the reaction time distributions it can be determined whether the two probed targets were processed in series or in parallel (Townsend & Nozawa, 1995). A simple way to do this would be to measure the mean reaction time for each of the three conditions and then calculate the mean interaction contrast (MIC; Houpt, Blaha, McIntire, Havig, & Townsend, 2013), sometimes also referred to as the interaction contrast of the means (Townsend & Wenger, 2004):

$$MIC = RT_{LL} + RT_{SS} - 2RT_{mixed} \tag{1}$$

where RT_{LL} is the mean reaction time for the LL condition, etc. A serial model predicts that since each target is analyzed in turn the total reaction time is equal to the sum of the reaction times for the detection of the darkening of each of the two targets. It follows that the MIC would therefore be equal to zero (Townsend & Wenger, 2004). Conversely, a MIC < 0 would exclude a serial account (Townsend & Wenger, 2004).

The problem with this analysis is that on its own the MIC cannot distinguish between a parallel process and a coactive process. In a coactive process the decision threshold is applied to the summation of the activity of the two inputs, rather than having a separate decision threshold for each input (Houpt et al., 2013). To distinguish between coactive processes and parallel processes, one needs to consider the survivor interaction contrast (SIC) (Townsend & Nozawa, 1995).

The survivor function, *S(t),* is the probability that an event, such as an observer's response, has not happened by time *t*. Letting *X* designate the occurrence of an observer's response, we can write the survivor function as

$$S_X(t) = \Pr\{X > t\} \tag{2}$$

Using the survivor functions for the three conditions listed above, we can calculate the survivor interaction contrast:

$$SIC = S_{LL}(t) + S_{SS}(t) - 2S_{mixed}(t) \tag{3}$$

In our experiment the observer was required to press a key whenever either target darkened. It follows that if both targets darken, the observers need notice only the darkening of just one of them before he can respond. In the literature such a task is described as obeying a self-terminating stopping rule to signify that in some circumstances the observer is able to make a decision without considering both inputs (Townsend & Nozawa, 1995). This stopping rule is sometimes also referred to as an OR stopping rule in analogy to the OR logic gate (Houpt et al., 2013).

The alternative to the self-terminating stopping rule is the exhaustive stopping rule (Townsend & Nozawa, 1995), otherwise known as the AND stopping rule (Houpt et al., 2013). According to this stopping rule, the observer is required to inspect both inputs before making a decision. Although our task did not require observers to do this, they might still have done so. As discussed below, we can use the SIC function to determine if this was the case.

In Fig. 4 we show the predicted SIC function for each of the two potential stopping rules (exhaustive or self-terminating) and for both serial and parallel processing (Townsend & Nozawa, 1995). Fig. 5 shows the prediction for the coactive model (Eidels, Houpt, Altieri, Pei, & Townsend, 2011; Townsend & Nozawa, 1995). Because coactive processing combines the two inputs before making a decision, it only ever makes just a single decision, so the concept of a stopping rule does not apply in its case.

Another useful measure that can characterize the operation of a system is its capacity. Capacity is a measure of the ability of a system to process stimuli simultaneously. It is calculated by measuring the change in the processing speed of a stimulus when it is simultaneously presented with additional stimuli. If there is no decrease in the speed of processing of the first stimulus when a second stimulus is presented, the system is said to

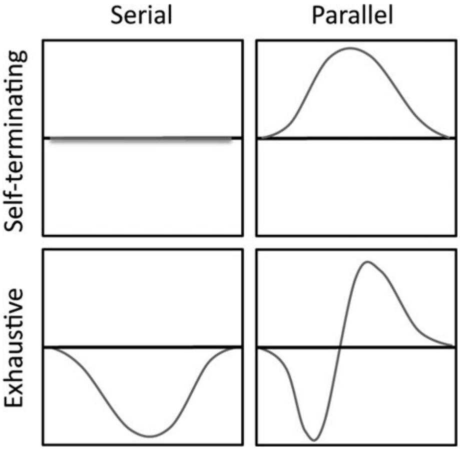

Fig. 4. Representative predictions for the survivor interaction contrast for serial and parallel processing for each of the two potential stopping rules.

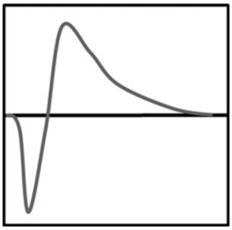

Fig. 5. Representative prediction for the survivor interaction contrast for a coactive process.

have unlimited capacity, denoted by a capacity coefficient of one. If the presentation of additional stimuli slows the processing of the first stimulus, the system is said to have limited capacity, denoted by a capacity coefficient < 1, but > 0. Alternatively, the speed of processing of the first stimulus might be increased by the presentation of additional stimuli. Such a system is said to have super capacity, denoted by a capacity coefficient > 1.

The actual formula used to calculate the capacity coefficient depends on the stopping rule required by the task. Because our task required an OR stopping rule, the formula we used to calculate capacity is given below:

$$C_{OR}(t) = \frac{\overset{\frown}{H}_{1\ldots n}(t)}{\sum_{i=1}^{n} \overset{\frown}{H}_{i}(t)} \tag{4}$$

where $\overset{\frown}{H}_{1\ldots n}(t)$ is the cumulative hazard function when all sources are present and $\overset{\frown}{H}_i(t)$ is the cumulative hazard function when only the *ith* source is present (Townsend & Nozawa, 1995).

2. Experiment 1: Simultaneous-sequential paradigm with targets whose unique identities were continuously visible

The purpose of this experiment was to apply the simultaneous-sequential technique to the MIT paradigm to determine whether the targets could be tracked in parallel. If they can, this would show that identity-location bindings can be updated in parallel.

2.1. Participants

The study consisted of 15 participants (5 female) aged 18–30 years ($M = 23.4$) with normal or corrected-to-normal visual acuity as verified by a Good-Lite near vision eye chart. All reported having normal color vision. These were all University of Melbourne undergraduate or graduate students, who were reimbursed \$15/h for their time. All provided informed written consent and the experiments were approved by the Department of Human Ethics Advisory Group in the School of Psychological Sciences at the University of Melbourne.

2.2. Apparatus and stimuli

The Stimuli were presented using the Psychophysics toolbox version 3 (Brainard, 1997; Pelli, 1997) and were presented 60 cm from the observer on a CRT monitor with 30° by 40° white display area. The experiment utilized the simultaneous-sequential paradigm as illustrated in Fig. 3. Eight disks, each subtending 1 degree (°) of visual angle, were presented. Two were orange, two were green, two were blue, and two were purple. A small, gray central fixation cross, subtending $1°\times1°$, was visible throughout each trial.

2.3. Procedure

Each participant took part in a total of 10 practice trials and 80 experimental trials, equally divided between the two conditions (simultaneous and sequential), presented in a random order. Participants were encouraged to take breaks between trials as needed. Participants were instructed to maintain fixation on the central cross throughout each trial to ensure that they did not sequentially foveate the disks. (Had they not obeyed this instruction, performance would have been superior in the sequential condition, the opposite of what was observed.) At the beginning of each trial, all objects were displayed and immediately began moving. Four of the disks were ringed with black circles for 4 s to indicate that these were the targets to be tracked. Each target was a different color. Disks moved and paused alternately (in either simultaneous or sequential fashion, as described in Fig. 3) for a random duration between 8 and 16 s. When in motion, disks moved in straight lines except that each disk and the fixation cross was surrounded by an invisible buffer that prevented the center-to-center separation of any two disks becoming $< 4°$ or any disk approaching within 4° of the fixation cross. At the end of each trial, all disks ceased moving and were masked with multicolor concentric ring masks for 0.25 s before turning black. The participants were then asked to click on a particular color target (e.g., the blue one) via a text instruction at the top of the screen. This request was repeated so that in total two targets were quizzed. The two targets to be quizzed were chosen randomly on each trial. The faster the targets move, the harder it is to track them (Alvarez & Franconeri, 2007; Bettencourt & Somers, 2009). The QUEST staircase routine (Watson & Pelli, 1983) was used to find the maximum speed at which, for each observer, the targets could still be tracked in each condition. The maximum tracking speed was defined to be that at which tracking accuracy was at 75%. Maximum tracking speed was used as the measure of tracking performance because, as discussed above, serial models and parallel models make very different predictions as to the maximum speed at which disks can be successfully tracked in the two conditions. It has been argued that it is not the tracking speed but the total distance traveled by the objects that determines the tracking accuracy (Franconeri, Sumeeth, & Scimeca, 2010). Because the trial duration was held constant, tracking speed correlated perfectly with the total distance traveled, so this possibility does not materially affect our analysis.

When piloting this experiment, it was found that some participants would sometimes verbally recite the color names while performing the tracking task. To prevent the use of this verbal rehearsal strategy, participants were required to continuously recite the word "animal" aloud throughout the duration of each trial. Observers were continuously monitored to ensure compliance with this verbal recitation task. Participants were not provided with feedback on their performance.

2.4. Results

As discussed previously, a standard serial model would predict that the maximum tracking speed in the sequential condition should be double that of the simultaneous condition. The results are shown in Fig. 6 and contradict this prediction. A within-subjects *t*-test showed no significant difference between the average maximum tracking speeds in the simultaneous and sequential motion conditions, $t(14) = 0.71$, $p = 0.492$, *Cohen's* $d = 0.217$ (paired *t*-test, equal variance not assumed, predetermined alpha level = 0.05). This finding is consistent with a parallel process.

3. Experiment 2: Simultaneous-sequential paradigm with targets whose unique identities were not continuously visible

In the previous experiment the identities of the objects were continuously visible during the tracking phase to mimic real life where the identities of objects are also continuously visible. Most MIT experiments have adopted this procedure (Cohen, Pinto, Howe, & Horowitz, 2011; Horowitz et al., 2007; Iordanescu et al., 2009; Oksama & Hyönä, 2008; Pinto, Howe, Cohen, & Horowitz, 2010). A drawback with this paradigm is that there is the potential for the observer to realize when a tracking error has occurred and to

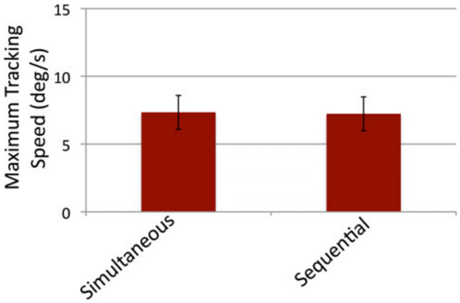

Fig. 6. Maximum speed at which the targets could be tracked in the two conditions in Experiment 1. Error bars, both here and in figures for subsequent experiments, represent the Cousineau–Morey within-subjects standard error of the mean (Cousineau, 2005; Morey, 2008).

rectify it. Indeed, MOMIT explicitly makes the assumption that observers will do this (Oksama & Hyönä, 2008). Consequently, the observer's performance on this task might reflect not only the observer's tracking ability but also his or her ability to recover from errors. Because we wanted a pure measure of tracking ability, in Experiment 2 we used a version of the MIT paradigm where all the objects became the same color after the initial target identification phase so that during the tracking phase they were all identical (Hudson, Howe, & Little, 2012; Pylyshyn, 2004), thereby making it impossible for the observer to recover from tracking errors. It is important to note that tracking in this experiment still required observers to continuously refresh the location-identity bindings of the targets in VSTM. Just because the target identities are not visible does not mean that location-identity bindings are not refreshed. Indeed, if observers were not able to refresh (i.e., update) these bindings, then they would be unable to identify the targets at the end of the trial.

We were also interested in determining if our results would depend on the number of targets the observer tracked. We therefore varied the target number between two and four targets and adopted a two-way factorial design target number (two or four targets) × motion condition (simultaneous vs. sequential).

3.1. Participants

Fifteen participants (six female) aged 18–33 years (M = 22.8) took part. As before, all had normal or corrected-to-normal vision and gave informed consent. None had participated in the previous experiment.

3.2. Apparatus and stimuli

The apparatus was the same as before. To address the possibility that the verbal recitation task used in Experiment 1 might have influenced tracking behavior, this verbal recitation was not used in this experiment. As this made tracking easier, the number of disks was increased to sixteen to maintain overall tracking difficulty. All disks were black

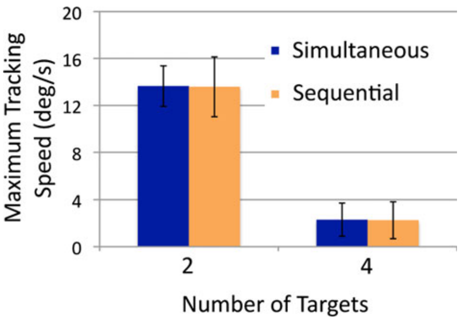

Fig. 7. Maximum tracking speed for the four conditions in Experiment 2.

except at the start of the trial when either two or four disks briefly became colored, each a different color, to indicate that these were the targets to be tracked. During the tracking phase, all the disks were identical.

3.3. Procedure

The procedure was similar to before. Each participant took part in a total of 10 practice trials and 160 experimental trials, equally divided between the four conditions, presented in a random order. We employed a 2 × 2 design where we crossed type of motion (simultaneous vs. sequential) with number of targets (2 or 4). For each of the four conditions, the QUEST staircase procedure was used to find the maximum tracking speed at which tracking accuracy was at 75%. As before, participants were instructed to maintain fixation on the central cross throughout each trial to ensure they did not sequentially foveate the disks.

3.4. Results

The results are shown in Fig. 7. The dependent variable is tracking speed. A two-way within-subjects analysis of variance (ANOVA) found a significant main effect for target number, with tracking performance significantly better in the two target conditions than in the four-target conditions, $F(1, 14) = 75.12$, $p < .001$, *partial* $\eta^2 = 0.86$. However, no main effect was found for motion type (simultaneous or sequential), $F(1, 14) = 0.01$, $p = .93$. Nor was there evidence of an interaction between motion type and target number, $F(1, 14) = 0.02$, $p = .902$.

This experiment confirmed the result of the previous experiment in that we found tracking performance not to be significantly different in the simultaneous and sequential conditions. This did not depend on the number of targets that were tracked. This null result is unlikely to have been caused by our experiment being underpowered because we were able to show a large effect of target number on tracking performance. As discussed previously, a standard serial model would predict that type of motion (i.e., sequential vs. simultaneous) should affect tracking performance as much as target number. Consequently, if we can find very strong evidence for the latter, we should be able to find evidence for the former had tracking been serial. Thus, this experiment also argues against a serial account of identity tracking.

4. Experiment 3: Converging evidence using systems factorial technology

The previous two experiments both used the same technique, the simultaneous-sequential paradigm, and found evidence that MIT occurs in parallel. In this experiment we used a complementary technique, systems factorial technology, to investigate the same question. To use systems factorial technology it was necessary to make all the objects identical during the tracking stage. However, this does not mean that observers did not need to continuously refresh the identity-location bindings. As explained earlier, unless the observer continuously refreshed these identity-location bindings, the observer will not know where the targets are at the end of the trial.

4.1. Participants

There were three participants (two female) aged 22–37 years (M = 27.0), with one of the authors (PH) participating. As before, all had normal or corrected-to-normal vision and had provided informed consent. None had participated in the previous two experiments.

4.2. Apparatus and stimuli

The apparatus was the same as before. There were two sets of four disks, with each set of disks confined to stay within its own 10° × 10° white square, bounded on all sides by a light gray border, located either to the left or right of a black 1° × 1° fixation cross. Each disk subtended 0.75° and moved at a preset speed 4°/s (i.e., the disks all moved continuously, they did not pause, and no QUEST staircase procedure was used in this experiment). Disks were surrounded by invisible buffers that prevented the center-to-center spacing between two disks becoming < 1.5°. Except at the beginning of the trial, all disks were light gray and identical to each other.

4.3. Procedure

At the start of the trial one disk turned red and another disk turned green to indicate that these were the targets to be tracked. Because it has previously been shown that tracking occurs quasi-independently in the left and right visual hemifields (Alvarez & Cavanagh, 2005; Hudson et al., 2012), both targets always occupied the same hemifield to avoid any possible hemifield-independence effects. Both disks then reverted to gray, so were identical to the other disks in the display and to each other. They then continued to move around the display. No verbal recitation task was used.

During the movement phase, each of the two target disks could darken by either a large amount (L), a small amount (S) or might not darken at all (N). (In each trial each disk could darken only once and if both target disks darkened they would do so simultaneously). We employed a double factorial paradigm, so in total there were nine conditions, as illustrated below. In total, each observer completed 2,250 trials with each condition being presented equally often. This ensured that there was no correlation between the darkening of one target and the darkening of the other target. The double factorial paradigm is a standard, recommended procedure, which allows us to calculate the mean interaction contrast, the survivor interaction contrast, and the capacity all from a single data set (Townsend & Nozawa, 1995). As discussed previously, in the moving phase there was no physical difference between the two target disks and they were each equally likely to occupy any region of the hemifield. For this reason, the LS condition and the SL condition are equivalent, so were combined and are referred to as a mixed trial in our subsequent analysis (Table 1).

If *either* of the disks that darkened were targets, then the observer was instructed to push a response button as quickly as possible. The response time was defined as the time from the initial darkening of the disk to the time the observer pushed the response button. If the observer pushed the response button, all the disks then stopped moving and the observer was asked to identify the location of a particular target, for example, the green one. In each trial, only one darkening episode occurred. After each trial, feedback was provided.

In the case that both the disks that darkened were targets, there were three possibilities. Either both target disks darkened by a large amount (a LL trial), one target darkened by a large amount and the other by a small amount (a mixed trial), or both targets darkened by only a small amount (a SS trial). For each of the three trial types the distribution of reaction times was recorded and converted into survivor functions and analyzed using the sft toolbox in the R statistical processing package (Houpt et al., 2013).

4.4. Results

Observers were able to identify the location of the queried target at the end of the trial with 95% accuracy averaged across the three observers, indicating that observers diligently maintained the identity-location binding for each target. We discarded any trial where the observer failed to correctly identify the location of the queried target at the end of the trial or where the observer responded but neither the targets had darkened (the average false-positive rate was just 2.7%). This ensured that we considered only trials where the observers had managed to maintain accurate identity-location bindings.

Central to systems factorial technology is the assumption that the manipulation of an input will affect the processing of that input but will not affect the processing of the other

Table 1
Showing the nine different conditions of the double factorial paradigm. Targets might darken by a large amount (L), a small amount (S), or might not darken at all (N)

Target 1, Target 2			
	L,L	L,S	L,N
	S,L	S,S	S,N
	N,L	N,S	N,N

input. This assumption is usually referred to as effective selective influence (Townsend & Nozawa, 1995). For our experiments, it means that the darkening of one of the targets should affect the processing of that target, that is, the detection of the darkening of that target, but not the processing of the other target. Effect selective influence implies that the survivor functions should be ordered as follows:

$$S_{LL}(t) < S_{mixed}(t) < S_{SS}(t) \tag{5}$$

Because it is not possible to directly test for selective influence, it is instead necessary to test for the implied ordering of survivor functions (Houpt et al., 2013). The survivor functions for the three observers are shown in top row of Fig. 8. For each observer we performed a series of Kolmogorov–Smirnov tests to test for the above predicted ordering, relative to the null hypothesis of equivalence. At the 0.05 level of significance we found that, for all observers, $S_{LL} < S_{mixed}$ *and* $S_{mixed} < S_{SS}$ (see the appendix for the results of the individual observers). We also found that for none of the observers were either of the following tests significant: $S_{LL} > S_{mixed}$ *and* $S_{mixed} > S_{SS}$. This demonstrates the ordering implied by effective selective influence.

For each observer, we calculated the mean interaction contrast (MIC). For all observers this was > 0 at the 0.05 significance level (adjusted rank transform test, average MIC = 0.0455). This allows us to reject serial processing but does not allow us to distinguish between parallel and coactive processing. To do this, we need to consider the survivor interaction contrasts shown in the bottom row of Fig. 8. Performing the Houpt–Townsend KS-SIC test at a significance level of 0.05 showed that each of these was at some point significantly > 0, but none of them was ever significantly < 0 (see Appendix). This allowed us to rule out coactive

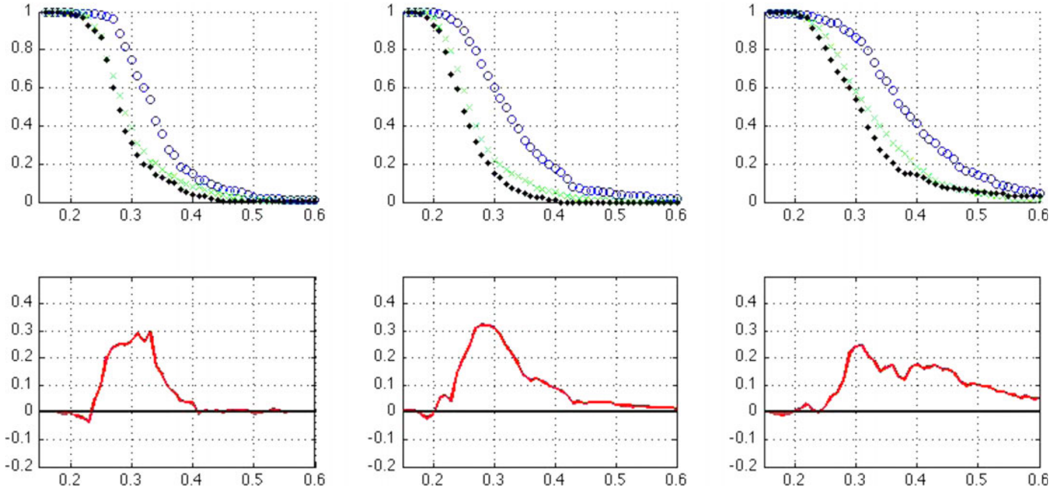

Fig. 8. Top row: survivor functions. The black dots represent the LL condition, the green crosses represent the mixed condition, and the blue circles represent the SS condition. Bottom row: survivor interaction contrasts.

processing and conclude that the processing was parallel with a self-terminating stopping rule, that is, an OR stopping rule (see Fig. 4 for further details).

We then calculated the capacity coefficient (Fig. 9). For all three observers, the capacity coefficient was found to be significantly < 0 under the Houpt–Townsend UCIP test. As such, our data are compatible both with an inhibitory, interactive parallel system (Eidels et al., 2011) and with a limited capacity independent parallel system (Townsend & Asby, 1983). Our data cannot be used to distinguish between these two types of model. This finding of a limited capacity is consistent with other studies that have also showed that tracking capacity is limited in that increasing the number of objects to be tracked; that is, increasing the tracking load makes it harder to keep track of the objects (Alvarez & Franconeri, 2007).

5. Discussion

The above three experiments have demonstrated that multiple identity tracking (MIT) is mediated by a parallel process. As MIT necessitates the continuous updating of identity-location bindings, this shows that identity-location bindings can be updated in parallel. This is surprising because other forms of the binding problem, such as the property binding problem, which is the mutual association of two properties of the same object, are thought to be mediated by a serial process in which only one object is attended at a time, thereby preventing attributes of different objects being conjoined (Treisman, 1996).

Experiments 1 and 2 utilized the simultaneous-sequential technique. This technique rests on the assumption that attention will be directed preferentially to the moving objects. On the one hand, it is known that motion onsets preferentially direct attention to moving targets (Abrams & Christ, 2003), and a previous MIT study has shown that observers preferentially attend to those targets that are in danger of being lost (Iordanescu et al., 2009), which in this case would be the moving targets. On the other hand, directing attention to the moving targets would require attention to be switched between the targets every 250 ms. While this would certainly imply rapid switching, the required switching frequency is not implausible (Egeth & Yantis, 1997; Horowitz, Wolfe, Alvarez, Cohen, & Kuzmova, 2009; Saarinen & Julesz, 1991; Wolfe, Alvarez, & Horowitz, 2000).

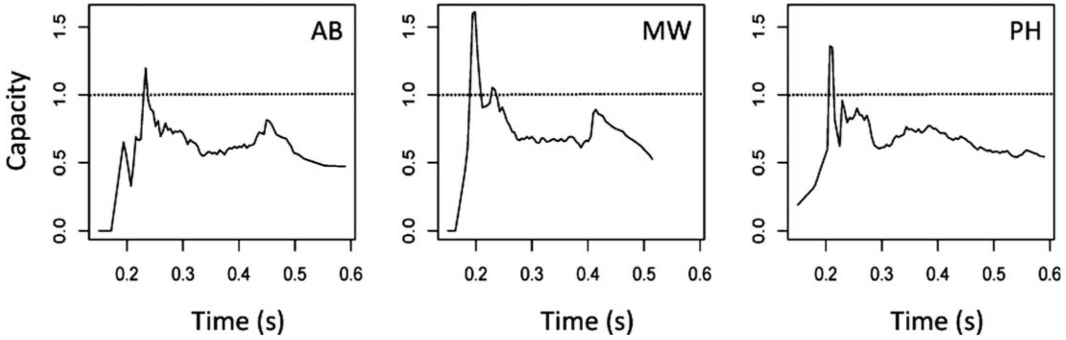

Fig. 9. The capacity coefficient as a function of time for the three observers.

In fact, Hogendoorn, Carlson, and Verstraten (2007) have provided evidence for much faster attentional switching times in the context of tracking, reporting that attention can sometimes be shifted between objects in as little as 30 ms.

Another possibility is that although observers can switch their attention between targets in a rapid fashion, they might not be able to direct their attention preferentially to the moving targets because they still need to continuously refresh the location-identity bindings of the stationary targets. We believe that this possibility is unlikely as the location-identity binding would need to be refreshed more often for moving targets as their location information becomes outdated more rapidly (Oksama & Hyönä, 2008). Consistent with this assumption, there is both EEG (Drew et al., 2011) and fMRI (Howe, Horowitz et al., 2009) evidence that during tracking attention is preferentially directed to moving targets as opposed to stationary ones. As such, Experiments 1 and 2 argued strongly against a serial account of identity tracking and in favor of a parallel account, albeit one with a limited capacity.

Experiment 3 utilized a different paradigm, systems factorial technology (Townsend & Nozawa, 1995), to address the same question from a different perspective. As discussed previously, this approach is complementary in that it makes different assumptions to the previous paradigm. In this paradigm the observer needed to both track and monitor the two targets, pressing a key whenever either of them darkened. Using this paradigm, we were able to demonstrate that both targets were attended simultaneously (Townsend & Wenger, 2004). Because it is generally accepted that tracking is mediated by attention (Cavanagh & Alvarez, 2005; Oksama & Hyönä, 2008), this data argue also against a serial account of MIT. For example, the MOMIT model explicitly assumes that tracking is achieved by attending to each target in turn, one target at a time, in a serial fashion. As such, it cannot account for our finding that the two targets were attended simultaneously. Indeed, no previously proposed serial model of tracking can account for our data because they all assume that attention is directed to only one object at a time (Pylyshyn & Storm, 1988; Tripathy, Ogmen, & Narasimhan, 2011; Yantis, 1992).

From our data, it is clear that the targets are both attended and tracked in parallel. There are a number of models of tracking that would appear to be consistent with this finding (Alvarez & Franconeri, 2007; Bettencourt & Somers, 2009; Cavanagh & Alvarez, 2005; Kazanovich & Borisyuk, 2006; Vul, Frank, Tenebbaum, & Alvarez, 2009). However, these models were designed only to address the MOT paradigm, so in their current state are unable to assign a unique identity to each target as is required for MIT. In principle there seems to be no reason why they could not be extended to do so.

5.1. Capacity limitations

Our third experiment demonstrated a capacity limitation: detecting a change to one target reduced the observer's ability to detect a change to a second target. This finding is consistent with previous studies that have shown that tracking one target makes it harder to track a second target, implying that tracking in general is capacity limited (Alvarez & Franconeri, 2007). It has previously been suggested that although targets are tracked in parallel, they are not attended continuously (Vul et al., 2009). Instead, they are periodically sampled with tracking accuracy determined by the sampling frequency. If this sampling frequency were to decrease as more targets were tracked, this would explain why tracking difficulty increases with an increasing number of targets (Vul et al., 2009). At least two previous studies have found evidence that attentional sampling frequencies do in fact decrease with increasing number of targets (Holcombe & Chen, 2013; Howard, Masom, & Holcombe, 2011).

5.2. *Formation versus maintenance of identity-location bindings*

While this study has shown that identity-location bindings can be maintained in parallel, this does not speak to the issue of whether their initial formation occurs in parallel. An experiment by Huang et al. (2007) addressed this issue using the simultaneous-sequential paradigm. Observers were briefly presented with two disks, each having a different color. These disks were either presented simultaneously or sequentially. After this initial presentation, the disks were masked, and the observer was shown a color and asked if either disks matched this color. It was found that accuracy was higher in the sequential condition than in the simultaneous condition, implying that the two disks were analyzed sequentially (Huang et al., 2007). Although there is now some debate whether this result applies when there are only two objects in the scene (Mance, Becker, & Liu, 2012), there seems to be general agreement that the processing is serial if there are more than two objects, providing each object has a different color. Thus, it seems that the initial analysis of objects generally occurs sequentially.

This finding could be explained in terms of object files (Kahneman, Treisman, & Gibbs, 1992). Object files are proposed to be episodic representations, distinct from long-term memory, within which successive states of an object are linked and integrated. An object file can contain a range of information about a specific object such as its features and location, but is addressed only via its location. It is assumed that the set-up of an object file is determined by attention (Kahneman et al., 1992) and occurs in a serial manner, one object at a time (Treisman & Gelade, 1980). The perception of multiple objects would require the construction of multiple object files, one for each object, so would be a serial processes, consistent with the findings of Huang et al. (2007). After the initial set-up, each object file would need to be continuously refreshed so that it could accurately represent the current state of the object, including any location changes. If the refreshing of the location information in these object files could occur in parallel, then this would explain the finding of our study that multiple identity-locations bindings can be maintained in parallel. Further work is needed to test this speculation. In particular, one would need to use systems factorial technology to test the claim that the initial representations of objects can be formed only one object at a time.

6. Conclusion

When an object is viewed, different features of that object are processed independently (Treisman & Gelade, 1980; Wolfe, 2012). To achieve a valid perception, the features of an object must be cognitively associated with it and distinguished from those of other objects (Treisman, 1996; Wolfe & Cave, 1999). It is assumed that such binding can generally occur only one object at a time (Treisman, 1996; Treisman & Gelade, 1980). As MIT requires the binding of each target's unique identity to its constantly changing location, it was assumed that this too could only be achieved one item at time, necessitating tracking to occur in a serial fashion as formalized, for example, by the MOMIT model (Oksama & Hyönä, 2008). We tested this assumption using both the simultaneous-sequential paradigm (Eriksen & Spencer, 1969; Shiffrin & Gardner, 1972) and systems factorial technology (Townsend & Nozawa, 1995; Townsend & Wenger, 2004). Because these two techniques make different assumptions and are based on different behavioral measures, if they both arrive at the same conclusion then this would amount to strong evidence for that conclusion. Both showed that identity tracking, and thus the updating of location-identity bindings, can occur in parallel.

Acknowledgments

This manuscript was funded by research grants to PH from the University of Melbourne. We thank Charlotte Hudson for assistance in the data collection.

Note

1. Traditionally one distinguishes between the LS condition, where the first target darkens by a large amount and the second darkens by a small amount, and the SL condition where the converse occurs. Because in our experiments both targets were equivalent in that they could both move anywhere within the visual hemifield, this distinction is not helpful. These two conditions were therefore combined and jointly referred to as the "mixed" condition.

References

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, *14*, 427–432.

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*, 637–643.

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*(14), 1–10.

Bettencourt, K. C., & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision*, *9*, 9.

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

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

Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 650–663.

Cohen, M. A., Pinto, Y., Howe, P. D., & Horowitz, T. S. (2011). The what-where tradeoff in multiple identity tracking. *Attention Perception and Psychophysics*, *73*, 1422–1434.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*(1), 42–45.

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Semin Neuroscience*, *2*, 263–275.

Di Lollo, V. (2012). The feature-binding problem is an ill-posed problem. *Trends in Cognitive Sciences*, *16*, 317–321.

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. *Journal of Neuroscience*, *31*, 659–668.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*, 269–297.

Eidels, A., Houpt, J. W., Altieri, N., Pei, L., & Townsend, J. T. (2011). Nice guys finish fast and bad guys finish last: Facilitatory vs. inhibitory interaction in parallel systems. *Journal of Mathematical Psychology*, *55*, 176–190.

Eriksen, C. W., & Spencer, T. (1969). Rate of information processing in visual perception: Some results and methodological considerations. *Journal of Experimental Psychology*, *79*, 1–16.

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

Fehd, H. M., & Seiffert, A. E. (2010). Looking at the center of the targets helps multiple object tracking. *J Vis*, *10*(4), 1911–1913.

Fific, M., & Townsend, J. T. (2010). Information-processing alternatives to holistic perception: Identifying the mechanisms of secondary-level holism within a categorical paradigm. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *36*, 1290–1313.

Fific, M., Townsend, J. T., & Eidels, A. (2008). Studying visual search using systems factorial methodology with target-distractor similar as the factor. *Perception and Psychophysics*, *70*, 583–603.

Franconeri, S., Sumeeth, J., & Scimeca, J. (2010). Tracking multiple objects is limited only by object spacing, not speed, time or capacity. *Psychological Science*, *21*, 920–925.

Hazeltine, R. E., Prinzmetal, W., & Elliott, W. (1997). If it's not there, where is it? Locating illusory conjunctions. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 263–277.

Hogendoorn, H., Carlson, T. A., & Verstraten, F. A. (2007). The time course of attentive tracking. *Journal of Vision*, *7*(2), 1–10.

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*(1), 1–19.

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.

Horowitz, T. S., Wolfe, J. M., Alvarez, G. A., Cohen, M. A., & Kuzmova, Y. I. (2009). The speed of free will. *Quarterly Journal of Experimental Psychology* (Colchester), *62*(11), 2262–2288.

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., & Townsend, J. T. (2013). Systems factorial technology with R. *Behavior Research Methods*, *46*(2), 307–330.

Houpt, T. C., & Townsend, J. T. (2010). The statistical properties of the survivor interaction contrasts. *Journal of Mathematical Psychology*, *54*, 446–453.

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

Howe, P. D., Cohen, M. A., Pinto, Y., & Horowitz, T. S. (2010). Distinguishing between parallel and serial accounts of multiple object tracking. *Journal of Vision 10*(8), article 11.

Howe, P., Evans, K., Pedersini, R., Cohen, M., Horowitz, T., & Wolfe, J. (2009). Attention: Selectice attention and consciousness. In *Encyclopedia for consciousness* (pp. 61–75). Oxford, England: Elsevier.

Howe, P. D., Horowitz, T. S., Morocz, I. A., Wolfe, J. M., & Livingstone, M. S. (2009). Using fMRI to distinguish components of the multiple object tracking task. *Journal of Vision*, *9*, 1–11.

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science*, *317*, 823–825.

Hudson, C., Howe, P. D., & Little, D. R. (2012). Hemifield effects in multiple identity tracking. *PLoS ONE*, *7*, e43796.

Huff, M., Papenmeier, F., Jahn, G., & Hess, R. F. (2010). Eye movements across viewpoint changes in mulitple object tracking. *Visual Cognition*, *18*(9), 1368–1391.

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision*, *9*, 1–12.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219.

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

Mance, I., Becker, M. W., & Liu, T. (2012). Parallel consolidation of simple features into visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 429–438.

Milner, P. M. (1974). A model for visual shape recognition. *Psychological Review*, *81*, 521–535.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.

Nozawa, G., Reuter-Lorenz, P. A., & Hughes, H. C. (1994). Parallel and serial processes in the human oculomotor system: Bimodal integration and express saccades. *Biological Cybernetics*, *72*, 19–34.

Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanisms independent of higher-order cognition? An individual difference approach. *Visual Cognition*, *11*, 631–671.

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

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

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*, 4.

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

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

Saarinen, J., & Julesz, B. (1991). The speed of attentional shifts in the visual field. *Proceedings of the National Academy of Sciences of the United States of America*, *88*, 1812–1814.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*, 1–46.

Shiffrin, R. M., & Gardner, F. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology*, *93*, 72–83.

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, *4*, 421–457.

Sung, K. (2008). Serial and parallel attentive visual searches: Evidence from cumulative distribution functions of response times. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1372–1388.

Townsend, J. T. (1992). On the proper scales for reaction time. In S. L. H. Geissler & J. T. Townsend (Eds.), *Cognition, information processing and psychophysics: Basic issues*. Hillsdale, NJ, Erlbaum. *3*, 105–120.

Townsend, J. T., & Asby, F. G. (1983). Methods of modeling capacity in simple processing systems. In J. C. F. Restle (Ed.), *Cognitive theory* (pp. 200–239). Hillsdale, N.J., Erlbaum.

Townsend, J. T., & Fific, M. (2004). Parallel versus serial processing and individual differences in high-speed search in human memory. *Perception and Psychophysics*, *66*, 953–962.

Townsend, J. T., & Nozawa, G. (1995). On the spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, *39*, 321–359.

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*, 1003–1035.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–178.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*, 107–141.

Tripathy, S. P., Ogmen, H., & Narasimhan, S. (2011). Multiple-object tracking: A serial attentional process? In C. Mole, D. Smithies, & W. Wu (Eds.), *Attention: Philosophical and psychological essays* (pp. 117–144). Oxford, England: Oxford University Press.

von der Malsburg, C. (1981). The correlation theory of brain function. MPI Biophysical Chemistry, Internal Report 81-2. Reprinted in *Models of neural networks* II (1994). J. L. van Hemmen, E. Domany, and K. Schulten (Eds.) Berlin: Springer.

von der Malsburg, C., & Schneider, W. (1986). A neural cocktail-party processor. *Biological Cybernetics*, *54*, 29–40.

Vul, E., Frank, M. C., Tenebbaum, J. B., & Alvarez, G. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Advances in Neural Information Processing Systems*, *22*, 1–9.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics*, *33*, 113–120.

Wolfe, J. M. (2012) The binding problem lives on: Comment on Di Lollo. *Trends in Cognitive Sciences 16*, 307–308; author reply, 308–309.

Wolfe, J. M., Alvarez, G. A., & Horowitz, T. S. (2000). Attention is fast but volition is slow. *Nature*, *406*, 691.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, *37*, 25–43.

Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron*, *24*(11–17), 111–125.

Yamani, Y., McCarley, J. S., Mounts, J. R., & Kramer, A. F. (2012). Spatial interference between attended items engenders serial visual processing. *Attention, Perception & Psychophysics*, *75*, 229–243.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, *24*, 295–340.

Zelinsky, G. J., & Neider, M. B. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition*, *16*(5), 553–566.

Appendix

Test for ordering of survivor functions (Kolmogorov–Smirnov test)

Subject	Test	Statistic	*p* Value
PH	$S_{LL} < S_{LS}$	0.097	.049
	$S_{LS} < S_{SS}$	0.360	<.001
	$S_{LL} > S_{LS}$	0.028	.775
	$S_{LS} > S_{SS}$	0.009	.975
MW	$S_{LL} < S_{LS}$	0.184	<.001
	$S_{LS} < S_{SS}$	0.381	<.001
	$S_{LL} > S_{LS}$	0.004	.995
	$S_{LS} > S_{SS}$	<0.001	1.000
AB	$S_{LL} < S_{LS}$	0.118	.011
	$S_{LS} < S_{SS}$	0.392	<.001
	$S_{LL} > S_{LS}$	0.041	.582
	$S_{LS} > S_{SS}$	<0.001	1.000

These tests show that for all three observers the ordering of the survivor functions is consistent with effective selective influence.

Tests of SIC (Houpt–Townsend KS-SIC test)

Subject	Test	Statistic	*p* value
PH	SIC > 0	0.325	<.001
	SIC < 0	0.009	.990
	MIC	0.065	<.001
MW	SIC > 0	0.266	<.001
	SIC < 0	0.029	.905
	MIC	0.033	<.001
AB	SIC > 0	0.408	<.001
	SIC < 0	0.041	.816
	MIC	0.043	<.001

These statistics show that for all three observers the survivor interaction contrast (SIC) contained a significant positive component but not a significant negative component. The mean interaction contrast (MIC) was significantly different from zero. Combined, these two sets of statistics allow us to conclude that observers used a parallel processing strategy and a self-terminating stopping rule.

Capacity (OR version—Houpt–Townsend UCIP test)

Subject	Test Statistic	*p* Value
PH	−5.07	< .001
MW	−4.22	< .001
AB	−5.57	< .001

For all three observers, the capacity was significantly < 0, demonstrating that the system has limited capacity.