Getting beyond the serial/parallel debate in visual search: a hybrid approach

Cathleen M. Moore and Jeremy M. Wolfe

Abstract
The question of whether visual search involves at least one item-by-item serial processing stage or whether instead it is an entirely parallel process has been debated for decades. Recently, estimates of 'attentional dwell-time', which is the time required to reallocate attention from one item to another, have been brought to bear on this question. Here, we review this and other classes of evidence that favor serial or parallel models of visual search, and conclude that hybrid models that are neither strictly serial nor strictly parallel are better candidates for describing human visual search. We end the chapter with a sketch of one such model, and some of its possibilities.

If you look for a 2 among the 5s in Fig. 9.1, it will take you longer to find the 2 in 1B than in 1A and even longer to confirm that there is no 2 present in 1C. How should we understand the dependence of response time (RT) in this sort of task on the number of items (set size)? The answer to this question has been phrased in essentially dichotomous terms for a generation. It could be that each item is processed one after the other in series (cf. e.g. Treisman and Gelade 1980; Wolfe et al. 1989; Wolfe 1994). It could be that information is accumulated from multiple items at the same time in parallel (Grossberg et al. 1994; Humphreys and Muller 1993; Palmer and McLean 1995). Which is it? Is attention deployed to one item at a time in visual search or is it distributed across many or all items?

These sound like dramatically different alternatives. Why then has it proved so difficult for proponents of serial models or proponents of parallel models to gain the upper hand in this debate? One explanation has been that, for all their apparent differences, serial and parallel models can be made to predict similar experimental outcomes (e.g. Townsend 1971, 1976, 1990). We will argue for a somewhat different position. There is a class of plausible models within which the distinction between serial and parallel deployment of attention is like the distinction between wave and particle theories of light. Measured one way, light behaves like a wave. Measured another way, it behaves like a particle. So far as we can tell, the reality is that light is both wave and particle. In a similar manner, we will argue that processing in visual search is, in a non-trivial sense, both serial and parallel. The analogy to light is intended as a loose one, in that light, so far as we can tell, is never particle and wave at the same time, and the models of search to which we refer are simultaneously serial and parallel.

This is not simply a claim that vision contains both serial and parallel components. It is relatively trivial to note that the earliest steps in visual processing are performed in parallel (for example, the registration of the image on the retina) and that many of the possible output steps (for example, eye movements, response execution) are necessarily serial. We are focused here on the role of selective attention, which is that process by which items are selected for the purpose of classifying them as target or non-target during visual search, which we will argue can be both serial and parallel in nature at the same time.

Before describing this alternative, however, we discuss some of the research that has given rise to this apparent dichotomy-in-need-of-resolution. This is not intended as an exhaustive survey of the literature, but rather as an illustration of some of the main points of unresolved contact between the two camps. After describing some of the theoretical reasons for proposing a serial process in visual search, we briefly review the empirical basis for inferring serial mechanisms from RT data in visual search experiments. Next, we discuss some of the empirical evidence that argues against serial processing in search, favouring instead entirely parallel models. Finally, we describe a way of looking at the data that seeks to explain how visual search can appear to be serial and parallel at the same time.

Why propose serial selection in visual search?
All else being equal, parallel processing would seem to be preferable to serial processing. Why do one thing at a time if you can do many? Recent ideological energy behind serial models of visual search comes from the conviction that, beyond a certain point, it becomes computationally impossible to perform vision in parallel (cf. e.g. Tsotsos 1990). Earlier, and more general, impetus for a serial stage to processing comes from the work of Broadbent (1958) and others, who began conceptualizing human information-processing in terms of formal information theory. Under that scheme, the human system was thought of as a limited-capacity information channel, which required that processing be reduced from parallel to serial (or nearly serial) at some point within the system.

Within the domain of visual processing, there are at least two specific aspects of search that have been hypothesized as requiring serial processing. First, Treisman and others have argued that to 'bind' the features of an object correctly together, the selection of individual objects is necessary (e.g. Treisman and Gelade 1980). Second, object recognition during search has a necessarily serial flavour to it (cf. e.g. Duncan 1990; Wolfe et al. 1989).

Figure 9.2 provides a standard illustration of the apparent need for serial selection when searching for a conjunction target. While it is easy enough to detect the black square in panels A and B, it is more difficult to do so in panel C. This, according to Treisman, is because the black square is defined with regard to its distractors on the basis of a single feature in each of the two upper displays, but only by the conjunction of features in the lower display. Therefore,
According to the theory, attention must be separately allocated to each individual feature to detect the black square as a black square.

Figure 9.3 illustrates a variation on the basic conjunction theme that suggests that the proper conjoining of features is preceded by a loose, unstructured, grouping of features on the basis of rough object representations. If you look for black horizontal lines, you will find it easier to find one in the right display than in the left display (Wolfe and Bennett 1997). This finding is consistent with the idea that preattentive processes (i.e., processes prior to the conjunction of features) parse the world up into candidate objects, which are essentially loose collections (think ‘lists’) of basic features. On the right, each object has only one orientation and only one colour (ignoring the central grey patch that is present in all items). So the preattentive object representations might consist of something like Object, = {black, vertical} or Object, = {white, horizontal}. In contrast on the left, each object is both black and white and both vertical and horizontal. So all the preattentive object representations would look something like Object, = {black, white, vertical, horizontal}. As such, all the objects on the left are preattentively identical, whereas all of the objects on the right are not.

The binding argument for seriality accounts for effects like those in Figs. 9.2 and 9.3 in so far as it holds that, until attention selects an item, it is impossible to determine what colour is bound to what orientation or shape. If attention is not allocated to individual items, binding failures or confusions can result in the illusory conjunction of features into items that are not in the display, but whose features are present across a number of different items in the display (Ivry and Prinzmetal 1990; Prinzmetal 1995; Treisman and Schmidt 1982). An example of this would be the illusory perception of a grey circle in Fig. 9.2C.

At a theoretical level, object recognition, like the conjunction of features, seems like a plausible candidate for serial processing. To recognize an object, it is necessary to form some connection between the visual representation of that object and its representation in memory. As will be discussed below, it may be that only one such link can be maintained at any instant in time (Wolfe et al. 2001; see also Carrier and Pashler 1996). To the extent that visual search requires object recognition, this would be another theoretical justification for a mandatory serial stage in the processing of items in visual search.

Theory aside, what are the empirical foundations of serial models? In the next section, we consider some of the evidence that has been taken as supportive of the idea that visual search involves a serial component.

**Empirical foundations of serial models**

One root of the dichotomy between parallel and serial processing during visual search can be found in the early work of Neisser (1967) and Treisman (Treisman and Gelade 1980), some of the theoretical consequences of which were described in the preceding section (see also: Egeth et al. 1972; Hoffman 1979; and van der Heijden 1975). Following the ideas of Broadbent (1958), Neisser introduced the distinction between a parallel ‘preattentive’ stage of processing and other ensuing bottleneck processes that can handle only one stimulus, or a very few stimuli, at a time. Attention, as we are using the term, refers to the process that selects items to pass through those bottlenecks.

Within the context of visual search, Treisman developed these ideas into Feature Integration Theory (FIT). In its early form (Treisman and Gelade 1980), FIT held that a limited set of visual search tasks could be performed in parallel. These were tasks in which the target was defined by a single basic stimulus feature, such as colour, size, or orientation (e.g., Figs. 9.2A and 9.2B). Other search tasks, notably those requiring the binding of simple features into more complex stimuli, required the serial deployment of attention from item to item until the target was found (e.g., Fig. 9.2C). FIT has since been modified to include the possibility that feature information can be used to guide the deployment of attention (Wolfe et al. 1989; Wolfe 1994; Treisman and Sato 1990). Thus for example, search for a red number among black and red letters might be serial; but it would be serial through the set of red letters. Attention would be guided toward the red items, and not ‘wasted’ on black items (Egeth et al. 1984; Moore and Egeth 1998; Zohary and Hochstein 1989). While models of this sort admit to an early parallel stage of processing, they also have a commitment to a later, item-by-item, serial deployment of attention. It is through this item-by-item engagement of processing that stimuli are classified as target or non-target during the search process.
How can such models be tested? The original Feature Integration Theory proposed that some searches were parallel and others were serial, so one could seek evidence concerning the existence of two different modes of search. Toward this end, the slope of the function relating RT to the number of items in the display (set size) was offered as a metric for categorizing specific search tasks as serial or parallel (cf. e.g. Treisman and Gelade 1980; Sternberg 1975). The logic is that if search is parallel (and unlimited in processing capacity), then adding items to the display should not affect the amount of time it takes to find the target. Therefore, the slopes relating RT to set size should be near zero. In contrast, if search is serial, then each item that is checked should add a more-or-less constant amount of time to the total time required to find the target. So in this case, RT should increase linearly with set size. A serial self-terminating model—one in which search ceases when the target is found—makes the additional, more specific, prediction that target-absent slopes will be twice as large as target-present slopes. This follows because if the target is present, then, on average, it will be found after checking half the items in the display. In contrast, in order to report that the target is absent, all the items in the display must be checked.

Consistent with the basic tenets of FIT, both patterns of results have been found (see Wolfe 1994; 1998 for reviews). That is, some searches have produced near zero RT by set size functions, and others have produced linearly increasing functions with something close to a 2:1 slope ratio between target-absent and target-present trials. These results have been taken as evidence that visual search sometimes includes an item-by-item serial component.

The search-slope logic is appealing at many levels, and has certainly been invoked in many studies. There are, however, serious problems with it as an arbiter of the serial/parallel debate (see for example Palmer 1994; Townsend 1971, 1990; Wolfe 1998). First of all, while a serial self-terminating model predicts that RT will increase as a linear function of set size, and that there will be a 2:1 slope ratio between target-absent and target-present trials, this logic does not work in reverse. That is, finding this pattern of results does not mean that search was necessarily serial and self-terminating. The pattern may have been caused by some other type of process. Townsend and others, for example, have shown that models that include no serial component, but do include a limited-capacity parallel component, can produce the same pattern of results (e.g. Townsend 1971, 1990; Ward and McClelland 1989). Moreover, on the basis of accuracy versions of this task, Palmer and others have shown that even models including only unlimited-capacity parallel processes may be able to produce the signature pattern of results (Eckstein 1998; Palmer and McClean 1995). One might appeal to parsimony, and argue that the assumptions that are necessary for the serial self-terminating model are fewer and simpler than those required for the alternative parallel models. However, even if true, this argument would not eliminate the logical problems. Moreover, despite Occam's adage, nature does not necessarily submit to the same priorities as those to which theoreticians aspire: more parsimonious does not necessarily imply more accurate.

In addition to the logical problems with inferring search mechanism from slope data, the data themselves raise nagging doubts about this enterprise. If search slopes, by themselves, were diagnostic of search mechanism, and searches could be divided into 'parallel' searches and 'serial' searches, then the distribution of slopes across the set of all search tasks might be expected to be bimodal. That is, there might be some principled slope value that could be established such that searches producing slopes above that level could be reliably labelled as 'serial', while those with slopes below this magic value could be labelled as 'parallel'. In fact, with something like this in mind, ten ms/item has become a conventional value for this magic slope. The data, however, reveal no such bimodality. Wolfe (1998) analysed the slopes of hundreds of visual-search experiments representing about 1,000,000 individual search trials. The data set included several different types of standard search tasks. Yet the distributions of slopes did not even hint at bimodality (see Fig. 9.4).

Note that the failure to find a bimodal distribution does not, by itself, falsify the hypothesis that there are 'parallel' and 'serial' searches. While a bimodal distribution would favour the conclusion that there were two different modes of search, the failure to find bimodal distributions does not support the conclusion that there is only one mode of search, though it is consistent with that conclusion. This follows because the distributions shown in Fig. 9.4 could represent a mixture of two underlying 'serial' and 'parallel' distributions that obscures the underlying distributions. While there are analytic tools for extracting component distributions from mixed distributions (Meyer et al. 1986; Yantis et al. 1992), unless the underlying distributions are substantially far apart, it is fairly difficult to do so in practice (Yantis et al. 1992). In any case, it is clear that the empirically unimodal distributions seen in Fig. 9.4 indicate that the mechanism of search cannot be trivially inferred (for example, by setting a criterion of 10 ms/item) from the slope of the RT x set size function.

In addition to the question of whether there are two qualitatively different modes of search, the data also raise doubts about the more specific conclusion that search is sometimes serial and self-terminating. First, in Wolfe's (1998) overview analysis of the many visual-search experiments, he also looked at the distribution of target-absent to target-present slope ratios. He found that, over a wide range of target-present slopes, the ratio was reliably greater than the 2.0 that the serial self-terminating models predict. Moreover, searches yielding increasing search slopes have produced target-present and target-absent slopes that are nearly identical (cf. e.g. Atkinson et al. 1969; Townsend and Roos 1973). Pashler (1987) has shown that while the 2:1 slope ratio between target-present trials can be observed across large ranges of display size, it can be shown to be closer to 1:1 when the range is small (i.e., from about 10 to 8 items). On the basis of these data, Pashler suggested the possibility of something like a serial-by-clumps model, wherein clumps of items are selected and processed in parallel, but the clumps are selected in series (see also Grossberg et al. 1994). Finally, Ward and McClelland (1989) noted that the variances that are typically observed across the target-present and target-absent conditions of visual-search experiments are often inconsistent with a simple serial self-terminating model.

Again, however, keeping with the theme that slope data from visual search tasks make for a poor arbiter of the serial versus parallel question, the failure to find 2:1 slope ratios is less conclusive than one might think or hope. RTs on target-absent trials are RTs of unsuccessful
searches for a target. If these were all simple exhaustive searches through the display, modeling would be uncomplicated. There is no reason to think, however, that these are all simple exhaustive searches. First, subjects make errors. In some cases, these errors seem to reflect an early abandonment of unpromising searches, which complicates matters (Chan and Wolfe 1996). Much worse, the simple 2:1 story relies on the assumption that rejected distractors are marked so that they are not revisited by attention during the search. This is a vexed topic (Klein 1988; Klein and MacInnes, in press; Klein and Taylor 1994; Wolfe and Polkorny 1990); but the very real possibility exists that any such marking is either imperfect (Arani et al. 1984) or nonexistent (Horowitz and Wolfe 1998). With all these complications, serial models do not make unambiguous predictions about slope ratios.

In summary, while there may be theoretical reasons to suppose that there is a serial stage during visual search and while RT search data can be modeled as a parallel stage feeding a serial stage, the RT data do not unambiguously support the existence of both stages. Other related evidence has been mustered in defense of a serial stage of processing (e.g. Kwak et al. 1991), but these have not been adequate to provide a resolution to the debate.

Empirical foundations for parallel models

Parallel models of visual search, in contrast to serial models, hold that attention can be distributed across multiple objects and that information about the identity of those objects can be developed simultaneously. There are many variants of parallel models (e.g. Grossberg et al. 1994; Bundesen 1990; Kinchla 1974; Ward and McClelland 1989). While the details of these models vary, they all reject the notion of a mandatory serial selection of individual items at any level of processing. A generic parallel model of the difficult search illustrated in Fig. 9.1 would have stimulus identity information accumulating for all objects at the same time. When an item crossed some threshold for identification as a 2, a target-present response would be generated. When all (or almost all) items had crossed a threshold for identification as 5’s a target-absent response would be generated. This then, in brief oversimplification, is the crux of the debate. Remaining with the 2 vs. 5 example, a parallel model would hold that it is possible to accumulate information concerning identification for multiple characters at one time, whereas a strict serial model would insist that identification of a 2 or a 5 requires attentional selection of that character and only that character for at least a moment of time. What evidence is there to support a parallel model over a serial model?

The unimodal distribution of search slopes that is shown in Fig. 9.4 raises the possibility of a single-factor model of search efficiency. Perhaps a limited-capacity parallel process mediates all visual searches. Different combinations of targets and distractors impose different demands on this process. Low-demand conditions produce efficient search with shallow slopes. High-demand conditions produce inefficient search with steep slopes, while intermediate conditions produce intermediate results (e.g. Duncan and Humphreys 1989; Humphreys and Muller 1993). A model of this sort would eliminate the need for the item-by-item serial deployment of attention in visual search. Consistent with such a model, several lines of investigation, which we describe in detail below, raise the possibility that stimuli can at least sometimes be identified in parallel. More than being consistent with a parallel limited-capacity model, definitive evidence on this point would constitute at least indirect evidence against serial models. This follows because if a process as late as stimulus identification can occur in parallel, then it seems unlikely that serial search should ever be required, except in cases that require individual eye fixations of each stimulus.

Using what we will refer to as the simultaneous/sequential method, Shiffrin and Gardner (1972; see also Eriksen and Spencer 1969) provided evidence suggesting that stimulus identification might occur in parallel. In Experiment III, for example, they used a visual search task in which the target was a T or an F, and the distractors were F/F hybrid characters. The task was to report whether the target was a T or an F. Display size was held constant at four items. What was varied in this study was the timing of information presentation within a trial. In the simultaneous condition, all the stimuli were presented at once. In the sequential condition, half the stimuli were presented at one time and then, after a 500 ms delay, the other half of the stimuli were presented. The total amount of time that a given piece of information was present was the same across the two conditions. The logic was that if the processing of one item interfered with the simultaneous processing of other items (for example, if processing had to be serial), and if attention could be switched within the time period between frames in the sequential condition (500 ms in this case), then there should be an advantage for the sequential condition over the simultaneous condition. Instead, Shiffrin and Gardner’s data were consistent with the idea that multiple stimuli can be processed simultaneously without interfering with each other, or that attention could not be moved in time; they found that target identification was no more accurate in the sequential condition than in the simultaneous condition.

Pashler and Badgio (1987) extended the work of Shiffrin and Gardner (1972) by using a task that was intended to ensure that all of the stimuli in the display had to be identified. In this task—known as the highest-digit task—displays of digits were presented, and subjects reported the identity of the highest digit in the display. Unless the highest digit in the display was also the highest digit in the whole set, all the digits in the display had to be identified in order to answer correctly. (Those trials on which the highest digit in the display was the highest in the set were analysed separately.) With regard to timing, Pashler and Badgio’s experiments were more like Experiments I and II of Shiffrin and Gardner’s study than Experiment III, in that stimulus presentations were separated in the sequential condition by only 50 ms (the stimulus duration). The results, however, were very similar. Specifically, across several different experiments, Pashler and Badgio found little or no advantage for the sequential condition over the simultaneous condition. These results suggest one of two things: processing all the way through stimulus identification occurred in parallel, unaffected by the processing of other stimuli, or attention could not be shifted from one frame to the next in the time available in the sequential condition.

For proponents of parallel models, the results from experiments using the simultaneous/sequential method suggest that processing all the way through stimulus identification can occur without an item-by-item serial engagement of individual stimuli, and therefore, that there is no need to propose a serial mechanism in visual search. As is so often the case in these serial/parallel debates, however, matters are not clear-cut. First, the studies that found little or no advantage for the sequential condition over the simultaneous condition used set sizes of only four items. When set size was increased to sizes that were closer to those used in standard visual-search tasks, a sequential advantage was observed (Fisher 1984). Second, sequential advantages were also observed when the target-distractor discrimination was more complex (Duncan 1987; Klein and Lane 1986; Shiu and Pashler, described in Pashler 1998). Finally, subjects were not performing all that well in either the sequential or the simultaneous versions of this task. For instance, in various versions of the Pashler and Badgio (1987) study, accuracy
was in the range of 50–67 per cent. This suggests that subjects could acquire information that was adequate to identify between 2 or 3 of the 4 items. In the sequential condition, the failure to process more items might reflect a failure to redepoly attention successfully from location to location at the required speed (see Shiffrin and Gardner 1972 for a similar argument). Even if attention can deploy quite rapidly (say, every 50 ms), there is evidence that it cannot be forced to specific locations at those high rates by exogenous prompts (e.g. Reeves and Sperling 1986; Weichselgarter and Sperling 1987; Wolfe and Alvarez 1999). We will return to this problem of the speed of attentional deployment later.

Another source of evidence cited as support for the idea that stimuli can be identified in parallel during visual search is in the spirit of the original search-slope logic. The idea is to introduce various stimulus manipulations and observe their effects on the pattern of RT. If a manipulation increases the time required for a serial step in processing, then each item in the display will take longer to process and the slope of the RT x set size function will increase. If a manipulation reduces the speed of a parallel stage, then the overall RT should increase, but it should not necessarily do so more for displays with many items than for displays with fewer items. Thus, the effects of the manipulation would be over-additive in the serial case and additive in the parallel case. Note, however, that a limited-capacity parallel process might (again) mimic a serial process if the basic cost in RT was greater for larger set sizes than smaller set sizes, not because of an item-by-item increment of cost, but because of a greater cost with greater numbers of items.

Applying this logic, Pashler and Badgio (1985) introduced manipulations of stimulus quality that were aimed at slowing down early perceptual processes and/or stimulus-identification processes. Examples of these manipulations include high-contrast versus low-contrast and added-noise versus no added-noises. In several different versions of the experiment, Pashler and Badgio obtained additive effects of stimulus quality and set size on RT, suggesting that the perceptual processes that were affected by the manipulations occurred in parallel.

Pashler and Badgio’s (1985) results rule out any model that holds that the perceptual processing of one item must wait until the identification of any other item that has been engaged is completed. Notice, however, that this is a stronger serial model than is usually assumed. Specifically, models of visual search that include a serial component, like Feature Integration Theory and Guided Search, also include an early parallel perceptual stage of processing. Pashler and Badgio’s data are consistent with a model that has an early parallel stage that is affected by the stimulus quality manipulations, and that feeds into a later serial stage. If stimuli are ‘cleaned up’ in the parallel stage then the cost of clean-up could be incurred once for the whole display, rather than for each item in turn as they are engaged by a later serial process. If one considers Pashler and Badgio’s results in light of Sternberg’s (1969) additive-factors logic, they suggest that the processes affected by the stimulus-quality manipulations and those affected by the set-size manipulation occur at two different stages of processing: one may be parallel and one may be item-by-item serial.

A variation of this technique applied by Egeth and Dagenbach (1991) gets around some of these difficulties of interpretation. They used a constant set size of two items, and manipulated the stimulus quality of the two items independently (Townsend and Nozawa 1988 described a similar method). Thus, there were three different types of display, one in which both items were of high quality, one in which both were of low quality, and one in which one item was of high quality and the other was of low quality. They reasoned that if the two items were identified serially, then the effects of the two quality manipulations (one for each item) should be additive. That is, relative to having one low-quality stimulus and one high-quality stimulus, having two high-quality stimuli should be twice as good, and having two low-quality stimuli should be twice as bad. In contrast, if the two items are identified in parallel, then effects of the two quality manipulations should be additive. Once the system has to deal with one low-quality stimulus, it will not incur any more cost for having to deal with two.

In two different applications of this technique—search for an X among Os (or vice versa) and search for a T among Ts (or vice versa)—Egeth and Dagenbach (1991) obtained under-additive effects of the two quality manipulations. If they had obtained only under-additive effects under all applications of their technique, then their diagnostic would be subject to the same interpretative worry as that of Pashler and Badgio (1985). That is, a model in which serial stimulus identification is preceded by a parallel ‘clean-up’ process could account for the results. Interestingly, however, they did not obtain only under-additive effects. Rather, when search was for a rotated T among rotated Ts (or vice versa), they obtained additive effects. A parallel-clean-up model cannot account for these results, because the clean-up procedure would have to be engaged whenever there was any low-quality stimulus, one or two. Instead, the results suggest that the target was identified through the engagement of a serial process. Again, however, as is perhaps always the case, a parallel limited-capacity model can also accommodate the results. It is possible that increasing numbers of low-quality stimuli result in increasingly high costs on processing of low-quality stimuli (for example, it takes more time to ‘clean up’ more stimuli). Such a model could produce additive effects like those observed for the rotated Ts and Ls search task.

Mordkoff, Yantis, and Egeth (1990, Experiment 3) provide perhaps the strongest evidence in favour of parallel identification of targets in a search task. They used a redundant-targets strategy. Their target was defined by a conjunction of features. Thus, under most models that include a serial component, it should have engaged serial processing. The twist to their study was that, on some of the trials, there were two targets in the display. As is usually the case with redundant targets, target-present responses were faster when there were two targets in the display than when there was only one (see also Pashler 1987; van der Heijden 1975). The redundant-target advantage alone, however, does not help to distinguish between parallel and serial models of search, because both predict it. For serial models, the probability that a given item that engages the serial process is a target is greater when there are two targets in the display than when there is only one. Therefore, a target will be found early more often when there are two than when there is only one, and assuming a self-terminating model, mean RT will be lower. For parallel models, depending on the specifics of the model, there are a number of reasons why a redundant-target trial might enjoy an advantage over single-target trials. One of those possibilities is that the evidence that accrues for the two targets in parallel somehow combines, allowing the ‘target-present’ threshold to be reached faster than if only one target were present. Such a model is known as a coactivation model, because multiple stimuli co-activate the target representation.

The insight that Mordkoff et al. (1990) had was to apply Miller’s (1982) Race Model to their situation. They noted that, even though serial and parallel models make similar predictions concerning mean response time, they make different predictions concerning the distributions of RTs. Specifically, under a serial model, the fastest RTs for the redundant-target trials can never be faster than the fastest RTs for the single-target trials. The lower mean RTs for redundant target trials would come from the fact that the fastest times occur more often on redundant-target trials than on the single-target trials; the fastest times, however, will not differ across the two conditions. In contrast, under a parallel model, in which the two targets both contribute to activation of the fastest response times for the
In fact, the target representation during a parallel identification process, rather than from serial
ation of items.

The advantage came from coactivation of redundant-target trials may well be faster than those on the single-target trials, suggesting that the advantage came from coactivation of the target representation during a parallel identification process, rather than from serial selec-

A potential limitation to Mordkoff et al.’s (1990) study is that, like the earlier simultaneous/
sequential experiments, it used only small set sizes (<=6). Nonetheless, to the extent that it is
evidence of a cooperative relationship between two or more items, it falsifies strict serial
models that hold that identification of one item is entirely independent of identification of
others.

The problem of attentional dwell-time

A different source of evidence that has been cited as favouring parallel models over serial
models concerns the attentional dwell-time. The attentional dwell-time is the amount of time
that is required for attention to be redeployed once it has already been deployed to some item or
location (Duncan et al. 1994; Moray 1969). Posner and others performed studies in which they
attempted to measure the time that is required to remove attention from one item and deploy it
to another. These estimates tended to be of the order of hundreds of milliseconds (e.g., Posner
1980; Posner and Cohen 1984; Posner and Presti 1987). Using rapid serial visual presentation (RSVP) Reeves and Sperling (1986) found that it took several hundred milliseconds to switch attention from one stream of stimuli to another. Similarly, many studies have shown that the identification of one target in an RSVP stream interferes with the identification of another target for several hundred milliseconds (e.g. Broadbent and Broadbent 1987; Raymond et al. 1992; Shapiro et al. 1994; Chun and Potter 1995), an effect referred to as the attentional blink (see many other chapters of the present volume).

In contrast to these long estimates of the attentional dwell time, serial models of attentional
deployment in visual search seem to require fairly rapid processing of items. In the absence of
required eye movements, searches described as ‘serial’ have slopes in the range of 20–30 ms/
item for target-present trials and about twice that for target-absent trials (e.g. Treisman and
Gelade 1980; Wolfe 1998). If, for simplicity, search is assumed to be serial and self-
terminating, then subjects must process half the items on a target-absent trial. That means that the rate of processing is given by twice the target-present slope, which should be roughly equal to the target-absent slope.

Estimating the rate of processing from slopes, however, is tricky. On the one hand, fifty
ms/item might be too fast an estimate of the rate of processing. Subjects do make errors, and we
may assume that some of these errors reflect a trade-off of accuracy for speed. Also, estimating
processing rate this way assumes that search is unguided, and that on target-absent trials, it
really is exhaustive. These assumptions are likely to be incorrect (Chun and Wolfe 1996). On
the other hand, 50 ms/item might be too slow an estimate of the rate of processing. This esti-
mate assumes that each rejected distractor is marked in a way that prevents the redeployment of
attention to rejected distractors. While ‘inhibition of return’ (IOR) has been proposed as such a
mechanism (Posner et al. 1985; Klein 1988; Tipper et al. 1996; see also Watson and
Humphreys 1997; Theeuwes et al. 1998), there are questions concerning this interpretation of
IOR (Klein and Taylor 1994; Wolfe and Pokorny 1990). Moreover, several studies suggest that
whatever labelling may occur during search, it must be incomplete (Arani et al. 1984; Courtney
and Guan 1998). Most extremely, results reported by Horowitz and Wolfe (1998) suggest that
there may be no labelling at all. If this is true, then serial models would predict that items
would be sampled at random from the display without regard to prior selection, yielding esti-
mates of about 25 ms/item.

While some of the specific claims surrounding the various estimates of processing rate in
search are controversial, all that matters for present purposes is that, even taking a broad view,
serial models require an ability to process an item every 25–60 milliseconds. Studies of the
‘attentional dwell time’ would seem to falsify such models by estimating that attention is tied
up with each stimulus for much longer than the models seem to allow.

Duncan, Ward, and Shapiro (1994; see also Ward et al. 1996) were the first to question
directly the feasibility of such a fast-switching attentional mechanism in visual search, by citing
long estimates of the attentional dwell time (see also Chapter 10 in the present volume). They
estimated the attentional dwell-time using a dual-task method in which two targets (T1 and T2)
were presented in two different locations on each trial. A mask followed each target, and, at
the end of a trial, subjects identified both targets as best as they could. The time between the
two targets (stimulus onset asynchrony, SOA) varied from 0 milliseconds to about 1 second. (Note
the similarity to the simultaneous/sequential method described earlier.) The key finding was
that, at the shorter SOAs, the identification accuracy for the second of the two targets (T2) was
impaired relative to a single-task (report only T2) control condition. T2 accuracy was especially
low at the shorter SOAs. Assuming that the dual-task interference arose because processing of
the first target (T1) interfered with the processing of the second target, Duncan et al. interpreted
the duration of this interference as a measure of the attentional dwell-time. In several different
experiments, substantial T2 impairment was observed as late as 300 ms after the presentation of
T1, and some interference was often still observed as much as 500 ms after the presentation of
T1. This implies that the attentional dwell-time is approximately 300–500 ms long, which is
clearly longer than the 25–60 ms estimate that serial models of visual search require.

How should we understand these long estimates of the attentional dwell-time, and how
should we reconcile them with the slopes of the visual search experiments? One approach
would be to accept the long dwell-time and declare that search slopes of less than 300–500 ms/
item can only be obtained by processing multiple items in parallel during the dwelling of atten-
tion, and that therefore there can be no item-by-item serial component to search. Some caution
is required, however, before leaping to this conclusion. To begin, it is possible that the espe-
cially long attentional dwell-time observed by Duncan et al. (1994; Ward et al. 1996) was
specific to the conditions under which it was measured, conditions that are different in many
ways from those of standard spatial visual-search. Moore, Egeth, Barglan, and Luck (1996), for
example, showed that, using the same procedure as Duncan et al., estimates of the attentional
dwell-time could be reduced from about 450 ms to about 200 ms simply by manipulating whether or not the first target was masked. In addition, the RSVP studies described above
suggest that rejecting targets, which is what is happening during visual search, can occur at a
rate as fast as 100 ms/item. If it could not, then finding even T1 in an attentional-blink task, in
which stimuli are often presented at this rate (e.g. in Chun and Potter 1995) would be
impossible.

In summary, it seems impossible to generalize from the extremely long estimates of dwelli-
time in Duncan et al.’s (1994; Ward et al. 1996) experiments, or the RSVP experiments (e.g.,
Raymond et al. 1992; Reeves and Sperling 1986; Shapiro et al. 1994) to standard visual search.
The conditions are very different, and there are strong indications that, as the conditions are
brought closer together, the estimates of the attentional dwell-time are also brought closer together. Still, it must be noted that in no case have estimates of dwell-time gone as low as the 25–60 ms processing times that are suggested by search slopes; this remains an inconsistency.

**Beyond serial and parallel**

The inconsistency with which we were left in the preceding section disappears if we make one crucial observation about search slopes. Slopes are not dwell-times. Slopes are rates. That is, a slope of 50 ms/item is not a claim that an item can be identified in 50 ms. It is a claim that 20 such items can be identified every second. It might take several hundred ms to process each item fully. As long as the beginning of the processing of item $N$ does not have to wait for the end of processing of item $N-1$, however, dwell-time—or total processing time—can be substantially longer than the processing rate.

An assembly line is a useful metaphor for illustrating this point (see Fig. 9.5). Think of the process of seeing and identifying a stimulus as being analogous to the processes of making a car on an assembly line. The line may be capable of delivering a car every ten minutes, but it does not follow from this that it takes only ten minutes to make a car. Rather, parts are fed into the system at one end. They are bound together in a process that takes an extended period of time, and cars are released at some rate (e.g., one car every ten minutes) at the other end of the system. Is this a serial process or a parallel process? Cars enter and emerge from the system in a serial manner. Moreover, many component processes could be strictly serial. For instance, it might be possible to paint only one car at a time. In these senses, the assembly line is a serial processor. However, if we ask how many cars are being built at the same time, it becomes clear that this is also a parallel processor.

Fig. 9.5 A cartoon illustrating the assembly-line analogy for the class of search models that are both serial and parallel. There are processes that take some finite amount of time, and for which items are engaged serially. However, the engagement of one item need not wait on the completion of these processes for a previously engaged object. This model is serial; items are engaged individually. This model is parallel; items are processed simultaneously. Such a model could yield a rate of processing (i.e., the number of items per unit time that are engaged and output by the system) of approximately 20 per second, or one every 50 ms, while yielding a dwell-time (i.e., total processing time) that is considerably longer (e.g., 500 ms or more).

Getting beyond the serial/parallel debate in visual search

One way of conceptualizing this type of model (though not the only way) is to recognize it as item-by-item serial, but functionally parallel. By this we mean that items engage processing individually, but are analysed, interpreted, and transformed, at least partially, at the same time. Models of this class lie between strict serial models and strict parallel models. Returning to our metaphor, a strict serial model would hold that only one car could be on the assembly line at a time. Under this model, the search slope rate would in fact match the dwell-time. So any reliable mismatch between those two measures would falsify it. In a thoroughly parallel model, each point on the assembly line would be capable of handling more than one item at a time, and all cars could start any given process simultaneously. So, any reliable evidence that any given step in the process is serial, in the ‘one item at a time’ sense, would falsify this model.

Note that the assembly-line model is not a two-stage model in the manner of Neisser (1967). It is not a proposal for distinct serial and parallel stages of processing. Rather, it is an architecture that is at once both serial and parallel. It does not preclude the possibility of other serial and parallel stages of processing lying before or after the assembly line. For instance, parallel preattentive processing of basic visual features could still occur prior to its beginning.

The assembly-line notion is not without precedent. Miller (1988, 1990, 1993) has discussed related issues regarding the question of whether information flow is continuous or discrete. While he draws a distinction between continuous versus discrete information flow, on the one hand, and sequential versus overlapping processing stages, on the other, many of the issues are similar. The main message is that it is a continuum rather than a dichotomy. Moreover, it is not even a simple continuum, because different aspects of the information flow can be, at the same time, both more continuous and more discrete. We are similarly arguing that visual processing can be, at the same time, both serial and parallel.

More directly relevant to the present discussion is that models resembling the assembly-line idea have been proposed for visual search in particular, yet have somehow failed to prevent the development of an apparent dichotomy between serial and parallel models. For example, Harris, Smith, and Bates (1979) proposed a model in which items are loaded in series at a very fast rate (10 ms/item) into a set of asynchronous parallel processors that take a minimum of 40 ms to ‘encode’ a stimulus. In their model, actual encoding time for an individual item depends on the number of items currently being processed in parallel.

Returning to the world of metaphor, Harris et al.’s (1979) model proposes a multi-lane assembly line, and serves to illustrate the fact that there are a dizzying number of theoretical options available to the modeller within this class of model. For instance, if items are loaded in series, must they remain in the same sequence until they are off-loaded at the other end, or is it possible for an easy item to be selected second, but recognized first (an implication of multi-lane architectures)? Can the same item be on the assembly line in more than one place at a time? This makes no sense in a real assembly line (unless one imagines the different parts of the unassembled car in different places at one time). In the case of the assembly line of attention, though, the ‘cars’ are stimuli in the visual image. As another example, if items are loaded every 50 ms or so and then processed for 200–500 ms, could an item be selected by attention at time 0 and then selected again at time 150 ms while the first ‘copy’ is still on the assembly line? Something like this possibility is implied by Horowitz and Wolfe’s (1998) claim that selected items are not marked or inhibited during the course of a visual search. This idea is also similar to the re-entrant object-substitution model that Di Lollo and Enns and colleagues have applied to the attentional blink and other visual phenomena (Di Lollo et al., 2000; Enns and Di Lollo 1997; Giesbrecht and Di Lollo 1998) Finally, if an object can be represented by two copies on
the assembly line, would it be recognized more quickly? If so, what if a type of object were represented twice, as in the redundant target experiments of Mordkoff et al. (1990)?

The resolution of questions like these is beyond the scope of this chapter. For present purposes, the important point is that an architecture of this sort would produce data that look either 'serial' or 'parallel' depending on how the experimental question is asked. In the next section we consider some of the data we reviewed above in the light of an assembly-line-like model.

What can an assembly-line type model do for the serial/parallel debate?

At this stage, we can offer some thoughts as to how models of the assembly-line class could deal with various phenomena. For example, we have already seen how such a model could accommodate the apparent conflict between rates of processing that are estimated from search slopes and the attentional dwell-time. Specifically, it is possible within such a model that items in a search display engage the system individually, at a rate of approximately 50 ms/item or so, and yet require several hundred ms for full processing.

Assembly-line type models might also be able to handle the data from the simultaneous/sequential method. Recall that the basic finding here was that small numbers of stimuli could be identified just as well when they were presented to subjects sequentially as when they were presented simultaneously (c.f. e.g. Pashler and Badgio 1987; Shiffrin and Gardner 1972). When larger numbers of stimuli were presented, however, sequentially presented stimuli were processed better than simultaneously presented stimuli (Fisher 1984). Also, when stimulus identification was made more complicated, sequentially presented stimuli were processed better than simultaneously presented stimuli (Duncan 1987; Kleiss and Lane 1986; Shiu and Pashler, described in Pashler 1998). All these findings can be accommodated within an assembly-line type of model. Assume that the serial engagement of stimuli occurs until all the slots on the line are filled. As stimuli are processed and released from the system, slots free up for new stimuli to be engaged. If the number of to-be-processed stimuli is smaller than the number of slots available on the line, then no limits to processing will be revealed. When more stimuli are presented, however, the line may fill up, providing an opportunity for processing limitations to be observed. These limits might occur for several different reasons. The first reason is straightforward: with more stimuli, more slots will be filled up, leaving fewer or perhaps even no slots available. Unengaged to-be-processed stimuli would then have to wait until previously engaged stimuli are released from the line, freeing up slots. Second, as more slots fill up, items may flow through the system at a slower rate. This would mean that previously-processed items would be released from the system at a slower rate, which in turn would mean that slots would be freed up for still-to-be-engaged stimuli at a slower rate. Exactly this sort of dependence of what they call 'encoding time' on the number of other stimuli engaged is incorporated in Harris et al.'s (1979) model. Finally, another way of slowing up the system, and therefore reducing the rate at which slots become available for still-to-be-engaged stimuli, would be to increase the complexity of the processing for a given stimulus. This would then lead to the observation of processing limits when the identification task is more complex (cf. e.g. Kleiss and Lane 1986). While this scenario clearly involves a number of assumptions, and requires further specification, it does illustrate how an assembly-line type of model, which is neither serial nor parallel, could, in principle, accommodate the overall pattern of results from the simultaneous/sequential literature.

Another result that an assembly-line type of model, like that of Harris et al. (1979), could accommodate is the basic redundant-targets advantage reported by Pashler (1987) and van der Heijden (1975), as well as the more specific characteristics of redundant-target effects reported by Mordkoff et al. (1990). Specifically, the two copies of the target would be engaged individually, but they would both be present within the system at the same time, albeit at different phases of processing, thereby providing the opportunity for the processing of one to affect the processing of the other. It is this simultaneous processing that allows for both the general advantage of two targets over one (Pashler 1987; van der Heijden 1975) and for the fastest of the two-target trials being faster than the fastest of any of the one-target trials (Mordkoff et al. 1990).

Exactly where in processing the advantage for redundant targets would come from, would depend on the specifics of the model and could, in principle, be identified through either analytic or simulation techniques, which are beyond the scope of the present chapter. Still, appealing once again to analogy, we can offer one possible scenario. Imagine that there are two types of cars that are built by the assembly line: an American version with the steering wheel on the left, and a British version with the steering wheel on the right. Every time a particular frame comes through the system, the steering-wheel assembler 'recognizes' the frame as belonging of a given type and sets itself up to, for example, put the wheel on the right. If the next car in line happens also to be of the British sort, then the set-up process need not be repeated, as it is already set to go. Now imagine that there is some characteristic of these cars for which there are many versions (for example, paint colour). The process responsible for that characteristic would benefit when individual cars (items in the display) are of the same version (redundant stimuli). Notice that according to this (admittedly loose) explanation, one would expect 'redundant distractor' advantages as well as redundant-target advantages, which do in fact occur (cf. e.g. Duncan and Humphreys 1989).

Object recognition as a possible item-by-item process within an assembly-line type of model

The preceding section has shown how assembly-line models might account for some of the evidence that points toward parallel processing in search. In this final section we consider object recognition as a possible example of a mandatory serial process within search.

Whether it is serial or parallel, object recognition is a mandatory step in most visual search tasks. In Fig. 9.5, recognition is cartooned as a link between the item on the assembly line and the cloud labelled 'memory'. Recognition must involve such a link. Consider the other possibilities: (1) A visual stimulus can be as fully processed as possible and yet, if it is not in contact with its representation in memory, it cannot be said to be recognized in the present moment. (2) By the same token, if a node in memory is active in the absence of the appropriate visual stimulus, the content of that node may be thought about or remembered, but it is not seen and recognized. Finally, (3), it cannot be that mere simultaneous activity of visual system and memory node constitutes recognition. After all, it is possible to see a chair and think about an elephant without spuriously 'recognizing' the chair as an elephant. Some link between the visual and mnemonic representations would seem to be required.

Finally, that being the case, how many such links can be maintained at one time? For the sake of simplicity, consider a scene containing a tree, a horse, and a rock. If multiple links can be maintained, then tree, horse, and rock could be recognized simultaneously. With the links thus established, suppose that we ask a subject if a horse were present in the display. The query 'horse'
would activate the horse node in memory. The horse node would already be linked to the perceptual representation of the horse in the display, and so the subject could make a positive response, unaffected by the presence of other objects in the scene. In contrast, if only one link can be maintained at a given time, then upon first presentation of the display and the query 'horse?' a standard visual search would be required to determine if a horse was present. In particular, the horse node in memory would need to be compared to the representation of each item in the display in turn, until the horse was "recognized" (i.e., was linked with the perceptual representation of the horse in the display). Suppose we added to the display a windmill, a rhinoceros, and a cow. It follows that increasing the number of objects would cause RTs to increase at the usual rate of 20–30 ms/item on target-present trials. Note, however, that this 20–30 ms is not dwell-time. There is much more to processing stimuli in a visual search task than the formation of links between memorial and perceptual representations.

How could we test such an account? The critical experiment involves repeated search through the same, unchanging scene. After multiple searches for horse, tree, bird, cow, etc., all the objects must have been recognized at one time or another. If multiple recognition links between vision and memory can be maintained, then search slopes should drop from 20–30 ms/item to near zero as search is repeated. If, on the other hand, links must be established individually, then search slopes should reach an asymptotic minimum. In a host of repeated search experiments (Wolfe et al., 2000), this is exactly what happened; that is, slopes simply did not drop below the standard 20–30 ms/item value. Specifically, if a search was inefficient on the first appearance of the stimuli, then it was inefficient on the fifth, fifteenth, and fiftieth search through the same stimuli. This is not to say that observers did not learn anything about the display. After a few repetitions, they could perform the task in the absence of the stimulus, just as you, as a reader, could answer if questioned about the presence of a horse in the hypothetical scene that you have not seen. Despite this memorization of the scene, the efficiency of target detection (i.e., the search slope) did not improve. If it were possible to maintain multiple recognition links, then search should become efficient with repeated search through the same stimuli. As it did not, it follows that multiple links are not possible, and that the act of recognition has an irreducibly serial aspect to it.

As a final point, notice that this specific type of assembly-line model can account for phenomena like the attentional blink. Consider what happens to items when they emerge off this assembly line. Some items, like the target in a visual search task, are taken up from the end of the line and made the subject of other processes (for example, response processes). Other items (for example the distractors) fall into a Jastrowian abyss and are lost (James 1883 [1890] Chapter 16). In the stable world of natural vision, a lost stimulus can be recovered by another search, if needed (O’Regan 1992). In the RSVP stream of an attentional blink experiment, however, if some act of processing T1 takes 300 ms, then 300 ms-worth of stimuli will go irredeemably into the abyss. A similar account can be offered for the failure to recognize changes in scenes (e.g. Pushler 1988; Rensink et al. 1997; Simons and Levin 1997) and the failure to recall aspects of a display that were unattended at the time of presentation (Mack and Rock 1998; Moore and Egeth 1997; Wolfe, 1999).

Conclusions

In summary, there is good evidence for serial processes of attentional selection in visual search tasks. There is also good evidence for parallel processing of multiple items in visual search tasks. In recent years, this has been seen as a contradiction, a conflict in need of a victory by one side or another. We believe, however, that Newell’s (1973, p. 283) worry that ‘you can’t play 20 questions with nature and win’ applies here, and in line with earlier thinking (e.g. Harris et al. 1979; Miller 1988, 1990), we propose that the mechanisms of selective attention in visual search are both serial and parallel at the same time. Different experiments will reveal a more serial or a more parallel pattern of results. We hold that future research should focus on constraining the still large set of possible architectures of this sort, rather than attempting to resolve a false dichotomy.

References

Visual attention moves no faster than the eyes

Robert Ward

Abstract

How quickly can visual attention move from one object to another? Evidence is reviewed here from behavioural and electrophysiological studies of visual search, from single-cell recordings of visual search, from behavioural and electrophysiological studies of covert and overt orienting, and from studies of attentional dwell time. In none of these studies or paradigms are movements of attention faster than approximately 150 ms identified; in most cases time to move attention appears to be 200 ms or more. Visual attention therefore appears to move no faster than the eyes.

We have only limited capacity to identify visually presented objects. Many experiments have demonstrated costs in accuracy and speed when multiple objects must be identified, as compared to single-object performance (e.g. Broadbent 1958; Duncan 1980; Treisman 1969). Because visual processing capacity is limited, it is best to process selectively the objects most relevant to current goals, and ignore irrelevant or distracting objects. Theories of visual attention describe such selective processing. A long-standing debate in visual attention, triggered in large part by results from visual search tasks, centres on how limited capacity is allocated for visual identification: serially to one item at a time, or in parallel to multiple items simultaneously? I will argue that shifts of visual attention occur no faster than eye movements, and that despite this temporal constraint greatly limits the space of possible attention models and the nature of capacity allocation. I first describe serial and parallel models, and then apply these characterizations to behavioural and neurophysiological data, including measures of cuing, visual search, and attentional dwell-time.

High-speed serial models. These models generally assume two stages of processing: a parallel preattentive stage followed by a serial stage of selective attention (Neisser 1967). Early preattentive processes are meant to segregate, group, and otherwise organize the visual array to facilitate the subsequent allocation of attention. They might even rank and prioritize regions of the visual input that are most promising for the current task (Cave and Wolfe 1990). However, beyond this role as an organizational front-end, preattentive processing is meant to be crucially limited. Only focused attention to an object can deliver the integrated and durable representations needed to support conscious awareness and overt report. For example, in many recent models, attention acts to bind separate feature dimensions, such as colour and shape, into a single integrated object description (Luck et al. 1997; Treisman and Gelade 1980; Treisman and Gormican 1988; Wolfe 1994). For such reasons, limited-capacity attentional processing is meant to be allocated serially, so that once processing of one object is complete, attention is reallocated.

Much of the empirical motivation for serial models has come from results of standard visual search tasks, in which subjects look for a single target among a variable number of distractors. In such tasks, search times for the target often increase linearly with the number of display items, and will often demonstrate the signature of a serial self-terminating search, an approximate 2:1 ratio of target-absent to target-present slopes. Slope ratios of roughly 2:1 have a straightforward explanation in models of serial, self-terminating search. When the target is