Visual Search as a Combination of Automatic and Attentive Processes

Chris Donkin (cdonkin@indiana.edu)
Rich Shiffrin (shiffrin@indiana.edu)
Department of Psychological & Brain Sciences, 1101 E. 10th St.
Bloomington, IN 47408 USA

Abstract

We present a model in which visual search behavior is assumed to result from a combination of controlled, serial search and automatic attraction of attention to target stimuli. The model provides a quantitative framework for how these different processes are combined, and despite a large number of constraints, it is highly successful in accounting for human search behavior at the level of full response time distributions and choice probabilities.

Keywords: visual search; computational modeling; automatic and attention processes; response times.

Visual search tasks usually require an observer to determine whether or not a pre-defined target object is present within a display of objects (the display size, D, is the number of objects in the display, and displays either contain all foils, or instead one target with the rest foils). Performance in such tasks is usually measured by response time, because accuracy tends to be quite high. The results have been used to understand the processes of visual search, the factors that determine attention allocation, and the use of automatic and parallel vs. controlled and serial processes of comparison. The last of these is the focus of this research.

Various factors contribute to whether search is automatic or controlled. Low-level perceptual differences between targets and foils, for instance a green target amidst red foils, facilitate automatic search behavior, produce response times that do not vary (much) with display size, and are often termed to 'popout' in line with subjective impressions (***Triesman & Gelade, 1980***). Even when targets do not 'popout' perceptually, consistent training (in which targets remain targets, and foils remain foils) in most cases gradually causes the targets to attract attention automatically, measured by the fact that the dependence on display size drops (Shiffrin & Schneider, 1977). However, the amount of such learning is dependent on the relation of the target shapes to foil shapes: For example, when a conjunction of features is needed to define targets, search generally appears more controlled (slower search rates) and automatic attraction of attention to targets is slower to develop.

When the plot of mean response time, RT, to D has a slope near zero, search is considered parallel and automatic; when it has a large slope (usually roughly linear) search is often assumed to be serial (one item in the display at a time), or controlled. If the slope of target absent responses is about twice that for target present responses search is usually assumed to terminate once a target is found in the serial set of comparisons.

Townsend and colleagues have argued convincingly, however, that analyses based on mean response times in standard visual search are relatively uninformative regarding the processes underlying search (e.g. Townsend & Nozawa, 1995). For example the function relating mean RT to D is insufficient for distinguishing serial from parallel search without strong additional assumptions. As we shall see, much more can be learned about visual search through analysis of full response time distributions.

Complementing research identifying conditions in which search might be either automatic or serial (Thornton & Gilden, 2007), is research developing a unified model which can account for all search behavior (e.g., Wolfe's guided search model, 2007). Our aim, in what follows, is to build upon these efforts and develop and implement a framework for how automatic and controlled search processes combine to produce the various types of observed visual search behavior. The model is fit to results partially reported in Cousineau and Shiffrin (2004), in which three participants received up to 80 sessions of training. It is similar to Wolfe's Guided Search theory in a number of respects, including a serial search process that is guided to target items by an automatic parallel process.

Model

When a display appears a set of consecutive serial comparisons without replacement is initiated. The order of comparisons is chosen by the observer, and is random with respect to the actual target position. The order of comparisons can be interrupted, however, when search is guided by a separate parallel process that forces the next comparison to a target position. Search terminates when a comparison to a target occurs with a positive (i.e. target present) response, or with a negative response (i.e. target absent) when all display positions are compared unsuccessfully. As we will discuss later, however, a separate decision is sometimes made to terminate before all comparisons are finished (i.e. early termination of search).

The Serial Comparison Process

Each comparison involves a decision as to whether an item in the display is either a target or a foil. We model this decision using a relatively simple evidence accumulation model, based on the Linear Ballistic Accumulator (LBA) model (Brown & Heathcote, 2008). Figure 1 contains a graphical depiction of the comparison process. We assume that evidence is accumulated for "target" or "foil" responses in separate, independent, accumulators. When a target is being compared evidence tends to accumulate quickly in the target accumulator and slowly in the foil accumulator. When a foil is being compared, the tendency reverses. Each accumulator has a threshold and the output of the comparison and its time is determined by the accumulator that reaches its threshold first (each comparison is therefore a race). It should be noted that this LBA model of a comparison closely mimics predictions of common alternative models for a comparison, such as the diffusion model of Ratcliff (1978). We use the LBA for reasons of computational convenience.

According to the LBA, evidence accumulates ballistically in each accumulator (i.e., without noise) at a linear rate until either threshold is reached. However, the rate at which each accumulator proceeds is chosen from a normal distribution, independently for the two accumulators. The winner of the race, and the time of the comparison, is therefore determined by the rate choices for that comparison (which depend on whether the comparison is of a target or foil) and the thresholds for each accumulator.

For simplicity we use only two Gaussians to determine the rates, each with a to-be-estimated variance, s. The target accumulator when a target is being compared, and the foil accumulator when a foil is being compared (i.e. the accumulators that would produce a correct response) have rates chosen from separate Gaussians with to-be-estimated means, v_T and v_F, respectively; the two remaining accumulators (those that would produce an incorrect response) have rates chosen from a Gaussian with fixed means, $1 - v_T$ and $1 - v_F$, for target and foil comparisons respectively. We assume that the threshold for the target and foil accumulators can be different, and thus estimate two threshold parameters, b_T and b_F.

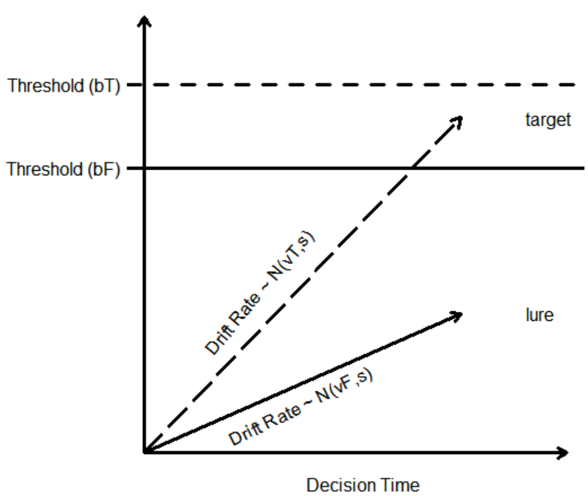

Figure 1 An LBA comparison process when a target is being compared (so the mean drift rate is higher for the target accumulator). The target accumulator and its threshold are given by the dashed lines, and the foil accumulator and its threshold are given by the solid lines. The drift rates shown are samples from normal distributions with the indicated means and common variance s. These sampled rates will cause the target accumulator to reach threshold first, so a correct result of the comparison will occur.

This gives a total of five parameters for the LBA comparison process. Looking ahead, let us note that in our application to the empirical data reported in Cousineau and Shiffrin (2004) all five are chosen to provide a best fit to the response time distributions and accuracy when display size is one, for each of the three participants. We are assuming, therefore, that comparisons are equivalent, and driven by the same parameters, regardless of the number of items in the display size.

If the target accumulator wins the race then the search ends and a positive response is initiated. If the foil accumulator wins the race, then if all display items have been compared a negative response if initiated. If there are more display items to be compared, then there are three possibilities: 1) Another comparison is initiated; the item to be compared is chosen by the participant, and is a random choice from the items that have not yet been compared. 2) Another comparison is initiated; the item to be compared is determined by an automatic parallel process, whose details will be given shortly. 3) The search is terminated early with a negative response, according to a process whose details will be given shortly. Note that even the selection of the first item to be compared can be driven by the parallel guiding process.

The total response time is a sum of the time occupied by the above serial comparison process, plus a taken for non-decision elements of making a response, such as the time taken to encode the stimuli and execute the motor response. The non decision time assumptions are given below.

The Parallel Guidance Process

There is considerable evidence in the literature to suggest that a purely serial model will often fail to account for visual search performance. For example, the model fails to predict that the time to find the target does not depend on display size when there is perceptual pop-out (e.g., Treisman & Gelade, 1980). Even when perceptual popout does not occur, a purely serial process can fail. This is the case for the data reported in Cousineau and Shiffrin (2004), where the purely serial model failed in several respects; most notably, participants responded much more quickly than would be expected from a purely serial self-terminating search. We note that positive responses exhibited multi-modal response time distributions, with the modes roughly corresponding to the serial position in which the target happened to be compared. However, the positions of the modes showed that the participants were comparing the target earlier in the serial search than would have been the case had each comparison been chosen randomly. Furthermore, these fast responses were clearly not guesses because their accuracy was essentially perfect.

To explain these faster than expected and accurate target present responses the model includes an automatic parallel process that guides search to the target position in a display, and is learned as training proceeds. In line with the research and model of Shiffrin and Schneider (1977), and given the fact that the Cousineau and Shiffrin (2004) study used consistent mapping in which targets remained targets and foils remained foils throughout training, we assume that targets come to attract attention automatically, and that this process operates in parallel across the entire display.

There are several plausible ways to operationalize a parallel process. For example, it could be allowed to race with the serial comparison process, with a target decision made when either process discovers a target. We decided instead to let the parallel process guide the next comparison of the serial process. The parallel process is initially weak and gradually strengthens over training. We thus think of it as unreliable from trial to trial, so that its tentative location of a target needs to be checked by a direct comparison. For simplicity this parallel guiding process is implemented as a single accumulator, which gathers evidence in parallel from the entire display concerning the presence and location of a target. If a target is present then the rate of evidence collected (which is again assumed to be linear and without noise) is selected from a normal distribution with mean v_{PAR} and standard deviation s_{PAR}. The accumulator is assumed to collect evidence for some time t_{PAR} before the first item is selected for comparison. If evidence is already at threshold b_{PAR} at this time, then the first comparison is guided, otherwise search continues as per usual and an item is selected at random (according to the observer's plan). When evidence in the parallel process does reach threshold, then the current serial comparison is allowed to finish, but if that comparison does not lead to a response, then the next comparison is guided to the position identified (correctly) by the parallel process.

The rate of accumulation in the parallel process should of course increase as training proceeds and automatic attention attraction to targets is learned. Also, although the Cousineau and Shiffrin (2004) study did not vary the target and foil stimulus properties, we note that the rate of accumulation parameter in the parallel guiding process should in principle vary with factors such as the perceptual similarity between targets and lures. Indeed, we find this is true when we fit response time distribution data reported by Wolfe, Horiwitz & Palmer (2010), in which the perceptual features of stimuli were manipulated. Unfortunately, we are unable to present these results here, but they will be available in a larger manuscript currently in preparation for submission.

Processes that Produce Rapid Negative Responses

We found that the model as described thus far could predict very well the distributions and accuracy levels for positive responses. However this model failed badly when applied to the negative response distributions (we note that the negative response times have proved a problem for other investigators, e.g., Thornton & Gilden, 2007; Wolfe, 2007). The main problem is its prediction of too large a proportion of very slow negative responses because these only occur when all comparisons fail. We believe it reasonable that the search process can be terminated with a negative response before all comparisons are completed. To take an extreme example, suppose there is a very large display (as in 'Find Waldo'). If say, 967 comparisons out of 1000 have been completed without finding a target, it would be reasonable to stop searching and respond negatively, because such a response would likely be correct. This reasoning is especially enhanced if a parallel guiding process is operating: the fact that this process has not yet reached threshold provides additional evidence that a target is not present. Cutting search short is also likely for participants who are motivated to finish the experimental session as rapidly as possible. With these factors in mind, we explored two possibilities for the way in which participants may terminate their search early – collapsing thresholds and early terminations.

Collapsing Thresholds

The first possibility is based on Thorton and Gilden's (2007) implementation of early terminations in a purely serial search model; that observers do search exhaustively through all items in the display (until a target is located) but that the threshold for responding negatively decreases as the number of items compared increases. In other words, we allow b_F to get smaller as comparisons continue: The response threshold for the target absent accumulator for the jth item compared, where $j>1$, was $b_F - (\Delta b_{MAX} - \Delta b_{ITEM}(D - j))$. Such an assumption has two effects: 1) It reduces the average amount of time required to respond negatively especially by decreasing the longest response times. 2) It decreases the variability in the predicted response time distributions, again because the longest response times tend to be eliminated.

Early Terminations

The second possibility is an early termination decision not tied to a particular evidence collection process, but instead to the display size and the number of comparisons or time taken searching without success. We implemented this idea by assuming a probability of terminating search with a negative decision that increases with the proportion of display items thus far compared unsuccessfully. This assumption will decrease the average response times, of course, but unlike collapsing thresholds, will increase the variance (because terminations that do not depend directly on the evidence being collected adds additional variability).

In line with the thought that participants will become increasingly likely to terminate with the increasing passage of time and unsuccessful comparisons, we set the probability of terminating early to be a logistic transformation of the proportion of the display items compared unsuccessfully thus far: $p_{NO} = (1\text{-}e^{(p-\mu)/\sigma})^{-1}$, where μ and σ are the location and scale parameters of the transform, and p is the proportion of the display thus far compared. This sigmoid function gets especially large as the search nears completion, thus making the probability of terminating early very large as the proportion of items searched increases. Note that we chose to set σ to be fixed across participants, allowing only μ to vary between participants.

Non-Decision Times and Switch Times

Fits of the model as described thus far revealed two small but systematic mispredictions that we fix by adding assumptions about non-decision time variation and switch times between comparisons.

First, the data show that the very fastest responses (the leading edge of the distributions) slowed as a function of display size. We therefore assumed that non-decision time varies from trial-to-trial according to a uniform distribution with mean T_{ER} and range s_{T}., but allowed T_{ER} to increase linearly with display size, with a parameter ΔT_{ER}.

Second, the modes of the observed response time distributions were farther apart than predicted by the model as described thus far. We therefore assumed that there is an extra time required to switch from one comparison to the next, t_{SWITCH}.

Fits to the Cousineau and Shiffrin (2004) Data

Cousineau and Shiffrin (2004) reported the results of a set of standard visual search conditions that were part of a much larger study. We have fit the model to many other conditions but will show here only fits of the model to response time distributions from each of the three participants for the standard conditions in training sessions 34 to 44. We show the detailed predictions for sessions 34 to 44 because inspection showed that learning had slowed enough to allow collapsing of the data across sessions to take place without undue distortions.

Methods

More details can be found in Cousineau and Shiffrin (2004). There were four target and four foil stimuli, all composed of a circle with short line segments (spokes) pointing outwards at eight different positions around the circle. Items maintained their role as targets and lures throughout the entire study. Target items were defined by a conjunction of features (i.e., the spokes on the circles), such that at least one foil item shared at least one feature with all target items. Targets were defined by a conjunction of features so as to inhibit perceptual pop-out. The extended and consistent training should have produced what Shiffrin and Schneider (1977) termed automatic categorization by which the set of four targets comes to act as a single category that can be compared in one step (analogous to searching for a letter among numbers without checking each possible letter). Thus in applying the model we assume that there is a single target rather than four.

Each trial began with a fixation star presented in the center of the display for 1000ms. The participants were then shown, for 500ms, a set of featureless stimuli (i.e., circles without spokes) where the stimuli for that trial were to be presented, after which time the features arrived (i.e., spokes were added to the empty circles) and remained until a response was made. Display size, D, was either one, two or four, with each display size occurring equally often within a block of trials. Stimuli were presented in the four corners of an imaginary square so that the entire display viewed at 50cm was within 2° vertically and 3° horizontally. For display sizes less than four, positions in the square were chosen randomly. One of the four target items was present on 50% of the trials in a block, with the order of target present and target absent trials chosen randomly, and with the target location being chosen randomly. Feedback on the speed and the accuracy of the response was given after each trial. Each block consisted of 108 trials, and each session had 6 blocks.

Table 1 Parameter estimates for each participant. Dashes (-) indicate parameters not used. [a] Indicates parameters whose units are seconds. [b] Indicates parameters whose units are per second. Other parameters have arbitrary units.

	Participant		
	A	B	C
v_T [b]	1	0.9	0.85
v_F [b]	0.95	0.92	0.82
s [b]	0.25	0.26	0.23
b_T	0.21	0.1	0.11
b_F	0.17	0.135	0.12
T_{ER} [a]	0.2	0.25	0.205
ΔT_{ER} [a]	0.02	0.025	0.03
s_T [a]	0.05	0.1	0.075
v_{PAR} [b]	-	4	7
t_{SWITCH} [a]	0.045	-	0.025
μ	-	0.57	0.46
Δb_{MAX}	0.075	-	-
Δb_{ITEM}	0.0325	-	-

Results and Model Fits

Figure 2 contains histograms of response time distributions for correct and incorrect responses (black and grey, respectively) for display sizes one, two and four (top, middle and bottom rows, respectively), for target present and target absent trials (left and right columns, respectively) for each of the three participants (A-C). The model predictions are given by the triangles

Probably most striking in the observed data are the distinct modes observed for target present trials when display size was greater than one, modes arguing strongly that search does include a serial comparison process. The procedure of identifying stimulus locations before each trial is atypical, but has the advantage of allowing the participant to plan an order of successive serial comparisons, thereby reducing some of the 'noise' that is found in typical search tasks. We believe that this method is what allowed us to see multiple modes in the target present response time distributions.

Participants B and C do not show modes as clearly as does A, and these participants appear to have been reliably guided to the correct location more often and sooner than was the case for A.

Fitting the Model

There are at most eleven freely estimated parameters per participant used to fit sessions 34-44, seven of which are associated with the comparisons to one display item in the serial process (a few other other parameters were fixed rather than estimated, as described below). The seven comparison parameters, v_T, v_L, s, b_T, b_L, T_{ER}, and s_T, were estimated from the data for display size of one (i.e., the top row of Figures 2A-2C). Fitting this way aligns with the assumption that the serial comparison process is identical for all items, regardless of the progress of the ongoing search. It also provides a great deal of constraint on the predictions of the model.

With these parameter values fixed, we next estimated the values of the residual time and switch time parameters, t_{SWITCH} and ΔT_{ER}. Finally we estimated parameters for the parallel guidance process. To minimize the number of free parameters required for this parallel process, we freely estimate only one parameter, v_{PAR}, used for all display sizes, while the other parameters were fixed at values of $b_{PAR} = .1$, $t_{PAR} = .12$ seconds and $s_{PAR} = 1.5$ per second, for all three participants.

Finally, we fit three different versions of the model to each participant – a version with only collapsing thresholds, a version with only our early termination process, and a version that included both of these processes. We found that the early termination process alone worked reasonably well for participants B and C, and that the collapsing thresholds process was required for participant A, and show predictions for these models.

The fits of the model are shown by the connected triangles in Figure 2. Model predictions were generated by simulating 20,000 trials per display size per target presence condition (i.e., present or absent). The parameters of the model were estimated by hand, in the order previously outlined, and were chosen to produce visual agreement between histograms of model predictions and observed data. Table 1 contains the parameter values that produced the predictions in the figures.

The parameter values show what is evident from inspection of the data: Participants B and C had much stronger parallel guidance processes than participant A. The rate of accumulation in the parallel guidance process was large for participant C, 7u/s, smaller for B, 4u/s, and very small for A (in fact we set it to 0 for A).

The choice to use just one early termination process for B and C leads to what is probably the only serious misprediction, the multi-modality predicted for their target absent responses. We found that adding a collapsing thresholds assumption in addition to early termination fixed this problem, but decided to show predictions of the simpler model.

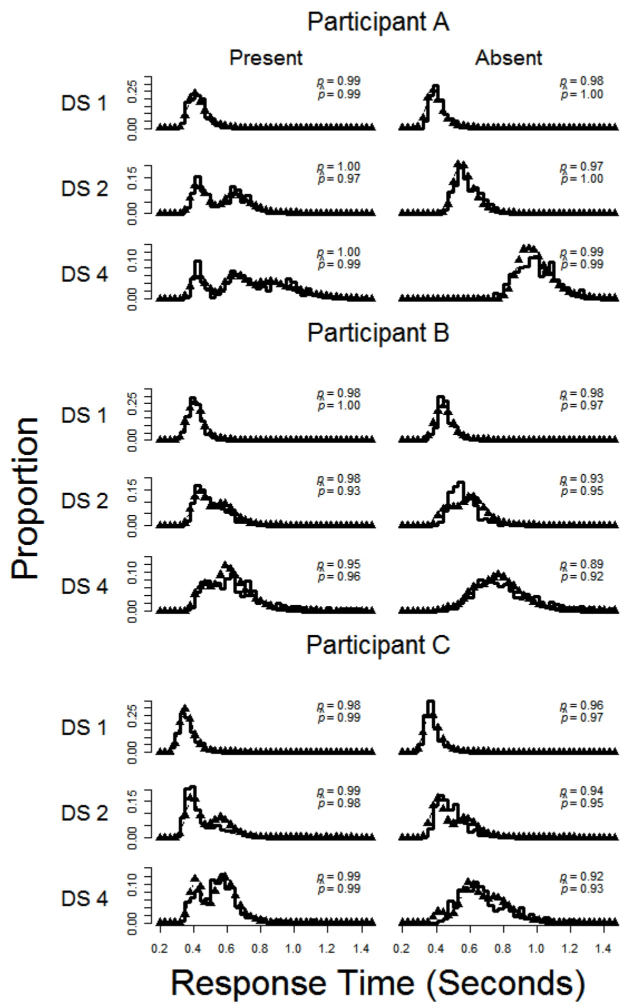

Figure 2 Response time distributions for correct responses in each of the display size conditions (rows) and target present or absent conditions (columns) in the simultaneous presentation (i.e., control) conditions for each of the three participants A-C. Observed data are represented by histograms and model predictions by joined triangles. The proportion of correct responses, both observed and predicted are shown by p and $\hat{p}$, respectively.

Discussion

The model has at its heart a parallel process that guides the order of serial comparisons. However, to fit the exact shape of each participant's full response time distributions requires auxiliary assumptions that do add complexity to the model. The extra complexity is in our view justified because this is essentially the first time that a model has been fit to complete distributions of visual search data, and because the assumptions are plausible and applied in reasonable ways across participants and display size conditions. Another and perhaps even stronger justification lies outside of the present article, because this model has been applied successfully to many more conditions with response time distributions collected by Cousineau and Shiffrin (2004) and to many additional conditions with response time distributions reported in Wolfe et al. (2010). A much longer article is being prepared reporting those applications.

Aside from the general success of this model combining parallel and serial processes in the form of a guided search, the model provides noteworthy insights into individual differences in the way the participants carry out visual search: Even though they carried out identical tasks, they varied enormously in the extent to which they allowed the parallel process to guide serial comparisons. Participants B and C enjoyed strong parallel guidance, suggesting that they had developed a form of perceptual attraction to the target stimuli, and allowed that attraction to guide comparisons. Participant A, on the other hand, appeared not to use such guidance, either because targets did not come to attract attention or because such learning was overwritten and ignored.

In a recent study of stereotype threat on learning in visual search Rydell et al. (in press) showed that women under stereotype threat did not show improvement over training in rates of search, although such improvement was seen for control groups of women. The authors hypothesized that the women under threat may have tried very hard to search without error and in a chosen order of comparison. Since guidance would have interfered with the chosen order, the learning of automatic attention attraction may have been inhibited. That learning had not occurred (as opposed to being ignored) was demonstrated in an incidental transfer task in which trained targets produced interference for the control women but none for the women under threat. It is possible that participant A inhibited learning for related reasons. The idea that participant A may have been 'trying too hard' is consistent with another finding suggested by the modeling, that A chose not to terminate search early. That is, instead of an arbitrary early termination late in search, A instead moved thresholds closer to the start point, so that the search did end with some specified resolution. Participants B and C terminated their search with a target absent response on almost every trial in which they had not found a target after two items had been scanned. One could argue that this behavior was appropriate for these participants because when the target was present they were reliably guided to the target before having to scan more than two items.

It is probably obvious both that the model development reported here was constrained and guided by the data and that a number of alternative assumptions are plausible and need to be tested. Such testing could and should be carried out on the present data sets, on the additional data sets to which the model has been applied but not reported in this paper, and on response time distributions collected in new studies designed to test particular assumptions. Nonetheless, the application to the fitting of full response time distributions is an excellent starting point, and has provided new insights into the nature of visual search and its component processes.

References

Brown, S. & Heathcote, A. (2008). The simplest complete model of choice reaction time: Linear ballistic accumulation. *Cognitive Psychology, 57,* 153-178.

Cousineau, D. & Shiffrin, R.M. (2004). Termination of a visual search with large display size effects. *Spatial Vision, 17,* 327-352.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59-108.

Rydell, R.J., Shiffrin, R., Boucher, K.L., Van Loo, K., & Rydell, M.T. (in press). Stereotype threat prevents perceptual learning. *Proceedings of the National Academy of Sciences*.

Shiffrin, R.M. & Schneider, W. (1977). Controlled and automatic human information processing: II Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127-190.

Thornton, T.L. & Gilden, D.L. (2007). Parallel and serial processes in visual search. *Psychological Review, 114,* 71-103.

Townsend, J.T. & Nozawa, G. (1995). Spatio-termporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology, 39,* 321-360.

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

Wolfe, J.M. (2007). Guided Search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 99-119). New York: Oxford.

Wolfe, J.M., Horowitz, T.S., & Palmer, E.M. (2010). Reaction time distributions constrain models of visual search. *Vision Research, 50,* 1304-1311.