Cognition 146 (2016) 393–409

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Position tracking and identity tracking are separate systems: Evidence from eye movements

Lauri Oksama [a,*], Jukka Hyönä [b]

[a] National Defence University, Finland
[b] University of Turku, Finland

ARTICLE INFO

Article history:
Received 28 June 2013
Revised 2 October 2015
Accepted 19 October 2015

Keywords:
Multiple identity tracking
Multiple object tracking
Eye movements
Visual attention
Visual scanning

ABSTRACT

How do we track multiple moving objects in our visual environment? Some investigators argue that tracking is based on a parallel mechanism (e.g., Cavanagh & Alvarez, 2005; Pylyshyn, 1989), others argue that tracking contains a serial component (e.g. Holcombe & Chen, 2013; Oksama & Hyönä, 2008). In the present study, we put previous theories into a direct test by registering observers' eye movements when they tracked identical moving targets (the MOT task) or when they tracked distinct object identities (the MIT task). The eye movement technique is a useful tool to study whether overt focal attention is exploited during tracking. We found a qualitative difference between these tasks in terms of eye movements. When the participants tracked only position information (MOT), the observers had a clear preference for keeping their eyes fixed for a rather long time on the same screen position. In contrast, active eye behavior was observed when the observers tracked the identities of moving objects (MIT). The participants updated over four target identities with overt attention shifts. These data suggest that there are two separate systems involved in multiple object tracking. The position tracking system keeps track of the positions of the moving targets in parallel without the need of overt attention shifts in the form of eye movements. On the other hand, the identity tracking system maintains identity–location bindings in a serial fashion by utilizing overt attention shifts.

1. Introduction

Keeping track of multiple moving objects is a central part of our everyday life. For example, a mother may be tracking the whereabouts of her children on a crowded beach, or a car driver approaching a busy intersection is monitoring other vehicles also manoeuvring through the intersection. Moreover, professionals, such as air traffic controllers and fighter pilots, constantly deal with similar dynamic visual environments. However, the demands of different tracking tasks may vary quite notably from each other. Sometimes it is sufficient that we are simply aware of the members of the target set as a whole, for example, when a soccer player is attending to the whereabouts of the opponent team members. Other times it is required that we are aware of the whereabouts of individual members of the target set, for example when a soccer player wishes to pass the ball to his team's top scorer.

In the research literature on tracking of moving objects, the former task bears similarity to the multiple object tracking (MOT) task (Pylyshyn & Storm, 1988), where observers track a set of targets that are visually identical to each other (likened to the tracking of a flock of white sheep). Thus, only location information needs to be encoded, whereas object features are irrelevant to the task (see Fig. 1, top). On the other hand, the latter task is similar to the multiple identity tracking (MIT) task (Horowitz et al., 2007; Oksama & Hyönä, 2004), where distinct objects (likened to individual soccer players) are tracked and where observers need to constantly bind and update identity information with location information (see Fig. 1, bottom). Thus, the MOT task is a position tracking task whereas the MIT task is an identity tracking task by nature.

Several theoretical controversies have emerged concerning the mechanisms of position and identity tracking. Firstly, is tracking achieved by a serial or by a parallel process? Secondly, do position tracking and identity tracking share a common mechanism or are they based on independent mechanisms?

1.1. Is multiple object tracking serial or parallel in nature?

Some investigators argue that tracking is based on a parallel mechanism (Alvarez & Cavanagh, 2005; Cavanagh & Alvarez,

* Corresponding author at: National Defence University, P.O. Box 5, FI-04401 Järvenpää, Finland.
E-mail address: loksama@utu.fi (L. Oksama).

http://dx.doi.org/10.1016/j.cognition.2015.10.016
0010-0277/

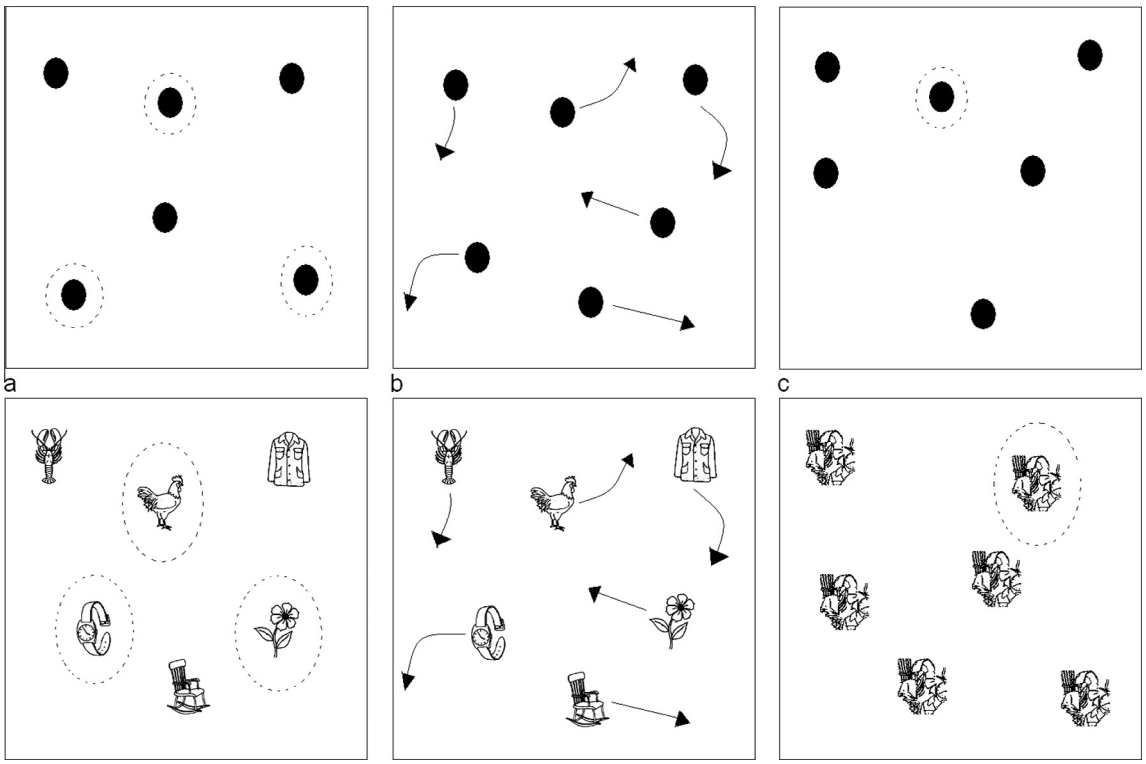

Fig. 1. A schematic depiction of the multiple object tracking task (MOT) with identical objects (top) and the multiple-identity tracking task (MIT, bottom) with distinct objects. Display a: six different objects are presented, and in this trial three of them are designated as targets by flashing a frame around them. Display b: all objects begin to move randomly about the screen. The participant's task is to track the location of the designated targets. Display c: when the motion stops, the participant is asked whether a flashed probe was among the target set flashed at the outset (MOT) or to report the identity of the masked probe (MIT). The target pictures of objects used in MIT were reprinted from Snodgrass and Vanderwart (1980), © 2007 Life Sciences Associates.

2005; Franconeri, Jonathan, & Scimeca, 2010; Howe, Cohen, Pinto, & Horowitz, 2010; Kazanovich & Borisyuk, 2006; Pylyshyn, 1989, 2001), others argue that tracking contains a significant serial component (d'Avossa, Shulman, Snyder, & Corbetta, 2006; Holcombe & Chen, 2013; Oksama & Hyönä, 2004, 2008; Tripathy, Ogmen, & Narasimhan, 2011). Parallel theories are typically based on data collected using the MOT paradigm (but see Howe & Ferguson, 2015), where observers track visually identical targets. According to the FINST theory (Pylyshyn & Storm, 1988), tracking is carried out in parallel for all targets within the capacity limit of about four items. Moreover, the tracking mechanism is assumed to operate pre-attentatively. According to the theory of Cavanagh and Alvarez (2005), tracking requires attention and is based on multiple attentional foci, between which limited attentional resources are allocated. Both versions of the parallel theory assume that serial switching of visual attention between target objects is not needed, either because attention is not needed or because object tracking is parallel in nature. Finally, Alvarez and Franconeri (2007) have proposed a model in which tracking is achieved by a flexibly allocated mental resource; however, they refrain from taking stand whether this resource is serial or parallel in nature.

Oksama and Hyönä (2008) have proposed a serial model especially designed for multiple identity tracking. According to their MOMIT model, observers use only one attentional focus, which needs to be shifted serially from one target to the next. When visual attention is focused on a target, its identity–location binding is created (or updated). In other words, binding identity with location is carried out individually for each target. As other, non-attended targets keep moving, it means that their bindings will be outdated and will not be updated until they are focally attended one at a time. It is further assumed that locations for the tracked targets are temporarily stored in visuo-spatial short-term memory (VSTM). This indexed location information (bound to identities) is then utilized by a mechanism that programs shifts of visual attention between targets. As targets move continuously, location information for all other than the focally attended targets are outdated. The magnitude of this location error is a key factor in predicting tracking accuracy as a function of object speed and target set-size. The size of the location error increases with an increase in target speed and set-size, which results in less efficient switching of attention between targets. Furthermore, it is assumed that serial shifting of attention is controlled partly with the help of this indexed location information stored in VSTM and partly with the help of peripheral vision. According to MOMIT, peripheral vision provides non-indexed (not bound to identities) location information about all moving objects in parallel.

Engineering models describe a generic model for visual sampling in dynamic situations. They are kin to MOMIT, as they are based on serial switching of focal attention. Seminal engineering models of supervisory control provide precise predictions about how often a dynamic display (e.g., flight instruments in a cockpit) should be sampled with focal attention and the eyes (e.g., Carbonell, 1966; Moray, 1984, 1986; Senders, 1964, 1983; Sheridan, 1970; see also Horrey, Wickens, & Consalus, 2006). According to Senders' (1964) model, in order to effectively monitor a dynamic display, it is necessary to sample the display at a rate twice its information bandwidth (bandwidth measured in events/s = Hz). For example, when relevant information in an information channel occurs at a rate of 1 Hz, the optimal observer should visually sample this channel at a rate of 2 Hz. Later Moray (1986) argued that the optimal sampling rate would be equal to the bandwidth. Sampling of dynamic display at this rate is necessary, because the more rapidly the dynamic signal varies, the more quickly it will become impossible to predict its current value on