1 Objects That Move

Attention is "the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought." At least that is how William James described it (James, 1890). James' description seems to imply that attention has a limited capacity of just one object or train of thought. James was joined at Harvard in 1892 by Hugo Münsterberg, who used moving stimuli to study attention. Münsterberg published a book in 1916, *The Photoplay: A Psychological Study*, which described his theory of the "moving pictures" (the cinema) and included a chapter on attention.

Münsterberg's book is insightful, but he did not address how attention operates in the presence of multiple moving stimuli. Much later, after World War II, the study of attention grew rapidly, and tachistoscopes became the

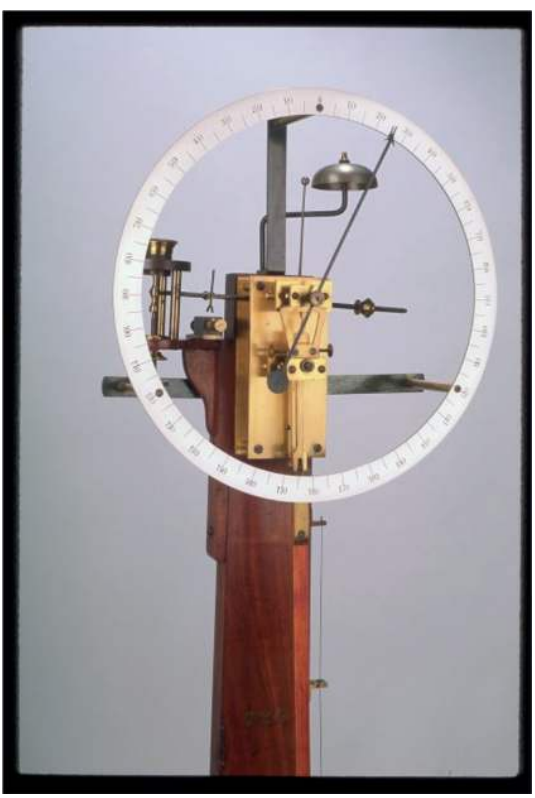
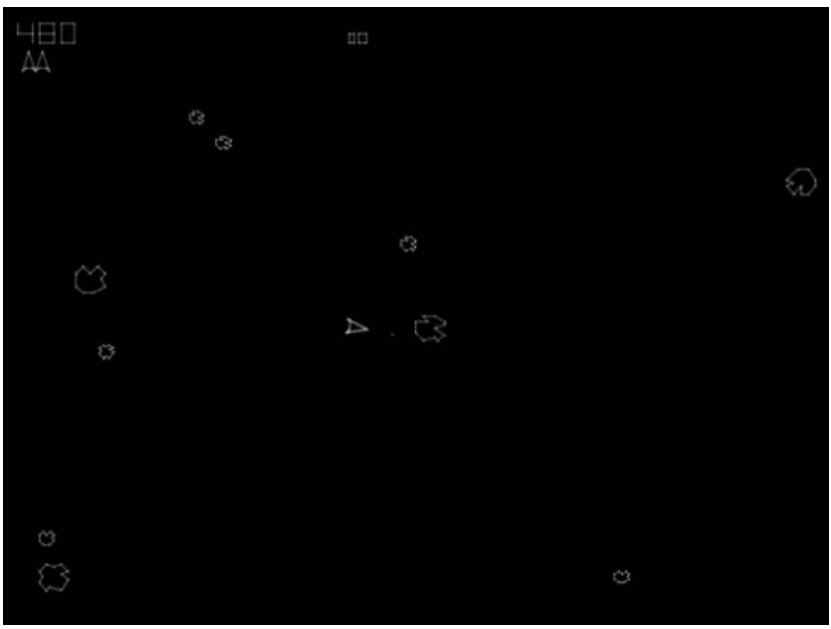


Figure 1 This "complication apparatus" from the Harvard laboratory of Hugo Münsterberg was used to measure the effect of attention to one stimulus on responses to another. A subject who focused on one of the numbers on the large dial was found to have a delayed reaction to the sound of the bell, and vice versa.



Video 1 Asteroids was released by Atari in 1979. Note: an animated version of the figure is available in the online resources (www.cambridge.org/Holcombe_supplementary)

standard laboratory presentation apparatus. These devices were limited in that they could not present motion: they were designed to present static stimuli very briefly. The dominance of stationary stimuli in the study of attention continued through the 1980s, even as the study of motion grew in a separate community of perception researchers.

The first popular home game system, the Atari, introduced the game Space Invaders in 1980 to millions of homes, including those of some of my childhood friends. Asteroids was ported to the Atari soon after, and it became one of my favorites.

When one plays Space Invaders or Asteroids (Video 1), multiple objects frequently move in the direction of one's avatar. Avoiding a collision seems to require monitoring more than one of these objects at a time. The ability of humans to do this was formally studied first by the Canadian psychologist and engineer Zenon Pylyshyn.

In the 1970s, Zenon Pylyshyn had been pondering the possibility of a primitive visual mechanism capable of "indexing and tracking features or feature-clusters" (he mentions this in Pylyshyn and Storm [1988]; I haven't been able to get copies of the 1970s reports that he refers to) as they moved. By 1988, Zenon Pylyshyn and Ron Storm formulated a way to empirically study

Video 2 A demonstration of the MOT task, created by Jiri Lukavsky. Note: an animated version of the figure is available in the online resources (www.cambridge.org/Holcombe_supplementary)

his hypothesized primitive visual mechanism, and they did a series of experiments on humans' ability to keep track of moving objects (Pylyshyn and Storm, 1988). On their Apple II+ computer, they created a display with ten identical objects moving on random trajectories, connected to a telegraph key with a timer to record response times. Pylyshyn and Storm also pioneered the use of an eyetracker to enforce fixation – in their experiments, movement of the eyes away from fixation terminated a trial. Thus they were able to investigate the ability to covertly (without eye movements) keep track of moving objects.

In a task that Pylyshyn and Storm dubbed multiple object tracking (MOT), up to five of ten displayed moving objects were designated as targets by flashing at the beginning of the trial. The targets then became identical to the remaining moving objects, the distractors, and moved about randomly. While viewing the display, people report having the experience of being aware, seemingly continually, of which objects are the targets and how they are moving about. In the movie embedded in Video 2, one is first asked to track a single target to become familiar with the task, and then subsequently four targets are indicated.

In addition to their demonstration that people could do the basic task, which in itself is quite important, Pylyshyn and Storm (1988) also showed that people are limited in *how many* targets they can faithfully track. In their experiments, Pylyshyn and Storm (1988) periodically flashed one of the moving objects, and if that object was a target, the participant was to press the telegraph key.

On trials with more targets, errors were much more common: while only 2% of target flashes were missed when only one of the ten objects was a target, 14% of target flashes were missed when five of the objects were targets.

The notion of keeping track of moving objects is familiar from certain situations in everyday life. If you've ever been responsible for more than one child while at a beach or a park, you know the feeling of continuously monitoring the locations of multiple moving objects. If you've ever played a team sport, you may recall the feeling of monitoring the positions of multiple opponents at the same time, perhaps the player with the ball and also a player they might pass the ball to. If you've ever wanted to speak to someone at a conference, you may know the feeling of monitoring the position and posture of that person relative to others they are chatting with, in order to best time your approach.

1.1 What's to Come

Despite advances in technology, the study of visual cognition continues to be dominated by experiments with stimuli that don't move. As we'll see in Section 10, putting objects in motion reveals that updating of their representations is not as effective as one might expect from studies with static stimuli. This suggests that with static objects, one can bring to bear additional processes, perhaps cognitive processes (Section 6), that motion helps to dissociate from lower-level tracking processes. It is these sorts of unique insights from MOT experiments that I have chosen to emphasize in this Element, together with the findings that I believe most constrain theories of how mental tracking processes work. I will argue that the following are the five most important findings in the MOT literature:

1. The number of moving objects humans can track is limited, but not to a particular number such as four or five (Section 3).
2. The number of targets has little effect on spatial interference, whereas it greatly increases temporal interference (Section 5).
3. Predictability of movement paths benefits tracking only for one or two targets, not for more (Section 6).
4. Tracking capacity is hemifield-specific: capacity nearly doubles when targets are presented in different hemifields (Section 9).
5. When tracking multiple targets, people often don't know which target is which, and updating of nonlocation features is poor (Section 10).

The organization of this Element was influenced by my desire to dispel common misconceptions about results in the literature, and to lay out the concepts needed to understand the implications of the empirical findings. In Section 13

I describe some broad lessons, including how best to study tracking in the future. We will start with the concept of limited capacity and bottlenecks in the brain.

2 Bottlenecks, Resources, and Capacity

Quickly, what is fourteen times thirteen? Calculating that in your head takes a while, at least a few seconds. And if I set you two such problems rather than just one, I'm confident that you would do those problems one at a time. Our minds seem to be completely incapable of doing two such problems simultaneously (Oberauer, 2002; Zylberberg et al., 2010). This limitation is remarkable given that each of our brains contains more than eighty billion neurons. The problem is not a lack of neurons, really, but how they are arranged – our mental architecture.

Multiplying and dividing two-digit numbers may not be something you attempt to do every day. You might think, then, that if you were doing lots of such problems each day, you could eventually do more than one at a time. This is probably wrong – consider that a task we do have daily practice with is reading. Despite years of reading dozens if not hundreds of words a day, the evidence suggests that humans can read at most only a few words at a time, and some research further indicates that we can really only read *one* word at a time (Reichle et al., 2009; White et al., 2018). At least some of the bottlenecks of human information processing, then, appear to be a fixed property of our processing architecture.

To flesh out what I mean by "bottleneck" here, consider a standard soft drink bottle. If you invert a full bottle, most of the liquid volume will be pressing down on the neck. The narrowness of the neck restricts the rate at which the liquid can exit the bottle. Similarly, a large volume of signals from sensory cortex ascending the cortical hierarchy press up against higher areas that are more limited in capacity.

The parallel processing happening in visual cortices, such as the multiple neurons dedicated to each patch of the visual field, gets a number of tasks done, so that higher stages don't have to do those tasks. These tasks appear to include the encoding of motion direction, color, and orientation throughout the visual field. Local and regional differencing operations happen for those features, resulting in salience, whereby odd features become conspicuous in our visual awareness. In Figure 2, for example, you should be able to find the blue objects very quickly.

For other judgments, higher, post-bottleneck brain areas that are very limited in capacity are critical. The visual word form area in the occipitotemporal sulcus of the left hemisphere, which seems to be needed to recognize words, is

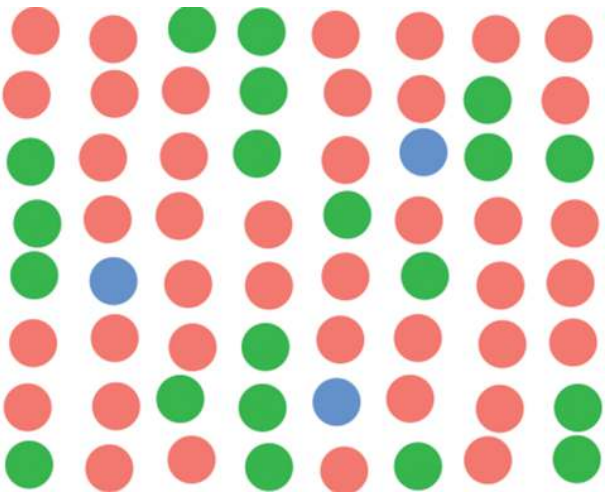


Figure 2 Thanks to featural attention (to color in this case), you should be able to find the blue circles very quickly

one example (White et al., 2019). Being limited in processing capacity to just one stimulus, the word recognition will not happen in a crowded scene until something selectively directs the visual signals from a word to the visual word form area. We often use the term *selective attention* to refer to this "something" that directs particular visual signals to the bottlenecks of limited-capacity processes. If there were no bottlenecks, there would be no need for selection for cognition (selection would be required when an action needed to be chosen).

So far the picture I have painted has been one of a torrent of visual signals impinging on a narrow bottleneck of signals that continue onward. But cortical processing is rarely a one-way street, and the way visual attention works is no exception. Visual attention seems to work partly by biasing processing within visual cortices, rather than leaving that unchanged and blocking all but a few signals at a later bottleneck stage. Thus, processing capacity may be restricted by limitations on control signals from high-level (possibly parietal) cortex that restrict processing capacity, as well as the more familiar idea of a structural bottleneck where ascending visual signals reach a lower-capacity neural mechanism.

To the extent control signals are a limitation, a resource metaphor can be apt. The control of selection may reflect a finite pool of neural resources in parietal cortex that bias which visual signals are cognitively processed. Thus I will sometimes use the term "limited resource" when referring to how we are restricted in how many visual representations are processed.

The word "resource" carries the appropriate connotation that people can choose how to apply their finite processing capacity; ordinarily a resource is

something that can be used in different ways. For example, the term suggests that one might use three-quarters of one's processing capacity for one target while using the other quarter for a second target. And indeed, there is evidence that people can favor one target over another when tracking both (Chen et al., 2013; Crowe et al., 2019).

While word recognition seems to be able to process only one stimulus at a time, other visual judgments may be limited in capacity relative to massively parallel sensory processing, but have a capacity greater than one. Object tracking seems to be one such ability. People appear able to track more than one target at the same time, although researchers haven't fully ruled out the possibility that tracking multiple objects happens via a one-by-one process that rapidly switches among the tracked objects.

The existence of processes with a capacity of just one object (I will introduce the term "System B" for this in Section 6) is a good reason to have a process that can keep track of the location of important objects in a scene. We are then always ready to rapidly shunt a subset of them to higher-level processing, rather than having to search for it.

3 The Biggest Myth of Object Tracking

What I consider to be the biggest myth about object tracking involves three misconceptions:

1. There is a fixed capacity limit of about four or five objects that can be tracked, after which performance falls rapidly.
2. A softer version of the above claim: that performance falls to a particular level once the number of targets is increased to four or five objects.
3. Different tasks show the same limit.

These three claims are widespread in the scholarly literature. A set of researchers writing about the "object tracking system" in 2010, for example, stated: "One of the defining properties of this system is that it is limited in capacity to three to four individuals at a time" (Piazza, 2010). Similarly, Fougnie and Marois (2006) wrote that "People's ability to attentively track a number of randomly moving objects among like distractors is limited to four or five items." This idea is sometimes perpetuated with more ambiguous statements such as "participants can track about four objects simultaneously" (Van der Burg et al., 2019).

Misconception #1 in my list, including the idea of a sharp fall in performance after a limit, is one aspect of the statements of the previous paragraph. This is fully explicit in one set of researchers' 2010 take on the literature,

when they wrote that "the main finding" of the object tracking literature is that "observers can accurately track approximately four objects and that once this limit is exceeded, accuracy declines precipitously" (Doran and Hoffman, 2010). Vaguer statements in other papers, such as "researchers have consistently found that approximately 4 objects can be tracked" (Alvarez and Franconeri, 2007) and "people typically can track four or five items" Chesney and Haladjian (2011), also bolster misconception #1 in the minds of readers.

To examine the evidence behind the claims of each of the quotations of the two preceding paragraphs, I have checked the evidence provided, and the papers cited, as well as the papers those cited papers cite. No paper contains any evidence supporting the claim that performance decreases very rapidly once the number of targets is increased above some value. Instead, a gradual decrease in performance is seen as the number of targets is increased, with no discontinuity, not even a conspicuous inflection point. For example, Oksama and Hyönä (2004), which is sometimes cited in this context, assessed performance with up to six targets. After a five-second phase of random motion of the multiple moving objects, one object was flashed repeatedly and participants hit a key to indicate whether they thought it was one of the targets. The number of trials that participants got wrong increased steadily with target number, from 3% incorrect with two targets to 16% incorrect with six targets.

Although Pylyshyn and Storm (1988) is the paper most frequently cited when a limit of four objects is claimed, even they found a quite gradual decrease in performance (their Figure 1) as the number of targets was increased, from one to five (five targets was the most that they tested). And nowhere in their paper did Pylyshyn and Storm (1988) state that there is a value beyond which performance rapidly declines. Six years later, however, Pylyshyn et al. (1994) did write that it is "possible to track about four randomly moving objects." By 2007, when he published his book *Things and Places: How the Mind Connects with the World*, Pylyshyn wrote sentences like "And as long as there are not more than 4 or 5 of these individuals the visual system can treat them as though it had a concept of 'individual object'" (Pylyshyn, 2007). I suspect that this sort of slide toward seeming to back a hard limit is caused in part by the desire for a simple story. It may also stem from an unconscious oversimplification of one's own data, and/or Pylyshyn's commitment to his theory that tracking is limited by a set of discrete mental pointers.

I have so far addressed only one aspect of the claim (misconception #1): that there is a limit after which performance decreases rapidly. Another aspect of misconception #1 is that the limit is consistently found to be four or five. This isn't viable if there is no limit after which performance decreases rapidly,