Published in final edited form as:
Nebr Symp Motiv. 2012 ; 59: 183–208.

When do I quit? The search termination problem in visual search

Jeremy M Wolfe
Professor of Ophthalmology & Radiology, Harvard Medical School and Brigham & Women's Hospital

Jeremy M Wolfe: wolfe@search.bwh.harvard.edu

Visual searches, great and small, are a continuous part of our lives. As this is being written, I have just searched for Gate 22B at the Denver Airport. I then proceeded to search for an electrical outlet, my power cord, the correct port on the laptop, the link to the internet, and so on. These searches are drawn from the subset of total searches for which I have introspective awareness and some memory. We engage in search because there is too much visual information to fully process. Even if the sign for Gate B22 is in my visual field, I still need to use attentional mechanisms to select that object from the welter of other stimuli on Concourse B because attention is required to read that sign (Rayner, 1983). Without worrying, for the present, about who this "I" is that is using attention, it makes some sense to imagine that I was asking my search engine to conduct these specific searches. Even if I am not engaged in what seems like deliberate search, covert attention is selecting one object after another or maybe a few objects at a time, much as the eyes are fixating on one thing after another. The deployments of attention may be based on the bottom-up, stimulus driven salience of the stimulus (Einhauser, Spain, & Perona, 2008; Foulsham & Underwood, 2008; Koch & Ullman, 1985; Masciocchi, Mihalas, Parkhurst, & Niebur, 2009) (Is that a bottom-up, attention-grabbing bird flying around in Concourse B? Yes, it is!). Alternatively, attention might be guided by top-down task demands (Jan Theeuwes, 2010), even if those top-down demands do not usually seem to rise to conscious awareness. Consider the searches that could be involved in avoiding obstacles as you navigate down the concourse (Hamid, Stankiewicz, & Hayhoe, 2010; Jovancevic-Misic & Hayhoe, 2009). The obstacles to be avoided might not be the most salient items but you manage to direct attention to them without introspective awareness of that search.

A vast set of research topics are present in this evocation of a trip down the airport concourse. Do we attend to objects or locations (Goldsmith, 1998; Logan, 1996; Roelfsema, Lammer, & Spekreijse, 1998; Yeari & Goldsmith, 2010)? What are the features that contribute to bottom-up salience (Jeremy M Wolfe & Horowitz, 2004)? Do those features really "capture" attention (Jan Theeuwes, 1995); Bacon & Egeth, 1994)? Do new objects capture attention? (Yantis & Jonides, 1996; Franconeri, Hollingworth, & Simons, 2005)? How is top-down control of selection organized (JM Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004); Hamker, 2006; Jan Theeuwes, 2010)? How do scene semantics guide the deployment of attention (Henderson & Ferreira, 2004; (Torralba, Oliva, Castelhano, & Henderson, 2006) (M. L. H. Vo & Henderson, 2009)? How is this implemented in the brain (Reynolds & Chelazzi, 2004; Buschman & Miller, 2009)? We could continue to list topics. For a daunting catalog from a computational viewpoint, see Tsotsos (Tsotsos, 2011). Each of these topics has generated a substantial research literature. In this chapter, however, we will focus on a different aspect of search that gets somewhat less attention. What happens when the search is unsuccessful? When is it time to abandon a search without having found a target? A moment's introspection reveals that, like successful searches, these abandoned searches occur all the time. Is there anyone I know in this airport waiting area? I can search for some period of time but, at some point, I need to give up and move on to the next task. How is that accomplished? If you find the target, there is an obvious signal that you are done. What is the signal that allows you to quit if no target is found? This problem of search termination is central to a variety of socially critical search tasks. Indeed, the airport is home to one of the signature examples of the search termination problem. Passing through security, your carry-on luggage is x-rayed and examined in a visual search for 'threats' like guns, bombs, and knives. Fortunately, most bags do not contain threats, meaning that, most of the time, the screener's task is to decide when it is time to abandon the search without finding a target. The stimulus is complex and could be examined for a long time, but, in that case, the line at the checkpoint would become unacceptably long. Of course, quitting too soon raises the possibility of missing a real threat (Wolfe, Horowitz, & Kenner, 2005), an error with far more consequence than a longer security line.

Similar search termination issues are raised in medical radiology. If you are screening mammograms for breast cancer, you do not want to miss any cancers but, at the same time, you need to get through all of the cases. When is it time to move to the next case? The radiology situation has some interesting characteristics that differ from the checkpoint search (beyond the obvious differences of stimulus materials). At the checkpoint, once a single gun is found, the search is done. In radiology, it is often important to find all of the signs of cancer (or whatever the radiologist may be looking for). Thus, in the medical setting, even if a 'target' has been found, there is still a search termination question. How sure are you that you have found everything that needs to be found in this image? The probability of missing a target is higher if another target has been found; a problem known as "satisfaction of search" (Berbaum, Franken, Dorfman, Caldwell, & Krupinski, 2000; Berbaum et al., 1990; Fleck, Samei, & Mitroff, 2010; Nodine, Krupinski, Kundel, Toto, & Herman, 1992).

In this chapter, we will focus on the fundamental mechanisms of search termination with allusions to these more applied topics but without a full treatment of them. We will trace the development of ideas about search termination from early ideas about serial exhaustive search to a more plausible account and some pointers toward possible future progress.

How shall we model the target absent trials?

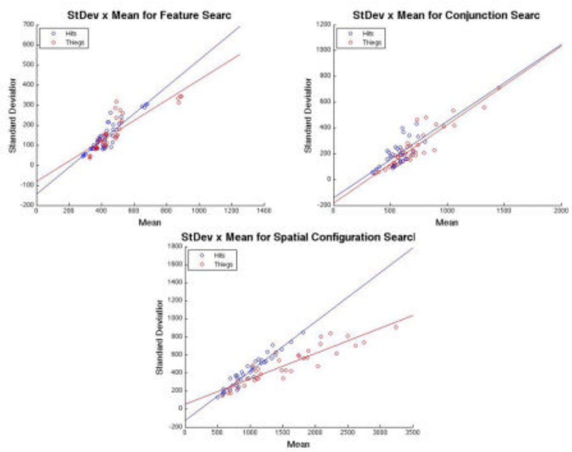
Model 1: Serial Self-terminating search

Consider a basic search task, as shown in Figure 1: Here you are looking for a target "T" among distractor, "L"s. In an actual experiment, we would arrange for the items to be large enough that acuity is not constraining performance. Typically, we would vary set size – the number of items in the display -and we would measure reaction time (RT) and accuracy. In cases where the display is visible until the observer responds, it is the RT data that are of most interest. An experiment of this sort would very typically produce data that look something like those shown in Figure 2. The measure of greatest interest is the slope of the RT x set size function. In a task like this, slopes are typically in the range of 20–40 msec/item for target present trials and something more than twice that for target absent trials.

This pattern of results suggested a serial self terminating search to Anne Treisman (Treisman & Gelade, 1980) following similar ideas in memory research (Sternberg, 1966). The idea, as illustrated in Figure 3 was simple and reasonable. Items would be selected, at random, one after another until the target was found or until all items were rejected. If there were N items, the target would be found, on average, after (N+1)/2 selections from the display. The display could be rejected after all N items were examined. The result should be a slope ratio of close to 2:1. Treisman's data were consistent with this 2:1 prediction.

There are problems with Model 1. With more extensive data sets, it turns out that the search ratio in search tasks of this sort is typically significantly greater than 2:1 (J M Wolfe, 1998).

For the data shown in Figure 2, for example, the hypothesis that Absent Slope = 2x(Present Slope) can be rejected (paired-t test, 2xPresent-Absent; $t(19)=2.5$, $p=0.023$). For these data, the average slope ratio is 2.5:1, very similar to what was found in Wolfe (1998). Note also that the variability of the absent trials is much higher than that of the present trials. This is also true for the RTs of individual observers contrary to what might be expected from a simple serial self-terminating model. After all, on absent trials, search always ends after all N items have been rejected while, on present trials, search could end after the first deployment of attention or the last or after any number of deployments between 1 and N. Playing Around with Mean vs. StDev:



The critical problem with a simple serial, self-terminating account is found in a classic experiment of Egeth, Virzi, and Garbart (1984). They didn’t use Ts among Ls but using those stimuli as an example, imagine that half the elements were red and you were told that the target was black. You would not spend time examining red items and you would not need to search those red items in order to declare that the target was absent on blank trials. In an experiment of this sort, search slopes on target present trials would be reduced by about half of what is shown in Figure 2. The absent slopes would be similarly reduced, suggesting that Os searched through only half the items. These and related results require a modification of the serial self-terminating model.

Model 2: Serial self-terminating search in a subset

The obvious modification in the basic serial self-terminating model is to propose that the search is terminated after an exhaustive search through the relevant subset. In the example given above, that would be the set of all red items. This model also runs into difficulties. One challenge comes from conjunction search tasks of the sort shown in Figure 4.

In this task, Os look for a target defined by the conjunction of two features; here, the light red vertical item. Treisman had originally proposed that conjunction searches produced the same pattern of results produced by searches like the T vs L example (Treisman & Gelade, 1980). However, subsequent research showed that conjunction searches could be much more efficient with shallower slopes (Nakayama & Silverman, 1986; Sagi, 1988; J.M. Wolfe, Cave, & Franzel, 1989; Zohary & Hochstein, 1989). How should observers perform on absent trials? If the relevant subset was the set of items that were EITHER red OR vertical, then the subset is the entire set – and that cannot be right. The slopes are too shallow. If the subset was the set of items that were BOTH red AND vertical, then the subset is empty on target absent trials and the slopes are too steep to support that assumption. Neither of these possibilities describes observers’ behavior. They produce slopes on absent trials that are about twice the slope of target present trials. An example, from the same observers, shown in Figure 2, is shown in Figure 5.

One might propose that Os searched through half the items, based, perhaps, on color. However, that version of a subset-search hypothesis can be rejected. When observers are forced to search through a subset based on color, performance looks very different from performance in standard conjunctions searches of the type illustrated in Figure 4 (Friedman-Hill & Wolfe, 1995). Moreover, some conjunction searches can be very efficient, with slopes near zero (J Theeuwes & Kooi, 1994). This is another challenge to any model that proposes that blank trials involve an exhaustive search through a feature-defined subset of items.

We will point to another challenge here and return to it later. A model based on exhaustive search through the set of items or some subset of items is plausible when the stimuli are well-isolated items on a blank background, as in typical laboratory search experiments (and as in Figures 2 & 4). It is much more difficult to implement such a model in a real scene because it is all but impossible to decide what the set size might be (Neider & Zelinsky, 2008; (Jeremy Wolfe et al., 2008). Look at the scene in front of you and try to decide what the set size might be. Still, an exhaustive search through some subset might still be a plausible model if a way could be found to define the subset.

Model 3: Serial Self-Terminating Search up to an activation boundary (Guided Search 2.0)

A version of this type of subset search was proposed in Guided Search 2.0 (J M Wolfe, 1994). In all of the incarnations of the Guided Search model, attention is guided by basic attributes of the stimulus such as color, orientation, size, et al. (Jeremy M Wolfe & Horowitz, 2004). As noted earlier, guidance comes in two forms. Attention is guided to an item in a bottom-up, stimulus driven manner if an item differs from its neighbors in a guiding attribute (red among green, vertical among horizontal, and so forth). As discussed extensively by Duncan and Humphreys (1989), the greater the difference between target and distractors, the easier a search will be (red among green is easier than red among orange). The greater the featural heterogeneity of the distractors, the harder the search will be (red among homogeneous orange distractors is easier than red among a variety of different colors).

Guidance can also be top-down, user driven. In Figure 4, bottom-up activity is essentially noise. Effective guidance to the light red vertical item comes from top-down guidance to red and to vertical. In Guided Search, each of these sources of guidance contributes to an overall activation map. Attention is directed to the most active item/location in that map. The map must be degraded by noise. Otherwise, a search like the conjunction search of Figure 4 should yield a slope of zero because the target is the only item with two target attributes. In the absence of noise, guidance to red and to vertical would lead directly to the one red vertical item first time, every time. Some of the noise will come from bottom-up activation. The juxtaposition of red and green or vertical and horizontal items makes those items salient. That salience is not useful.

When the sources of guidance, useful and otherwise are summed up and some noise is added, the result will be that targets in a search like the conjunction search of Figure 4 will have some activation drawn from a distribution and different distractor types will have activations drawn from lower but overlapping distributions. If attention is directed to the most activated item, that first item will be the target on some trials but on other trials some distractors will be examined before the target is reached. Returning to the absent trials, a reasonable approach would be to set an activation threshold below which only very few targets are ever found. That threshold could define the subset on each trial and unsuccessful searches could end after an exhaustive search through that subset. If a target was not found in the set of items above the threshold activation, then it is time to quit.

Data like those shown in Figure 3 constrain the placement of the activation threshold in this model. If the threshold is set to examine only items with high activation, then search will be abandoned on too many target-present trials before the target is attended and the miss error rate will be too high. If the threshold is set too low, few targets will be missed but the RTs will be too long. Given some assumptions about the noise in the activation values, it was possible to use one set of parameters to simulate a substantial set of search experiments in Guided Search 2.0, producing reasonable simulated target present and absent RTs (J M Wolfe, 1994).

There is an important assumption underlying this model and, indeed, all of the models that propose some sort of exhaustive search through some set of items on absent trials. To do an exhaustive search, one needs to know which distractors have been rejected. Put differently, search needs to sample without replacement from the display. Many models of search assumed such sampling and a mechanism, inhibition of return, had been proposed (R. Klein, 1988). Unfortunately, the assumption does not appear to be correct.

In 1998, Horowitz and Wolfe (1998) tried to test the assumption directly. They asked, what would happen if search were forced to be sampling with replacement? Their “dynamic search” method is illustrated in Figure 6.

In Dynamic Search, observers see a sequence of frames. The items are the same on each frame but they are randomly replotted each time. A target item will be present on every frame of a target present trial and on no frames in absent trials. Dynamic search must require sampling with replacement, unless the search can be accomplished in a single frame. It can then be compared to a standard, static search condition. If, as required by the models sketched so far, rejected distractors are remembered in static search - that is, if static search is sampling without replacement - then there is a clear prediction for the relationship of slopes in the dynamic and static conditions. If standard, static search produces a slope of N msec/item, dynamic search, sampling with replacement, should produce a slope of 2N msec/item (Horowitz & Wolfe, 2003).

The results rejected this hypothesis. In Horowitz and Wolfe (1998), the frame rate was 10 Hz. The slopes on target present trials were essentially the same in static and dynamic conditions. Horowitz and Wolfe reasoned that dynamic search *had* to be search with replacement. Thus, if static search produced the same result, it followed that static search was also search with replacement and they titled their paper “Visual search has no memory”. Vigorous controversy ensued (Dodd, Castel, & Pratt, 2003; Gilchrist & Harvey, 2006; Horowitz & Wolfe, 2003; A Kristjansson, 2000; Peterson, Kramer, Wang, Irwin, & McCarley, 2001; Shore & Klein, 2000). Figure 7 shows the results of a replication of the original dynamic search result from Horowitz and Wolfe (2003). This time the frame rate was 2Hz, large set sizes were used, and in one version of the dynamic condition, targets could only appear in a few display positions, unknown to the observer. This was done to thwart “sit and wait” strategies in which the observer might pick one location and simply wait for the randomly plotted target to appear (von Muhlenen, Muller, & Muller, 2003). The results, shown in Figure 7, again show dynamic and static search having similar slopes.

If visual search really had no memory, one would think that perseveration would be a serious problem. Imagine that there was one salient distractor in the display. In the no memory account, what would keep attention from continuously revisiting that item? Moreover, the papers cited above that responded to the original “no memory” claim, make a case there is at least some limited memory in visual search. For methodological reasons, it is hard to differentiate between the consequences of a little memory and no memory in the dynamic search task. Perhaps the most plausible position is that “inhibition of return is a foraging facilitator in visual search”(R. M. Klein & MacInnes, 1999). That is, perfect memory for the rejected distractors does not exist but there is enough inhibition to prevent perseveration and to bias attention toward new items. This seems reasonable but, returning to the problem of absent trials, the models presented thus far rely on perfect memory for rejected distractors, and that does not exist. A different type of model is needed.

Model 4: Timing or counting models

Even if the observer cannot rely on perfect memory for every deployment of attention, there is little doubt that he is accumulating some information about the ongoing searches. Suppose that an observer had information about the mean time required to find the target and the variance of that time. He could set a threshold in time rather than in activation. “If I search for N msec without finding the target, the probability that I will find a target is low enough that I might as well quit.” Observers would not, in fact, need to compute the mean and variance (implicitly or explicitly). Suppose that an adaptive process changed the quitting time on blank trials based on feedback from the ongoing sequence of trials. Threshold would be decreased and observers would quit more rapidly after correct responses and the threshold would increase after errors. If the step size on this ‘staircase’ is set appropriately, it would estimate a quitting threshold that would produce an acceptable error rate (Chun & Wolfe, 1996). As an alternative to measuring time, the observer could count rejected distractors (sampled with or without replacement, it would not matter) and could quit after sampling some threshold number of items. As with a timing threshold, a counting threshold could be based on the number of items sampled in order to find targets on previous trials. These timing or counting models can be implemented with diffusion (Ratcliff, 1978) or accumulation (Brown & Heathcote, 2008; Donkin, Brown, Heathcote, & Wagenmakers, 2011) methods. In either case, search is terminated when the accumulating or diffusing signal reaches a termination threshold. That threshold, as noted, would go up in response to error and down in response to correct responses.

Chun and Wolfe (1995) looked for evidence for this adaptive mechanism. They ran observers in a triple conjunction (color X size X shape) task at a single set size of 25. Os made 3.3% miss errors in an easy version of the task and 8.0% errors in a harder version. Chun and Wolfe looked at RT as a function of the position of a trial, relative to a miss error. The results, replotted from the original paper, are shown in Figure 8. It can be seen that RTs become faster after correct trials and markedly slower after a miss error.

In a function of this sort, we can see the searcher, estimating how long it should take to complete a search with an acceptable number of errors. However, there are a number of complications. First, the time for a given search is obviously dependent on how many items are present in a display. A timing or counting threshold that was established using one set size would be obviously incorrect for another set size. In practice, any such quitting threshold must be set relative to the set size on the current trial. We know this because performance on absent trials in standard search tasks does not change markedly whether set sizes are blocked or mixed. (J M Wolfe, Horowitz, Palmer, Michod, & VanWert, 2010). This raises a second problem. If observers can adjust the quitting criterion based on the set size, they must be able to derive that set size at the start of the trial. Since we know that exact counting is only possible in the subitizing range of up to about 4 items (Trick & Pylyshyn, 1994), set size must be an estimate based on our ability to roughly enumerate larger number of items (Krueger, 1984; Dehaene, 1997).

To recap, the working model would now say that the observer, at the start of a trial, makes an estimate of the set size and then sets a quitting threshold. It could be a counting threshold. In that case, the threshold would be set as some constant times the estimated set size. The constant would go down if the search could be based on a subset (I only need to search the 25% of objects that are green). It would also depend on whether search was sampling with or without replacement or somewhere in between. Alternatively, search termination could be based on a timing threshold, based on a calculation of the average time per item that must be devoted to the display, in order to produce a reasonable error rate. Models of this sort will run into problems when the observer is confronted with a real scene, as opposed to a display containing a countable number of items. As noted earlier, we simply have no idea how to count the number of searchable items in a real scene. We have some ideas about how to approach this problem. Observers can probably extract an "effective" set size (Neider & Zelinsky, 2008) from the scene based on a variety of rapidly computed aspects of the gist of the scene (J M Wolfe, Vo, Evans, & Greene, 2011); (M. L.-H. Vo & Henderson, 2010); Oliva 2005). Thus, for example, if you are looking for your thumb drive, objects the size of your computer screen probably do not enter into the calculation of effective set size. Moreover, this size constraint is probably calculated in three dimensions and not just in the image plane. Layout in depth is calculated quickly (Greene & Oliva, 2009) and so the book, located across the room, that happens to subtend the same visual angle as a much closer missing thumb drive, nevertheless, is not a candidate for search because it is the wrong size in 3D even if it would be plausible in 2D (Sherman, Greene, & Wolfe, 2011).

Beyond figuring out the effective set size in a scene, other properties of the scene will be important as well. Guidance by basic features like color will enter into the calculation of a quitting time. If you are looking for your car, it will make a great deal of a difference if the car is an unusual shade of lime green or not. If that unusual color is not present in the visual field at all, you are likely to be able to abandon the search for the lime green car rapidly. The search for a more generic silver gray car will not be abandoned so quickly because your initial assessment of the scene will give you more hope that it is present, even if it is not. In addition, clutter and crowding become issues in real scenes (Mary J Bravo & Farid, 2004; Felisberti, Solomon, & Morgan, 2005; Levi, 2008; Rosenholtz, Chan, & Balas, 2009; Vlaskamp & Hooge, 2006). Even if all the other factors are controlled, intuition holds that the search for a fully visible carrot peeler will be harder in a jumbled kitchen drawer than in a neat one. No one really knows how to compute clutter or crowding for these purposes, though progress is being made (Mary J. Bravo & Farid, 2008); Rosenholtz et al., 2009).

Nevertheless, we can modify the current story to run as follows: When the scene (or an artificial search display) is presented to a viewer and a search task is defined, a quitting threshold is set based on an assessment of the gist of the scene. That gist will include an estimate of the number of candidate targets. Candidate targets will be defined by their basic features and a variety of scene-based properties. The threshold will be further adjusted on the basis of an estimate of the ease with which candidate targets can be located and analyzed amidst whatever noise, clutter, or other obstruction is present. Diffusion or accumulator models can use a signal that measures time or that counts rejected items. Search can be terminated when that signal reaches the quitting threshold.

Target Prevalence

One factor that has not been mentioned but that has a substantial effect on search termination is the likelihood that a target is present. Returning to that carrot peeler, you should search for a longer time in the jumbled gadget drawer in your kitchen than you should search in a drawer in your office, even if we arrange for the two drawers to be visually equivalent. The prior probability of target presence is simply much higher in the former case than the latter. Moving from intuition to data, Wolfe, Horowitz, and Kenner

(2005) had observers search for black and white objects on a noise background. In different blocks, the targets were present on 50%, 10% or 1% of trials. These are quite laborious experiments since it takes 2000 trials to collect a mere 20 target present trials at 1% prevalence. Nevertheless, even with the limited statistical power imposed by the relatively small number of present trials, the results are dramatically clear, as shown in Figures 9 and 10.

Figure 9 shows the miss (“false negative”) error rates. Miss errors are much higher at low prevalence.

Figure 10 shows the RT data. The data for 50% target prevalence show the typical RT x set size functions with absent trials being slower than present and having a slope of somewhat more than twice the present trial slope. In dramatic contrast, in this experiment, the absent RTs are actually shorter, on average than the present trial RTs. This RT result is somewhat more dramatic than usual. However, it is obvious that prevalence has a very substantial effect on target absent trials and that effect is not accounted for in the model sketched above. The basic prevalence result has been replicated many times (M S Fleck & S R Mitroff, 2007; Godwin et al., 2010; Lau & Huang, 2010); Kunar, Rich, & Wolfe, 2010; VanWert, Wolfe, & Horowitz 2009). There is always a rise in miss errors and a fall in target absent RTs as prevalence falls.

These effects of target frequency have been anticipated in tasks other than visual search. What has been called Hick-Hyman Law proposes RT increases with the number of alternatives (takes longer to respond with one of four keys than with one of two) and this has been taken to reflect a general relationship between stimulus frequency and RT (Hick, 1952; Hyman, 1953; (V. Maljkovic & Martini, 2005). Moreover, in the vigilance literature, RT has been shown to increase as signal frequency decreases (Parasuraman & Davies, 1976). The vigilance literature also documents the rise in miss errors as signal frequency decreases (Colquhoun & Baddeley, 1967; Mackworth, 1970).

The prevalence effect is a potentially important phenomenon beyond the lab because a number of critical search tasks are low prevalence search tasks. Clear examples include medical screening (Ethell & Manning, 2001; Gur et al., 2003; Kundel, 1982) and airport baggage screening. In each case, the target is very rare and in each case, miss errors are very undesirable. At the same time, the professionals doing these tasks are under time pressure to get through the workload. Is low target prevalence a source of errors in the field in any of these domains? Experiments are in progress as this is being written. We do know that expertise is not insulation against these effects. In one experiment, two groups of cytology technicians, who read Pap smear, cervical cancer tests, examined photomicrographs of cells. Each group read one set of stimuli at 50% prevalence and another at low prevalence, either 2% or 5%. One group simply rated slides on a 4-point normal/abnormal scale. The other group also localized apparent abnormalities. In the first group, false negatives/miss errors were 17% at higher prevalence and 30% at low prevalence. In the second group, false negative rates rose from 27% to 42% (One cannot make comparisons between the two groups because the stimulus sets were different). Incomplete data collection strongly suggests that other search experts will prove to be just as vulnerable to the prevalence effect as observers in the lab.

With miss errors going up at low prevalence and RT going down, an obvious thought is that the prevalence effect is nothing but a speed-accuracy tradeoff. Fleck and Mitroff (2007) made an argument of this sort. It was based on data that showed that they could eliminate the prevalence effect if they simply allowed observers to rescind responses that they knew were in error. Everyone who has done visual search RT studies knows this phenomenon.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

You commit yourself to making target absent motor response. Then you find the target a moment later but it turns out to be a moment too late to recall the motor act.

There are a number of reasons to think that, while errors of this sort occur, they are not responsible for the main prevalence effect of interest. Most importantly, a speed-accuracy trade-off should represent a loss of sensitivity at low prevalence. (NOTE: We are using “sensitivity” to refer to what is indexed by d’ or the area under an ROC curve in signal detection experiments. This is different from the usage in the medical community where sensitivity refers to the “hit” rate (P(correct|target-present)). The medical literature uses “specificity” to refer to the true negative rate (P(correct|target-absent)). In our original experiments and in the Fleck and Mitroff (2007) study, there were very few false alarm errors, making signal detection measures unreliable. When we used a simulated baggage search task that produced false alarms, it became clear that prevalence had its primary effect on response criterion, not on d’. The data shown in Figure 11 illustrate the point. The data come from an experiment in which prevalence varied sinusoidally over the course of 1000 trials from near 1.0 to near zero and back to 1.0 (J M Wolfe & VanWert, 2010). Each data point represents 50 trials from each of 13 observers. The color and shape coding of the points show the prevalence for those 50 trials. Clearly, the data points slide along a receiver operating characteristic (ROC) curve with low prevalence conditions characterized by high miss errors and low false alarms (conservative criterion) and high prevalence showing low miss errors and high false alarms (liberal criterion).

Figure 12 shows sensitivity (d’) and criterion (c) for the data presented in Figure 11. In a wide range of prevalence experiments, d’ tends to be somewhat higher at low prevalence than at higher prevalence (Kundel, 2000; (J M Wolfe et al., 2007). As can be seen, this appears to be the case in Figure 12. The effect is significant if all data points are included (r-sq = =0.40, p=0.0025). If the extreme points are excluded on grounds that they are very unstable, the relationship is marginal (for prevalence between 0.1 and 0.85, r-sq=.32, p=0.053). This is probably an artifact of the underlying assumption that the “signal” and “noise” distributions are of equal variance. ROCs like the one shown in Figure 11 become straight lines if plotted on Z-transformed axes. If the variance of the signal and noise distributions are the same, the resulting zROC has a slope of 1. Slopes in these prevalence experiments tend to be less than 1 (0.6 for the data in Figure 11). However, any of the various ways to deal with unequal variance retain a strong relationship of criterion to prevalence (in various approaches, r-sq > 0.75, all p < 0.0001). The dependence of criterion on prevalence in search is anticipated in non-search tasks where changes in event frequency produce criterion shifts, rather than changes in sensitivity (Healy & Kubovy, 1981; Swets & Kristofferson, 1970).

Another line of evidence suggesting that prevalence effects are not simple speed-accuracy tradeoffs comes from the RT data. If prevalence effects were simple speed-accuracy tradeoffs, we might expect that target-present responses or, perhaps, all responses would become very fast at very high prevalence when observers could respond “present” with a good chance of being correct, no matter how quickly they pressed the key. However, as can be seen in Figure 13, there is a fairly modest effect of prevalence on target present trial RTs. There is a much larger effect on absent trials with very slow RTs accompanying the highest prevalence.

Prevalence is changing both speed and accuracy. However, as discussed in Wolfe and VanWert (2010), we need to think about two different criteria, each of them influenced by target prevalence. Thus far, we have been talking about a search termination threshold or criterion. When some accumulating quantity like elapsed time or number of items attended reaches that threshold, search ends, presumably with an “absent” response on most trials.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

While this accumulation to a termination threshold is ongoing, there are other decisions that need to be made about every attended item. Is that item a target? These 2-alternative forced-choice decisions must have their own criterion. Prevalence alters both the search termination threshold and the target/non-target decision. As prevalence goes down, the search termination threshold comes down; observers are willing to abandon search sooner. At the same time, the 2AFC target/non-target decision criterion becomes more conservative; observers are less willing to declare that an item is a target. There are various lines of evidence to suggest that these are dissociable criteria. One of these is Experiment 2 of Wolfe et al. (2007). In an effort, to "cure" the prevalence effect, Wolfe et al., forced observers to slow down their responses. An initial block of 50% prevalence trials established the average time required to find targets. In a subsequent low prevalence block, the computer produced a warning whenever an absent trial RT was below 1.3X of that average. This had the desired effect – observers learned to slow their average target absent RTs by over a second (from <1500 msec to > 2500 msec). Once behavior stabilized, they required warnings on only a few trials. Apparently, they had reset the termination threshold. However, there was no significant impact on errors. Miss errors remained much higher at low prevalence than at higher prevalence. While this was a setback in the quest to find a way to reduce miss errors, it does show that time to search and decisions about attended items are governed by dissociable responses to prevalence.

Estimating Prevalence

As noted above, the observer must estimate the set size, crowdedness, and/or clutter in a display in order to set a search termination criterion. Similarly, the observer must estimate the prevalence, if prevalence is going to have an influence on performance. One could imagine the estimate of prevalence being set by the prior history of the search. The frequency with which you find the target before the current trial would produce the estimate of the prevailing prevalence on this trial. Alternatively (or additionally), the estimate of prevalence could be based on top-down, semantic knowledge from outside of the search itself. That is, you could be told that the prevalence is 2% or 50%. Under real world conditions, versions of both types of information are present. A radiologist knows that breast cancer is rare in a screening population and knows that she has found very few cancers in this collection of cases. Lau and Haung (2010) found no effect of explicit instruction on error rate and concluded that the prevalence effect is based entirely on past history with the task. However, Ishibashi et al, (2011) have subsequently reported a small effect of instruction on RT.

If prevalence effects are based on prior history, how much prior history is being taken into consideration. In the experiment described above (Figs 11–13), Wolfe and VanWert (2010) varied prevalence sinusoidally over 1000 trials and found that error rates and RTs also varied in a roughly sinusoidal fashion. Given this variation in prevalence, each prevalence value was experienced twice, once as prevalence was falling and once as it was rising back to 1.0. Based on the difference between performance at the same prevalence value in the rising versus the falling portion of the sinusoid, Wolfe and VanWert concluded that Os were using a prevalence estimate based on 40–50 trials. However, this may not be the best estimate of what we can call the "prevalence window" because the prevalence is changing and it is changing in a predictable manner.

How wide is the prevalence window when the prevalence for a block of trials is fixed? Even if overall prevalence is fixed, there will be local variations. Consider 8 successive trials from an experiment with an overall prevalence of 50%. Chance variation might produce 3 target present trials in one sequence, 7 in another, and so forth. Figure 14 shows the effects of just such random fluctuations in local prevalence for a window of 8 trials averaged over the data from 20 observers. The data happen to be taken from a search for a 2 among 5s (a task that will produce very few false alarm errors). Other data sets produce similar results. This particular data set includes variation in set size (5–20), which will introduce large variability into the absent RTs. Nevertheless, as can be seen, there is a substantial and significant effect of random fluctuations in local prevalence on absent trials RTs ($p = 0.04$). Target present RTs show no dependence on local prevalence ($p=0.83$).

To estimate the prevalence window, we measured the correlation between RT and the local prevalence for windows of different sizes. Figure 15 gives the results of this analysis for a different data set; this time, a large data set where we collected 4000 trials per subject per condition in order to examine RT distributions (J. M. Wolfe, Palmer, & Horowitz, 2010). The large number of trials improves the statistical power of the analysis since, as can be seen on the y-axis of Figure 15, while prevalence effects are reliable and quite large, they do not account for much of the variance in an experiment of this sort. Here the maximum correlation occurs in the range of 5–8 trials, suggesting a fairly small prevalence window. Interestingly, this is comparable to the range for priming of pop-out (V Maljkovic & Nakayama, 1994). In that paradigm, the color of items going back about 8 trials into the past has an impact on the RT of a pop-out color search on the current trial.

Value: One more factor

The discussion thus far has failed to consider how badly you want to find whatever it is that you might be looking for. It seems quite obvious that this, too, will have an effect on search termination. You would search for a lost 20 dollar bill for longer than you would search for a 1 dollar bill. This example has the added virtue of making it intuitively clear that the effect of value on RT will be an effect on target-absent RTs. The 1 and 20 dollar bills are essentially the same visual stimulus. Thus, the time to find the bill, if the bill is successfully found, is unlikely to depend on its value. It is the time devoted to unsuccessful search that will be influenced by value.

There has been a recent uptick in interest in the effects of reward in visual search (Hearns & Moss, 1968; Hickey, Chelazzi, & Theeuwes, 2011; Hickey & Theeuwes, 2008; (A. Kristjansson, Sigurjonsdottir, & Driver, 2010). However, as with most other topics in search, much less attention has been devoted to the impact on target absent trials. There has been some discussion of reward in the context of the prevalence. After all, if one is concerned that low prevalence is pushing observers toward elevated miss errors, one should be able to move them the other direction on the ROC by changing the payoff. There is some evidence that prevalence effects are resistant to manipulations of payoff (Healy & Kubovy, 1981; Maddox, 2002) and Wolfe et al. (2007) argued that it would not work in settings like airport security and medical screening. However, more recent work shows that, if deployed correctly, payoff manipulations can affect the error rates, counteracting the prevalence effect (Navalpakkam, Koch, & Perona, 2009). There is a need for work on reward effects on reaction time.

Moving to the next field

The model we have been sketching asserts that the observer in a visual search task is monitoring the time spent searching or the amount of searching that has been done. Search is terminated when the relevant quantity reaches a search termination threshold. On a given trial, that threshold is set by an estimate of the number of items in the display and an assessment of the difficulties imposed by crowding and clutter. The threshold is also influenced by the likelihood that a target is present. This estimate of prevalence seems to be based on recent search history and, perhaps, on something more like semantic knowledge. You don't need to have looked for President Obama multiple times in order to understand that he is unlikely to be in your kitchen. Finally, the termination threshold is influenced by the intrinsic value of the search target.

Now let us re-imagine the task. Suppose that each search display is a patch of a habitat in which some animal is searching for its food. In each patch, there either is or is not a food item. That assumption, convenient for 2AFC tasks, might not be entirely realistic but, if one imagines fairly sparse, evenly distributed food and patches of the right size, it is not a bad assumption. Described this way, visual search has much in common with foraging problems, as studied in Behavioral Ecology (Stephens & Krebs, 1986). The search termination problem becomes what is known as the “patch leaving” problem. When should our animal stop searching/foraging in one patch and move to the next?

Unlike visual search, where search termination is a bit of an orphan problem, patch leaving in behavioral ecology has attracted a lot of attention. Many accounts are versions of Charnov’s Marginal Value Theorem (Charnov, 1976) which asserts that the animal should move when the marginal rate, the rate at which resources are being extracted from the patch, drops below the average rate of return for the environment. If you imagine picking berries from a bush, you pick at some rate. At some point, the rate begins to drop as the bush is depleted. It is time to move to another bush once the rate drops to a point below the average rate at which your berry bucket is filling up. If it takes a long time to get to the next bush, you should stay longer on the current bush because that travel time reduces the average rate of return.

Simple versions of the marginal value theorem assume that the average rate is known and uniform (Pyke, Pulliam, & Charnov, 1977). Realistic complications ensue if you endow the animal with an ability to sense the distribution of resource in an uneven habitat. Other variables might include how long it takes to consume an item or whether one type of item is more common than another.

It is not hard to map foraging variables to visual search variables. The various factors that influence the slope of RT x set size functions are influencing the observer’s rate of return; how many targets he eats per unit time. Endowing the observer with preattentive processes that give that observer the numerosity of a display and guide his attention to likely targets are like the processes that would allow an animal to notice that one patch appears to be more promising than another. Attentional limits have been proposed to constrain behavior in foraging as well as in the search domain (Dukas, 2002, 2004). This is not to say that there is a trivial equivalence of issues in search and in foraging. However, the rich theoretical work in behavioral ecology provides a promising habitat for visual search researchers. The control and comparative ease with which visual search data can be collected represents an opportunity to test some of those theories.

A brief conclusion

The search termination problem is important. Many searches get terminated without success. Searches for unknown numbers of targets always face a termination problem. Search too long and you are perseverating. Quit too fast and you are leaving too many targets undiscovered. If the search is a search for a mate or food or cancer or a bomb in luggage, the costs of poor performance can be very great. The topic is under-researched in visual search but a basic model can be outlined on the basis of what we know (and can be implemented, at least in one incarnation in the supplement to Wolfe and Van Wert, 2010). Figure 16 gives a final summary.

We envision a search termination decision, based on how many items have been searched or how much time has been spent in search. For any given search, there will be a search threshold, expressed in time or item units. Information about time or items will accumulate toward that threshold in a noisy fashion (green arrows). The resulting distribution of RTs will be positively skewed (Palmer, Horowitz, Torralba, & Wolfe, 2011; Van Zandt, 2002). If the threshold is more liberal, observers will quit more rapidly. Factors that will move the threshold to a more conservative position would be: 1) A greater number of relevant items (larger set size, larger number of items with the correct features, etc), 2) More crowding and clutter, making it harder to get information out of the image, 3) Higher target prevalence, and 4) Higher value.

Analogous problems exist in other domains. Here, we briefly considered the relationship to patch leaving in behavioral ecology. Ideas brought in from these neighboring fields should allow us to make progress in figuring out when to quit.

Acknowledgments

Paper written for the 59th Annual Nebraska Symposium on Motivation: The Influence of Attention, Learning, and Motivation on Visual Search. Supported by ONR: N000141010278, NIH: EY017001 and DHS: 02-G-010

References

Bacon WF, Egeth HE. Overriding stimulus-driven attentional capture. Perception and Psychophysics. 1994; 55(5):485–496. [PubMed: 8008550]

Berbaum KS, Franken EA Jr, Dorfman DD, Caldwell RT, Krupinski EA. Role of faulty decision making in the satisfaction of search effect in chest radiography. Academic Radiology. 2000; 7(12): 1098–1106. [PubMed: 11131054]

Berbaum KS, Franken EA Jr, Dorfman DD, Rooholamini SA, Kathol MH, Barloon TJ, et al. Satisfaction of search in diagnostic radiology. Invest Radiol. 1990; 25(2):133–140. [PubMed: 2312249]

Bravo MJ, Farid H. Search for a category target in clutter. Perception. 2004; 33(5):643–652. [PubMed: 15330362]

Bravo MJ, Farid H. A scale invariant measure of clutter. Journal of Vision. 2008; 8(1):1–9. [PubMed: 18318626]

Brown SD, Heathcote A. The simplest complete model of choice response time: Linear ballistic accumulation. Cognitive Psychology. 2008; 57(3):153–178. [PubMed: 18243170]

Buschman TJ, Miller EK. Serial, Covert Shifts of Attention during Visual Search Are Reflected by the Frontal Eye Fields and Correlated with Population Oscillations. Neuron. 2009; 63:386–396. [PubMed: 19679077]

Charnov EL. Optimal foraging, the marginal value theorem. Theoretical Population Biology. 1976; 9:129–136. [PubMed: 1273796]

Chun MM, Wolfe JM. Just say no: How are visual searches terminated when there is no target present? Cognitive Psychology. 1996; 30:39–78. [PubMed: 8635311]

Colquhoun WP, Baddeley AD. Influence of signal probability during pretraining on vigilance decrement. J Exp Psychol. 1967; 73(1):153–155. [PubMed: 6031659]

Dehaene, S. The number sense. New York, Cambridge (UK): Oxford University Press, Penguin press; 1997.

Dodd MD, Castel AD, Pratt J. Inhibition of return with rapid serial shifts of attention: Implications for memory and visual search. Perception & Psychophysics. 2003; 65(7):1126–1135. [PubMed: 14674638]

Donkin C, Brown S, Heathcote A, Wagenmakers EJ. Diffusion versus linear ballistic accumulation: different models but the same conclusions about psychological processes? Psychonomic Bulletin & Review. 2011; 18(1):61–69. [PubMed: 21327360]

Dukas R. Behavioural and ecological consequences of limited attention. Philos Trans R Soc Lond B Biol Sci. 2002; 357(1427):1539–1547. [PubMed: 12495511]

Dukas R. Causes and consequences of limited attention. Brain Behav Evol. 2004; 63(4):197–210. [PubMed: 15084813]

Duncan J, Humphreys GW. Visual search and stimulus similarity. Psychological Review. 1989; 96:433–458. [PubMed: 2756067]

Egeth HE, Virzi RA, Garbart H. Searching for conjunctively defined targets. J Exp Psychol: Human Perception and Performance. 1984; 10:32–39.

Einhauser W, Spain M, Perona P. Objects predict fixations better than early saliency. Journal of Vision. 2008; 8(14):1–26.

Ethell, SC.; Manning, D. Effects of prevalence on visual search and decision making in fracture detection. In: Krupinski, EA.; Chakraborty, DP., editors. Image Perception and Performance, Proceedings of SPIE. Vol. 4324. SPIE; 2001. p. 249-257.

Felisberti FM, Solomon JA, Morgan MJ. The role of target salience in crowding. [Article]. Perception. 2005; 34(7):823–833. [PubMed: 16124268]

Fleck MS, Mitroff SR. Rare targets are rarely missed in correctable search. Psychol Sci. 2007; 18(11): 943–947. [PubMed: 17958706]

Fleck MS, Mitroff SR. Rare targets rarely missed in correctable search. Psychol Sci. 2007; 18(11): 943–947. [PubMed: 17958706]

Fleck MS, Samei E, Mitroff SR. Generalized “Satisfaction of Search”: Adverse Influences on Dual-Target Search Accuracy. Journal of Experimental Psychology: Applied. 2010; 16(1):60–71.10.1037/a0018629 [PubMed: 20350044]

Foulsham T, Underwood G. What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. J Vis. 2008; 8(2):6, 1–17. [PubMed: 18318632]

Franconeri SL, Hollingworth A, Simons DJ. Do new objects capture attention? Psychol Sci. 2005; 16(4):275–281. [PubMed: 15828974]

Friedman-Hill SR, Wolfe JM. Second-order parallel processing: Visual search for the odd item in a subset. J Experimental Psychology: Human Perception and Performance. 1995; 21(3):531–551.

Gilchrist ID, Harvey M. Evidence for a systematic component within scanpaths in visual search. Visual Cognition. 2006; 14(5–7)

Godwin HJ, Menneer T, Cave KR, Helman S, Way RL, Donnelly N. The impact of Relative Prevalence on dual-target search for threat items from airport X-ray screening. Acta Psychologica. 2010; 134(1):79–84.10.1016/j.actpsy.2009.12.009 [PubMed: 20129597]

Goldsmith M. What’s in a location? Comparing object-based and space-based models of feature integration in visual search. J Experimental Psychology: General. 1998; 127(2):189–219.

Greene MR, Oliva A. The briefest of glances: the time course of natural scene understanding. Psychol Sci. 2009; 20(4):464–472. [PubMed: 19399976]

Gur D, Rockette HE, Armfield DR, Blachar A, Bogan JK, Brancatelli G, et al. Prevalence effect in a laboratory environment. Radiology. 2003; 228(1):10–14. [PubMed: 12832568]

Hamid SN, Stankiewicz B, Hayhoe M. Gaze patterns in navigation: encoding information in large-scale environments. J Vis. 2010; 10(12):28. [PubMed: 21047760]

Hamker F. Modeling feature-based attention as an active top-down inference process. Biosystems. 2006; 86(1–3):91–99. [PubMed: 16843589]

Healy AF, Kubovy M. Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. J Exp Psychol: Human Learning and Memory. 1981; 7(5): 344–354.

Hearns JF, Moss SM. EFFECT OF METHOD OF PAYOFF ON DETECTION OF TARGETS IN A VISUAL SEARCH TASK. Journal of Experimental Psychology. 1968; 78(4P1):569. [PubMed: 5705867]

Henderson, JM.; Ferreira, F. Scene perception for psycholinguists. In: Henderson, JM.; Ferreira, F., editors. The interface of language, vision, and action: Eye movements and the visual world. New York: Psychology Press; 2004. p. 1-58.

Hick WE. On the rate of gain of information. Quart J Experimental Psychology. 1952; 4:11–26.

Hickey C, Chelazzi L, Theeuwes J. Reward has a residual impact on target selection in visual search, but not on the suppression of distractors. Visual Cognition. 2011; 19(1):117–128.

Hickey C, Theeuwes J. Reward primes visual search. Perception. 2008; 37:46–46.

Horowitz TS, Wolfe JM. Visual search has no memory. Nature. 1998; 394(Aug 6):575–577. [PubMed: 9707117]

Horowitz TS, Wolfe JM. Memory for rejected distractors in visual search? Visual Cognition. 2003; 10(3):257–298.

Hyman R. Stimulus information as a determinant of reaction time. J Exp Psychol. 1953; 45(3):188–196. [PubMed: 13052851]

Jovancevic-Misic J, Hayhoe M. Adaptive gaze control in natural environments. J Neurosci. 2009; 29(19):6234–6238. [PubMed: 19439601]

Klein R. Inhibitory tagging system facilitates visual search. Nature. 1988; 334:430–431. [PubMed: 3405288]

Klein RM, MacInnes WJ. Inhibition of return is a foraging facilitator in visual search. Psychological Science. 1999; 10(July):346–352.

Koch C, Ullman S. Shifts in selective visual attention: Towards the underlying neural circuitry. Human Neurobiology. 1985; 4:219–227. [PubMed: 3836989]

Kristjansson A. In search of rememberance: Evidence for memory in visual search. Psychological Science. 2000; 11(4):328–332. [PubMed: 11273394]

Kristjansson A, Sigurjonsdottir O, Driver J. Fortune and reversals of fortune in visual search: Reward contingencies for pop-out targets affect search efficiency and target repetition effects. Atten Percept Psychophys. 2010; 72(5):1229–1236. [PubMed: 20601703]

Krueger LE. Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. Perception & Psychophysics. 1984; 35(6):536–542. [PubMed: 6483555]

Kunar MA, Rich AN, Wolfe JM. Spatial and temporal separation fails to counteract the effects of low prevalence in visual search. Visual Cognition. 2010; 18(6):881–897. [PubMed: 21442052]

Kundel HL. Disease prevalence and radiological decision making. Invest Radiol. 1982; 17(1):107–109. [PubMed: 7076429]

Kundel, HL. Disease prevalence and the index of detectability: a survey of studies of lung cancer detection by chest radiography. In: Krupinski, EA., editor. Medical Imaging 2000: Image Perception and Performance. Vol. 3981. 2000. p. 135-144.

Lau JS, Huang L. The prevalence effect is determined by past experience, not future prospects. Vision Res. 2010; 50(15):1469–1474. [PubMed: 20438744]

Levi DM. Crowding-An essential bottleneck for object recognition: A mini-review. Vision Res. 2008; 48(5):635–654. [PubMed: 18226828]

Logan GD. The CODE theory of visual attention: An integration of space-based and object-based attention. Psychological Review. 1996; 103(4):603–649. [PubMed: 8888649]

Mackworth, J. Vigilance and Attention. Harmondsworth, England: Penguin Books; 1970.

Maddox WT. Toward a unified theory of decision criterion learning in perceptual categorization. J Exp Anal Behav. 2002; 78(3):567–595. [PubMed: 12507020]

Maljkovic V, Martini P. Implicit short-term memory and event frequency effects in visual search. Vision Res. 2005; 45(21):2831–2846. [PubMed: 16023172]

Maljkovic V, Nakayama K. Priming of popout: I. Role of features. Memory & Cognition. 1994; 22(6): 657–672. [PubMed: 7808275]

Masciocchi CM, Mihalas S, Parkhurst D, Niebur E. Everyone knows what is interesting: Salient locations which should be fixated. Journal of Vision. 2009; 9(11):1–22. [PubMed: 20053088]

Nakayama K, Silverman GH. Serial and parallel processing of visual feature conjunctions. Nature. 1986; 320:264–265. [PubMed: 3960106]

Navalpakkam V, Koch C, Perona P. Homo economicus in visual search. Journal of Vision. 2009; 9(1)

Neider MB, Zelinsky GJ. Exploring set size effects in scenes: Identifying the objects of search. Visual Cognition. 2008; 16(1):1– 10.

Nodine CF, Krupinski EA, Kundel HL, Toto L, Herman GT. Satisfaction of search (SOS). Invest Radiol. 1992; 27(7):571–573. [PubMed: 1644560]

Oliva, A. Gist of the scene. In: Itti, L.; Rees, G.; Tsotsos, J., editors. Neurobiology of attention. San Diego, CA: Academic Press/Elsevier; 2005. p. 251-257.

Palmer EM, Horowitz TS, Torralba A, Wolfe JM. What are the shapes of response time distributions in visual search? Journal of Experimental Psychology: Human Perception and Performance. 2011; 37(1):58–71. [PubMed: 21090905]

Parasuraman R, Davies DR. Decision theory analysis of response latencies in vigilance. J Exp Psychol Hum Percept Perform. 1976; 2(4):578–590. [PubMed: 1011007]

Peterson MS, Kramer AF, Wang RF, Irwin DE, McCarley JS. Visual search has memory. Psychological Science. 2001; 12(4):287–292. [PubMed: 11476094]

Pyke GH, Pulliam HR, Charnov EL. Optimal Foraging: A Selective Review of Theory and Tests. The Quarterly Review of Biology. 1977; 52(2):137–154.

Ratcliff R. A theory of memory retrieval. Psych Review. 1978; 85(2):59–108.

Rayner, K. Visual selection in reading, picture perception, and visual search. In: Bouma, H.; Bouwhuis, BG., editors. Attention and Performance X. Vol. 10. Hillside, NJ: L Erlbaum Assoc; 1983. p. 67-96.

Reynolds JH, Chelazzi L. Attentional modulation of visual processing. Annu Rev Neurosci. 2004; 27:611–647. [PubMed: 15217345]

Roelfsema PR, Lammer VAF, Spekreijse H. Object-based attention in the primary visual cortex of the macaque monkey. Nature. 1998; 395:376. [PubMed: 9759726]

Rosenholtz R, Chan S, Balas B. A crowded model of visual search. Journal of Vision. 2009; 9(8): 1197–1197.

Sagi D. The combination of spatial frequency and orientation is effortlessly perceived. Perception and Psychophysics. 1988; 43:601–603. [PubMed: 3399359]

Sherman AM, Greene MR, Wolfe JM. Depth and Size Information Reduce Effective Set Size for Visual Search in Real-World Scenes. J of Vision, VSS abstracts. 2011

Shore DI, Klein RM. On the manifestations of memory in visual search. Spat Vis. 2000; 14(1):59–75. [PubMed: 11334182]

Stephens, DW.; Krebs, JR. Foraging Theory. Princeton, NJ: Princeton U. Press; 1986.

Sternberg S. High-speed scanning in human memory. Science. 1966; 153:652–654. [PubMed: 5939936]

Swets JA, Kristofferson AB. Attention. Annu Rev Psychol. 1970; 21:339–366. [PubMed: 4904374]

Theeuwes J. Abrupt luminance change pops out; abrupt color change does not. Perception & Psychophysics. 1995; 57(5):637–644. [PubMed: 7644324]

Theeuwes J. Top-down and bottom-up control of visual selection. Acta Psychologica. 2010; 135(2): 77–99.10.1016/j.actpsy.2010.02.006 [PubMed: 20507828]

Theeuwes J, Kooi JL. Parallel search for a conjunction of shape and contrast polarity. Vision Research. 1994; 34(22):3013–3016. [PubMed: 7975336]

Torralba A, Oliva A, Castelhano MS, Henderson JM. Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search. Psychological Review. 2006; 113(4):766–786. [PubMed: 17014302]

Treisman A, Gelade G. A feature-integration theory of attention. Cognitive Psychology. 1980; 12:97–136. [PubMed: 7351125]

Trick LM, Pylyshyn ZW. Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. Psych Review. 1994; 101(1):80–102.

Tsotsos, J. A Computational Perspective on Visual Attention. Cambridge, MA: MIT Press; 2011.

Van Zandt, T. Analysis of response time distributions. In: Pashler, H.; Wixted, J., editors. Stevens' handbook of experimental psychology (3rd ed.), Vol. 4: Methodology in experimental psychology. New York, NY: John Wiley & Sons, Inc; 2002. p. 461-516.

VanWert MJ, Wolfe JM, Horowitz TS. Even in correctable search, some types of rare targets are frequently missed. Attention, Perception & Psychophysics. 2009; 71(3):541–553.

Vlaskamp BN, Hooge IT. Crowding degrades saccadic search performance. Vision Res. 2006; 46(3): 417–425. [PubMed: 15893785]

Vo MLH, Henderson JM. The time course of initial scene processing for eye movement guidance in natural scene search. Journal of Vision. 2010; 10(3):1–13. [PubMed: 20377291]

Vo MLH, Henderson JM. Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. Journal of Vision. 2009; 9(3):1–15. [PubMed: 19757963]

von Muhlenen A, Muller HJ, Muller D. Sit-and-wait strategies in dynamic visual search. Psychol Sci. 2003; 14(4):309–314. [PubMed: 12807402]

Wolfe J, Alvarez G, Rosenholtz R, Oliva A, Torralba A, Kuzmova Y, et al. Search for arbitrary objects in natural scenes is remarkably efficient. Journal of Vision. 2008; 8(6):1103–1103.

Wolfe J, Horowitz T, Kenner NM, Hyle M, Vasan N. How fast can you change your mind? The speed of top-down guidance in visual search. Vision Research. 2004; 44(12):1411–1426. [PubMed: 15066400]

Wolfe JM. Guided Search 2.0: A revised model of visual search. Psychonomic Bulletin and Review. 1994; 1(2):202–238. [PubMed: 24203471]

Wolfe JM. What do 1,000,000 trials tell us about visual search? Psychological Science. 1998; 9(1):33–39.

Wolfe JM, Cave KR, Franzel SL. Guided Search: An alternative to the Feature Integration model for visual search. J Exp Psychol - Human Perception and Perf. 1989; 15:419–433.

Wolfe JM, Horowitz TS. What attributes guide the deployment of visual attention and how do they do it? Nature Reviews Neuroscience. 2004; 5(6):495–501.

Wolfe JM, Horowitz TS, Kenner NM. Rare items often missed in visual searches. Nature. 2005; 435:439–440. [PubMed: 15917795]

Wolfe, JM.; Horowitz, TS.; Palmer, EM.; Michod, KO.; VanWert, MJ. Getting in to Guided Search. In: Coltheart, V., editor. Tutorials in Visual Cognition. Hove, Sussex: Psychology Press; 2010. p. 93-120.

Wolfe JM, Horowitz TS, VanWert MJ, Kenner NM, Place SS, Kibbi N. Low target prevalence is a stubborn source of errors in visual search tasks. JEP: General. 2007; 136(4):623–638.

Wolfe JM, Palmer EM, Horowitz TS. Reaction time distributions constrain models of visual search. Vision Res. 2010; 50(14):1304–1311. [PubMed: 19895828]

Wolfe JM, VanWert MJ. Varying target prevalence reveals two, dissociable decision criteria in visual search. Curr Biol. 2010; 20(2):121–124. [PubMed: 20079642]

Wolfe JM, Vo MLH, Evans KK, Greene MR. Visual search in scenes involves selective and non-selective pathways. Trends Cogn Sci. 2011; 15(2):77–84. [PubMed: 21227734]

Yantis S, Jonides J. Attentional capture by abrupt onsets: New perceptual objects or visual masking. J Experimental Psychology: Human Perception and Performance. 1996; 22(6):1505–1513.

Yeari M, Goldsmith M. Is object-based attention mandatory? Strategic control over mode of attention. Journal of Experimental Psychology: Human Perception and Performance. 2010; 36(3):565–579. [PubMed: 20515189]

Zohary E, Hochstein S. How serial is serial processing in vision? Perception. 1989; 18:191–200. [PubMed: 2771604]

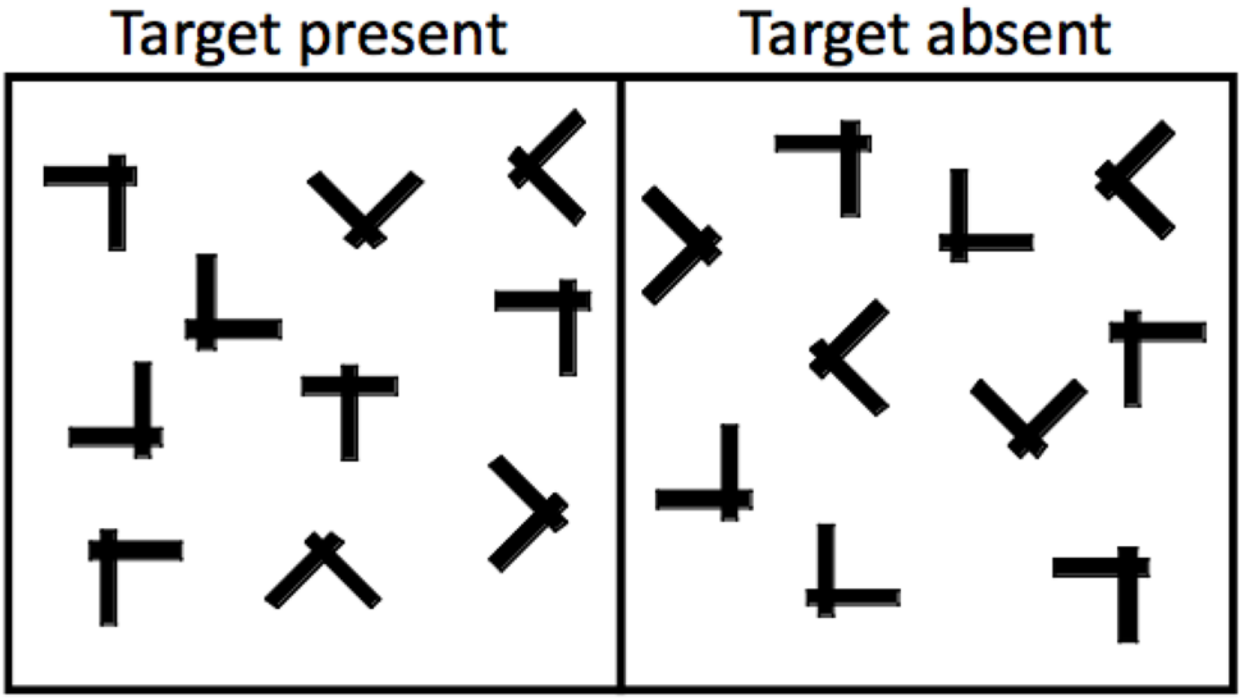


Figure 1.
A classic visual search for a T among Ls.

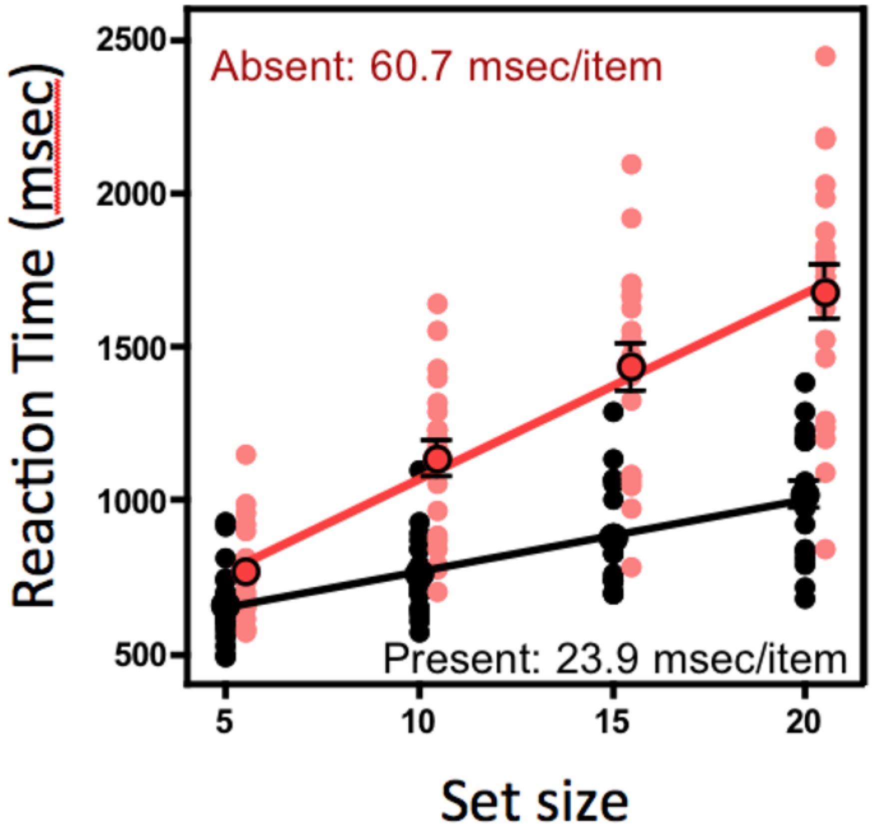


Figure 2.
Data from 20 observers performing a search for a T among Ls. Dark black spots show average correct target present RTs for each observer. Light red dots show correct target absent averages. Larger symbols are group averages. Error bars are +/− 1 s.e.m. Lines are best fit regression lines through the average points.

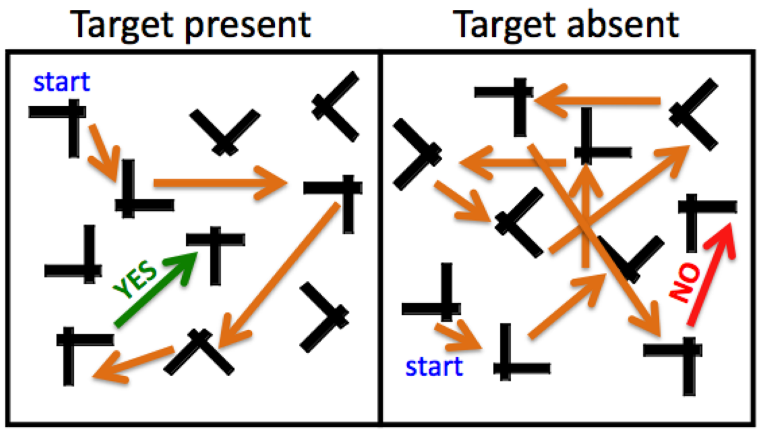


Figure 3.
Cartooned deployments of attention in a serial, self-terminating search.

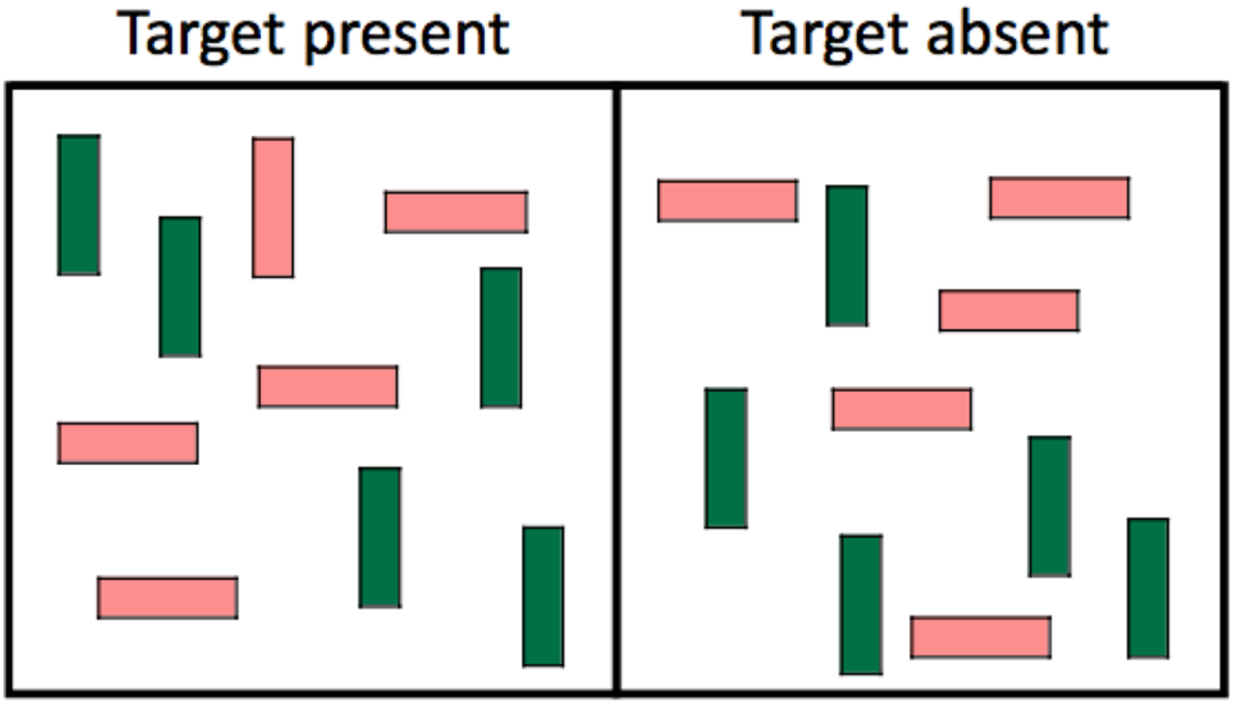


Figure 4.
A conjunction search for the light red vertical item among light red horizontal and dark green vertical distractors.

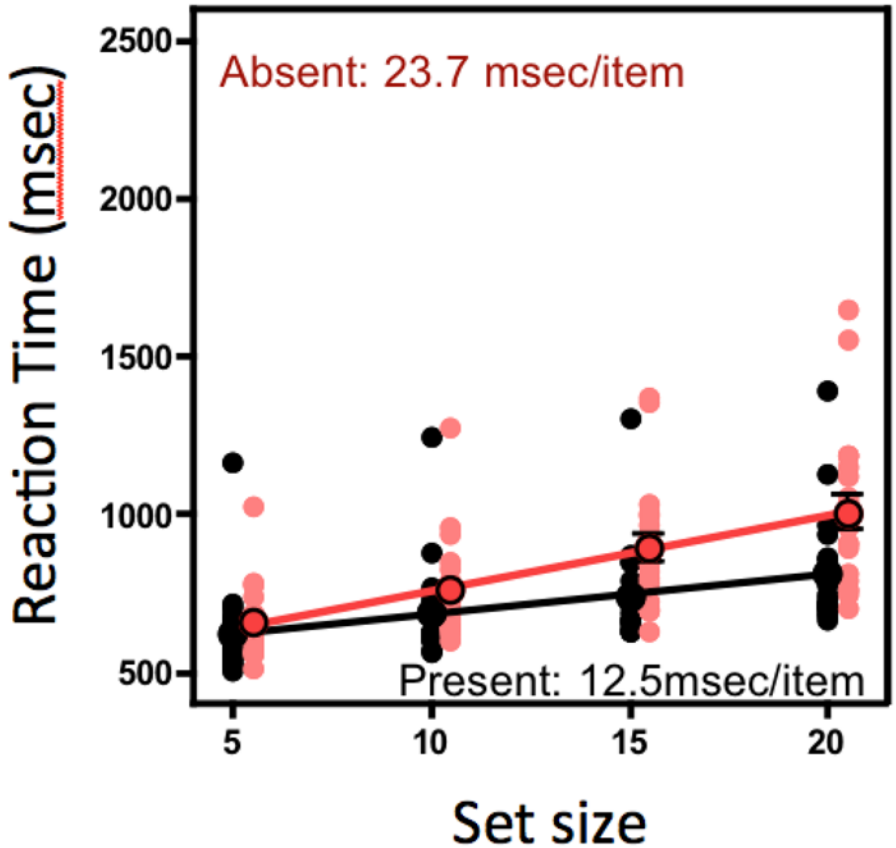


Figure 5.
RT x set size data for a color x orientation conjunction search. Each dot represents the average RT for one observer at one set size. Light red dots show absent trials. Black dots show target present trials. Larger symbols are group averages. Error bars are +/− 1 s.e.m. Lines are best fit regressions for the average points. The scale is the same as in Figure 2 for purposes of comparison.

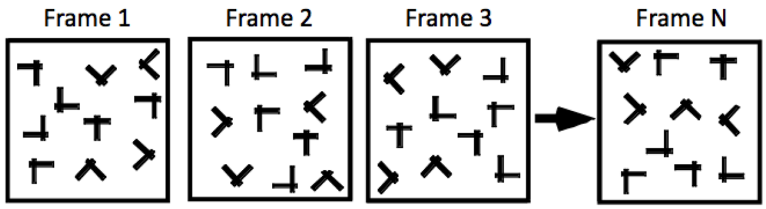


Figure 6.
Dynamic Search: All items are randomly replotted on each frame (every 100 ms). A target, if present, is present on every frame.

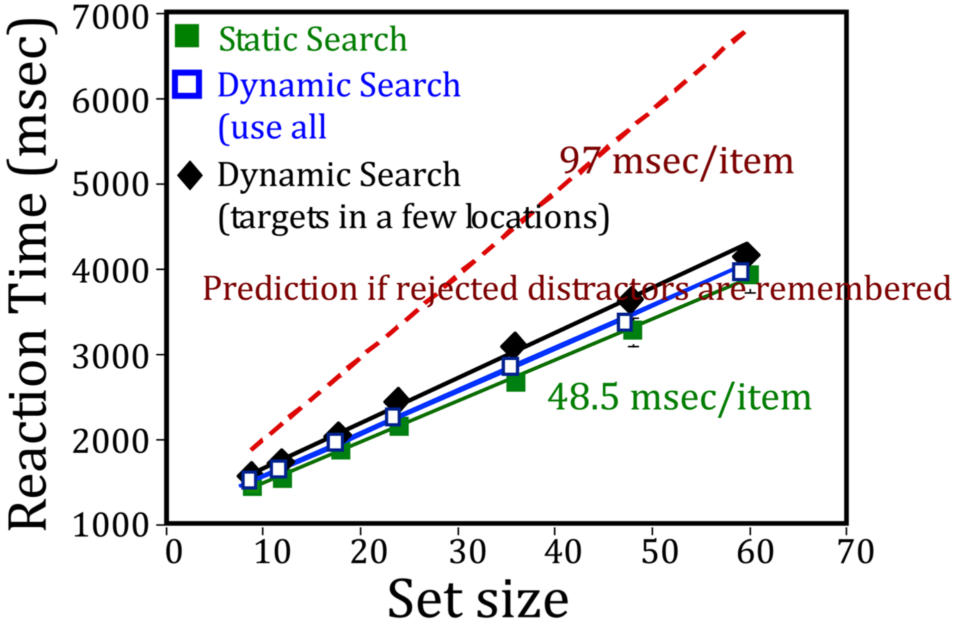


Figure 7.

Dynamic search produces similar results to static search. Green, filled squares show standard static search. Blue, open squares show dynamic search with random replotting of items. Black filled diamonds show a version of dynamic search with targets constrained to appear in a few locations. The red dashed line shows the predicted dynamic slope if static search has full memory for rejected distractors. Replotted from Horowitz and Wolfe (2003).

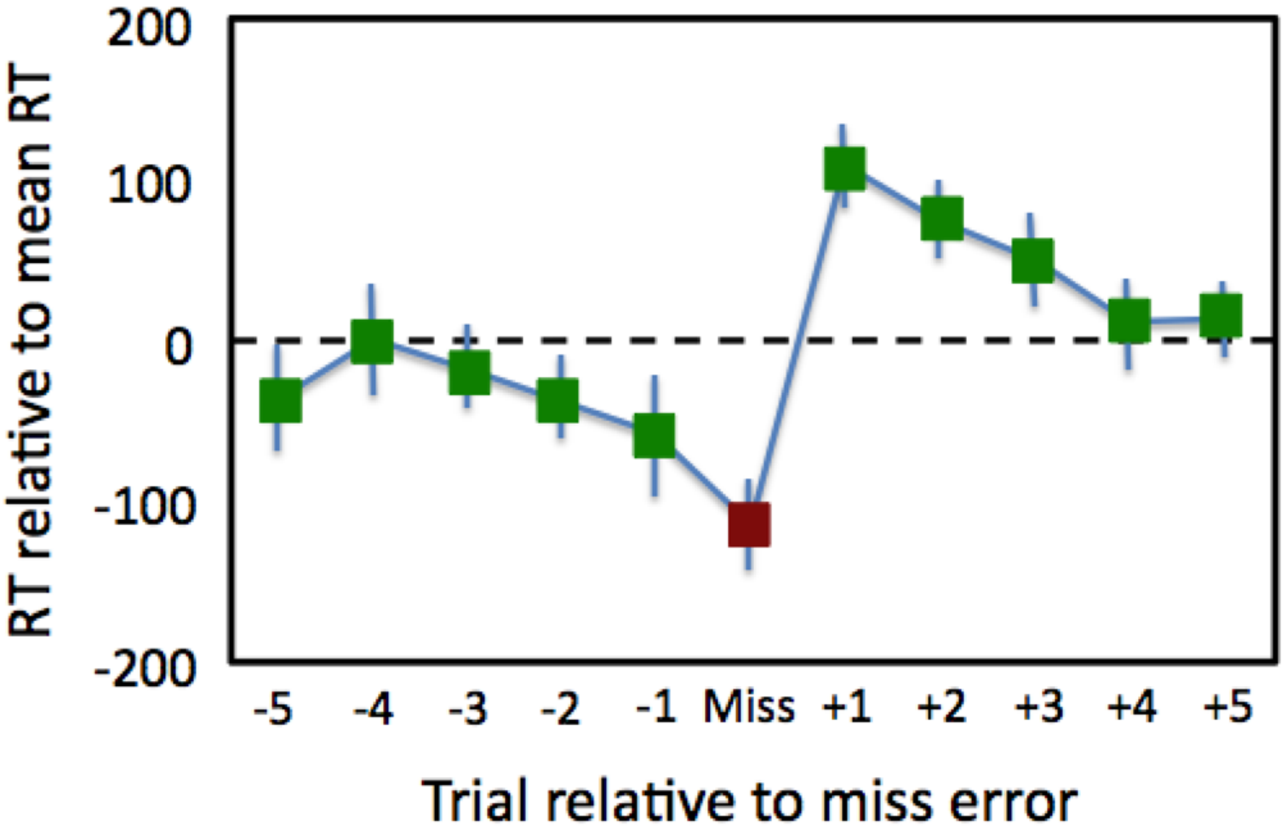


Figure 8.
Change RT relative to mean RT in a triple conjunction task, plotted with trials aligned to Miss errors.

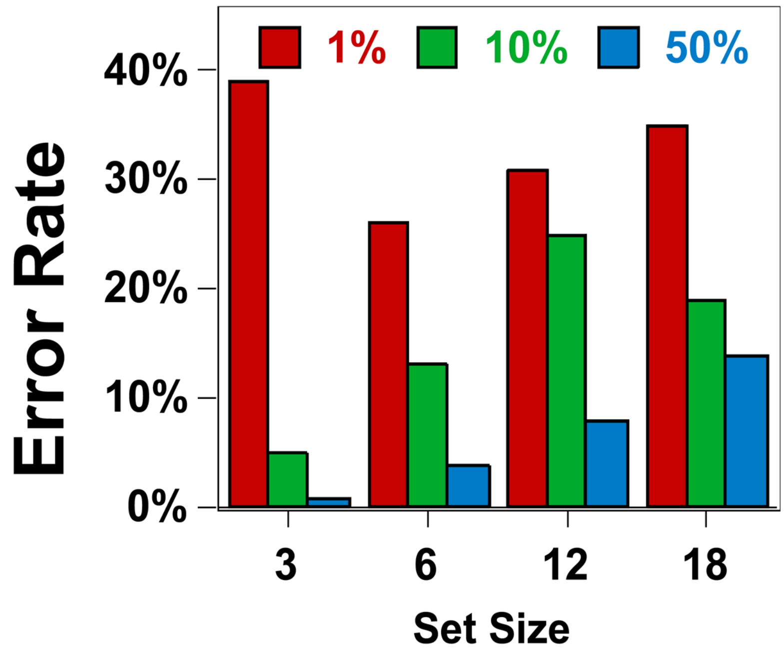


Figure 9.
Miss error rates as a function of target prevalence and set size (results redrawn from Wolfe et al., Nature, 2005).

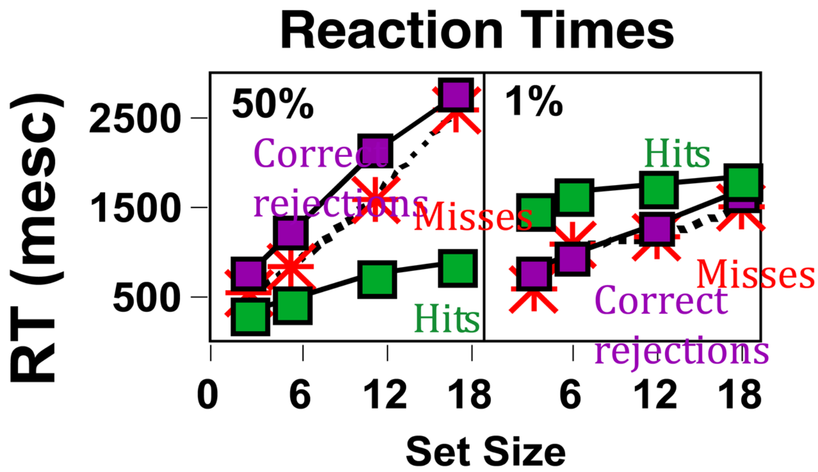


Figure 10.
RT x set size functions for 50% and 1% prevalence for the data shown in Figure 9. Green squares show “hit” RTs, purple-Correct absent trials, and red asterisks show miss error RTs.

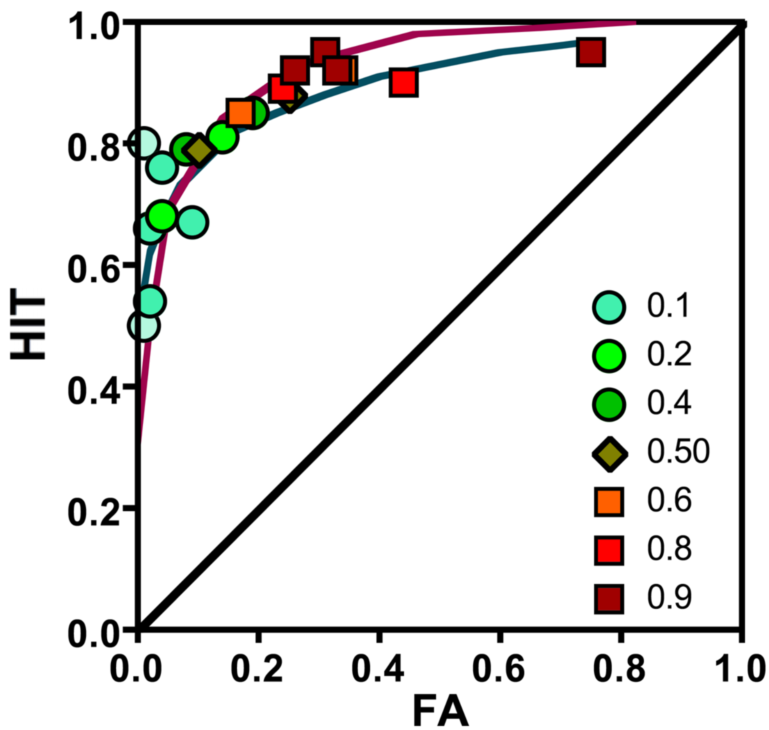


Figure 11.

Receiver operating characterstic curve for an experiment with variable prevalence. Green circles indicate lower prevalence (<0.5). Red squares indicate higher prevalence (> 0.5). Data are taken from Wolfe and VanWert, 2010. Red ROC assumes an equal variance. Blue ROC assumes unequal variance with a zROC slope of 0.6 (see text).

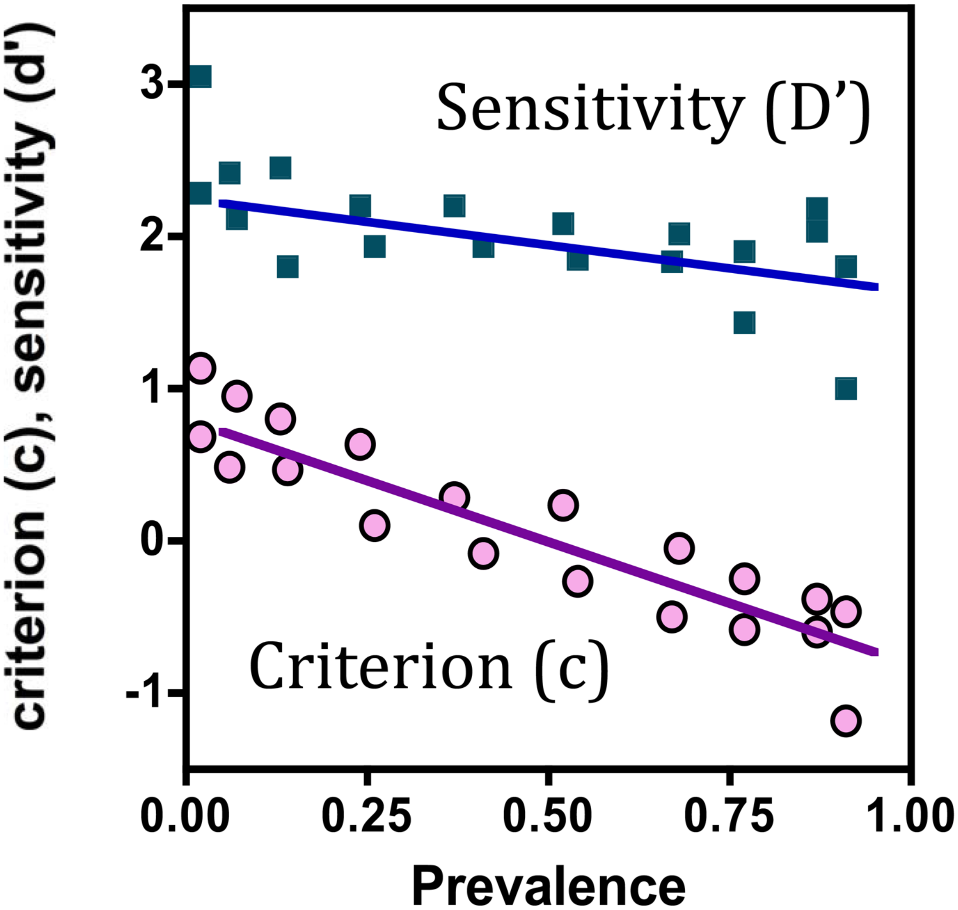


Figure 12.
Change in sensitivity (d′) and criterion (c) as a function of prevalence in data taken from Wolfe and VanWert (2010).

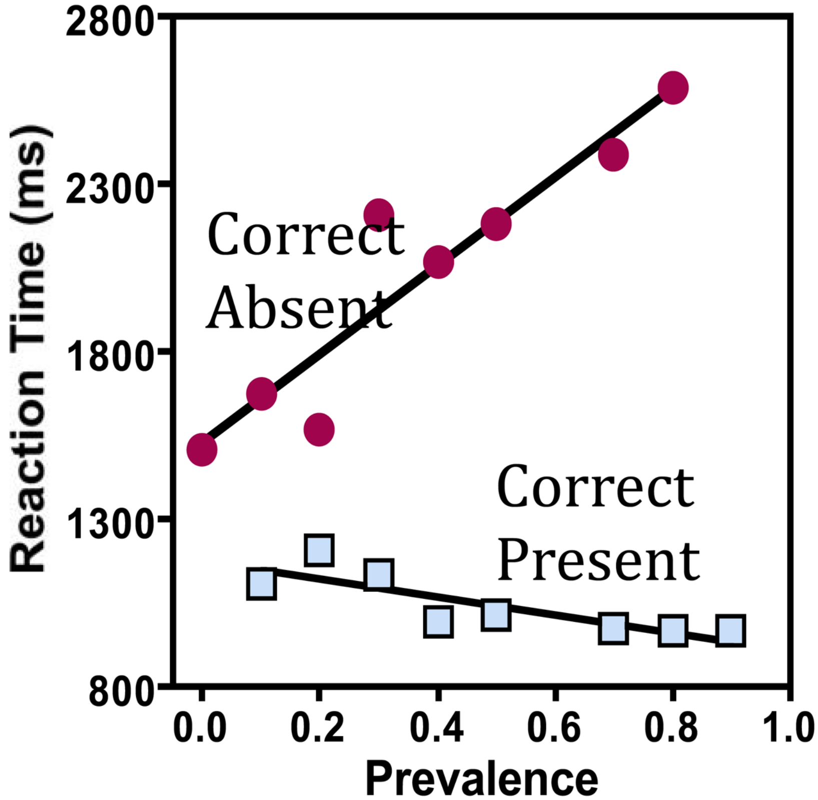


Figure 13.
Reaction time as a function of prevalence. Average data derived from Wolfe and VanWert (2010).

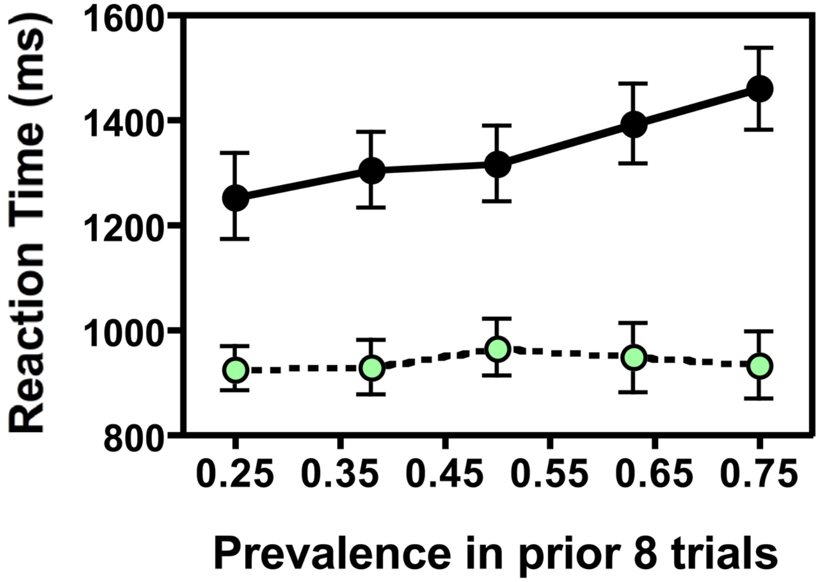


Figure 14.
RT as a function of local prevalence; in this case, the prevalence as calculated from the preceding 8 trials.

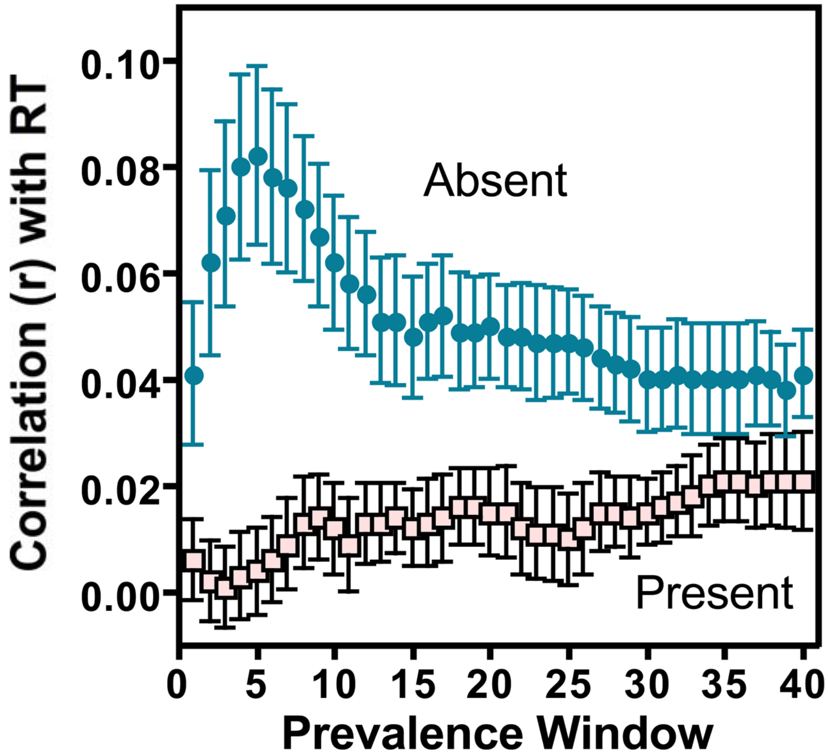


Figure 15.
Correlation of the current RT with the prevalence over the previous N trials (the "window size").

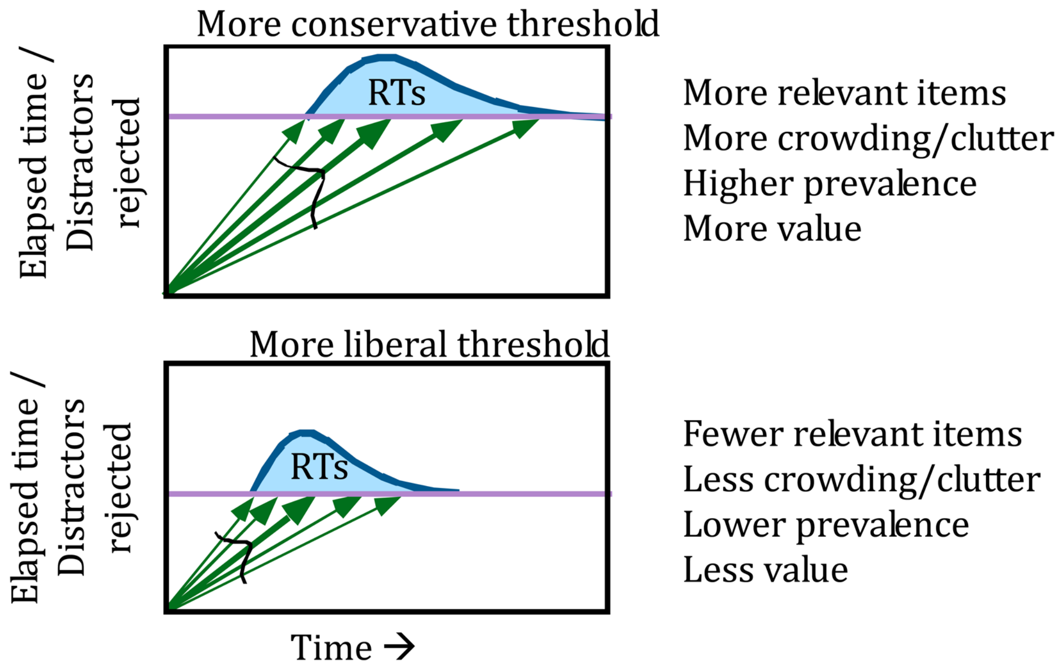


Figure 16.

Factors constraining search termination.