
Parallel Neural Network Models of Visual Search

Seth A. Herd

Department of Psychology
University of Colorado Boulder
345 UCB
Boulder, CO 80309
sethherd@psych.colorado.edu

Randall C. O'Reilly

Department of Psychology
University of Colorado Boulder
345 UCB
Boulder, CO 80309
oreilly@psych.colorado.edu

Abstract

In the ongoing debate regarding the serial and/or parallel properties of visual search, neural network models have previously supported theories of search as parallel. Here we replicate and extend the results of Deco and Zihl (2001), reproducing behavioral feature and conjunction search slopes using models with parallel architecture and moderate biological detail. We find that search is substantially faster in many conditions if a less conservative criteria is used, allowing multiple attentional fixations. Our results show that assuming a parallel location process often leads to a semi-serial search process similar to the Guided Search model.

1 Introduction

There is a long standing debate regarding the nature of visual search: do we locate an object of behavioral importance by moving attention across a scene serially, one object at a time, or by processing everything in that scene at once, in parallel (see Wolfe (1998) for a review). Computational models have recently become useful for exploring this question in combination with behavioral findings.

In a standard visual search (VS) experiment, the participant is asked to search a display for a particular target object among many distractors, and quickly decide whether the target is present. For very easy search tasks, such as finding a single red object among many green ones (feature search), the search time does not depend on the number of distractors. For most searches, however, extra distractors slow search by an amount roughly proportional to their number.

This led to a Feature Integration Theory (FIT), which states that people shift attention serially from one object to the next, deciding for each whether it is the target (Treisman & Gelade, 1980). This process was said to be necessary when combinations of object features (color, shape, size, orientation, etc.) differentiate targets from distractors, e.g., searching for a red X among green X's and red O's (conjunction search).

However, these results could also be the result of inefficient parallel search processes. Theories of this type are supported by a variety of evidence (Duncan & Humphreys, 1989; Chelazzi, 1999). The nature of visual search as serial, parallel, or a mix of both is still a matter of active speculation and debate on both behavioral (Wolfe, 1998) and physiological grounds (Woodman & Luck, 1999; Chelazzi, 1999). We hope to clarify the contribution of

neural network models to that debate. In order to do so, we ignore many interesting exceptions and addendums to the simple summary above, and focus on those central issues.

Recently, a neural network model of parallel search dynamics was presented (Deco & Zihl, 2001). This was simple but effective, reproducing search slopes characteristic of several types of search. Of primary interest, that model reproduced the finding of feature search times independent of number of objects in the search display, and conjunction searches times linearly dependent on number of objects. The model was a neurodynamical network model in which units are thought of as populations of neurons. It used a system of coupled differential equations based on neural population dynamics.

That model reproduced, in neural network form, parallel theories with no serial aspects (Duncan & Humphreys, 1989). These theories essentially say that finding a target location is a diffusion process, and each additional distractor adds noise, but only if admitted to that process by some amount of top-down support. Only the target location is supported by the target template in feature search, while one feature of every distractor representation is supported in conjunction search.

Like those earlier verbal theories, the model of Deco and Zihl (2001) explains behavioral experiments that were originally considered to be evidence of serial movement of attention (Treisman & Gelade, 1980) using a completely parallel process. It lends the credence of a concrete, replicable computational model to earlier verbal theories, in part by making the assumptions of such theories concrete. We further explored a computational model of this type, and discovered a relevant and probably general feature of its behavior.

2 Methods

First we set out to replicate the results of Deco and Zihl (2001), using a different modeling framework. This allowed us to verify our understanding of the computational principles involved, and to verify that these principles could apply outside of the specific set of equations they used. We used the Leabra modeling framework, previously used to model a wide range of psychological phenomena (O'Reilly & Munakata, 2000; O'Reilly, 1998). The Leabra framework is designed to mimic principles of cortical processing. Units are based on the dynamics of single pyramidal neurons, and use the point neuron approximation (including ion currents and membrane potential), as opposed to the population neurodynamics of the previous model. Like the previous model, our framework uses an activity rate code rather than spiking neurons.

We constructed a model similar to that of Deco and Zihl (2001) in structure (figure 1). Our model, like theirs, had two broad levels: a retinotopic feature level, in which each unit represents a specific feature in a specific location, and a feature-insensitive location level. The original model used a separate layer for each feature dimension, in order to accurately account for triple conjunction search results. We do not address those results, and it proved adequate to allow all features to remain in one common layer, in line with the organization of the visual system in which organization by feature representation is loose at most.

One important feature of the Leabra modeling framework is the abstracted inhibition rule, K winners take all (kWTA). Inhibition for each layer is chosen to allow k units to become active. Inhibition is set between the excitatory input level of the K and K+1 units (where K is rank order from most to least input current). This rule seems more desirable than applying a common inhibitory unit to all of each layer, as in the previous model (Deco & Zihl, 2001). A common inhibitory unit becomes more active with more input (more distractors). This behavior, when a fixed stopping criteria is applied, will produce the observed effect of conjunction search costs, but seems to be at odds both with biology and with common theories of visual search. Inhibitory interneurons in visual cortex receive input from only

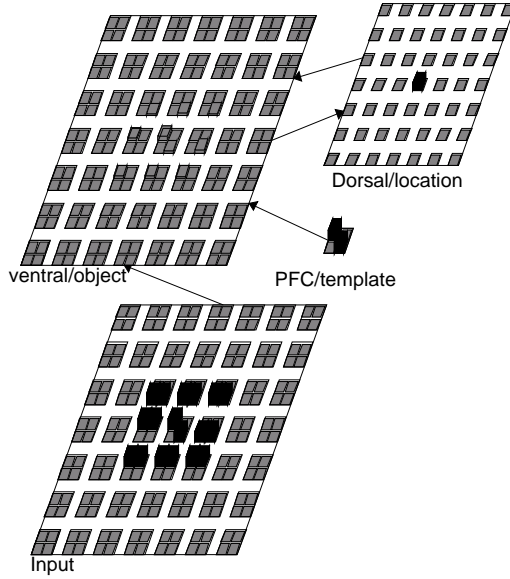


Figure 1: Input layer is externally set to represent a 9-object conjunction search with the target in the center. In the input and object layers, two units in each four-unit location group represent different colors, while the other two represent different shapes. The connection from input to object layer is one-to-one, with uniform weights. Connection from object layer to location layer is one-to-one in that all units at a given location project to the location layer, and receive feedback connections from that layer. Connections from PFC/Template layer to object layer are one-to-one in that each unit in the PFC/Template layer connects to one corresponding unit in the Object/Ventral layer at each location. The response criteria is the activation of any location unit above a threshold of .5. This response corresponds to completing the focus of spatial attention upon a certain location.

a small fraction of any one visual area (Salin & Bullier, 1995). Also, this explanation is at odds with the common theory that search costs are a result of noise (Duncan & Humphreys, 1989; Chelazzi, 1999). The kWTA rule does not tend toward higher inhibition with more units active, and therefore seems to be both more biologically plausible, and a better fit with existing theories.

3 Results

First, we observed that our model produced results quantitatively similar to those of the previous model, and to behavioral results. We obtained nearly flat search slopes in the feature search condition, and a linear increase in time to settle with additional distractors in conjunction search (figure 2).

In addition, it was apparent that this linear increase was driven by noise, as the search cost per distractor varied with the amount of gaussian noise applied to the net input current on each time step (figure 2). This gives an explanation of varying behavioral search slopes as a larger signal/noise ratio for more easily discriminated stimuli.

The model's dynamics are as follows: during feature search, the single unit in the object layer representing the searched for feature is seen to activate fairly quickly, while few units corresponding to distractor features activate more slowly,

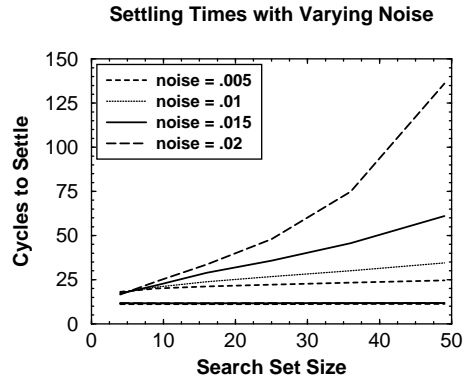


Figure 2: Settling times for our model. Sloped lines are conjunction search; lower flat lines are feature searches. The amount of noise affects the settling slope for conjunction search, but does not affect the feature search settling time.

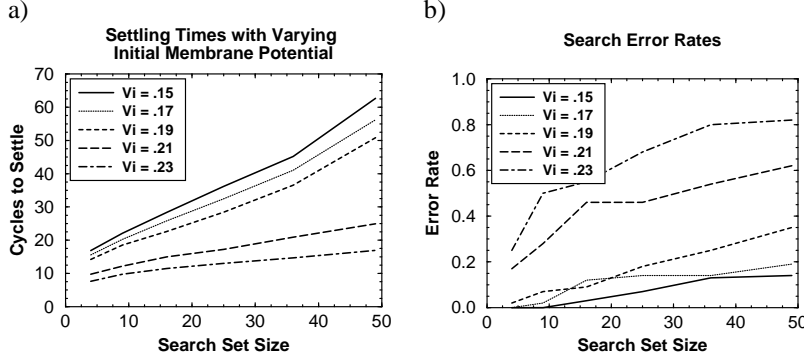


Figure 3: a) Location process times for varying starting states. Locating a potential target is dramatically speeded by larger starting membrane potentials, corresponding to a lowered threshold. b) Error rates rise rapidly as the location process becomes faster.

and then become active only briefly through noise. The target unit's activity fairly quickly activates the corresponding location unit, ending the trial. During conjunction searches, all units in both target and distractor locations that represent target features become active, and their activity level fluctuates. There are two units representing target features in the location of the target, and over time these will usually out-compete the one active unit in distractor locations. This competition takes longer to conclude when more distractors are present, since each distractor location occasionally and briefly out-competes the target location.

Because this model is essentially a diffusion process model in which information is accumulated over time in a noisy environment, we reasoned that it should be possible to speed the settling process at the cost of accuracy. Many variables could affect the system in this way. We chose to vary the starting value of the membrane potential. This has the effect of placing the system closer to settling, giving it a shorter time to accumulate evidence. It also seems that this is a likely variable for online adjustment by the cognitive system; providing extra diffuse input before a trial will provide a baseline activity level, and put the system closer to its response threshold.

Raising the system's baseline activity level produced a dramatic speedup of settling, at the cost of an equally dramatic reduction in accuracy (figure 3). Our preliminary work suggests that this behavior is very general to this class of models, persisting through a wide range of parameters and architectures. However, it remains to test the generality of these results more thoroughly.

Is this reduction in accuracy disastrous for the performance of the system? It is if we assume that every missed location is a missed trial; behavioral performance usually shows a less than 10% error rate. However, if we instead assume that the system checks the accuracy of its response with an object identification process, then chooses a new location if that object does not match the target template, then risking wrong location guesses could be a good strategy.

If every missed location process results in a repeat of that process, the total search time will be given by $(\text{location time} + \text{identification time}) / (1 - P(\text{error}))$, since the series $1 + x + x^2 + x^3 + x^4 \dots$ converges to $1/(1 - x)$ for $x < 1$. That series corresponds to the total number of location processes that will be completed on average when $x = P(\text{error})$, or alternately, one plus the average number of errors per trial.

The speedup of search proved so dramatic that the system can afford one or even more missed attentional fixations, depending on assumptions about how long the identification

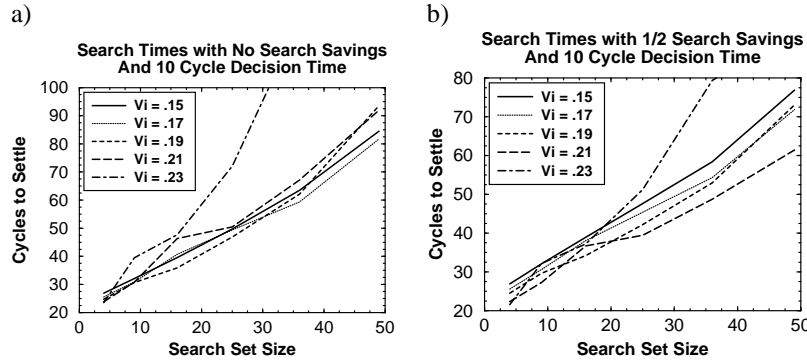


Figure 4: a) total conjunction search times under the consideration that there is an object identification process that takes the equivalent of 10 processing cycles of the model, about the same amount of time the system takes to settle on the location of the target in a feature search. b) total conjunction search times under the same considerations for identification times, but with the more reasonable assumption that some location information is retained so that additional location processes take 1/2 the time of the first

process takes, and the signal strength and noise level. Figure 4a gives the total search times under the assumption that an identification process takes 10 extra cycles. Even though that process is fairly costly, it can be seen that less conservative location processes are competitive with those that locate the target on the first try.

This assumption is probably still too conservative; it seems unlikely that no information is retained from the location process after the first settling process. If we assume that later location processes take 1/2 the time of the first, due to retained information, search efficiency is biased even further toward processes that make some mistakes in the interest of a faster location process (figure 4b). In this case, an intermediate parameter setting is the most efficient over the whole range of display sizes, while the most efficient search parameters vary with changing display size. Of course each missed location process results in a wasted object identification process, so if object identification is very slow relative to the location process, a conservative (and therefore parallel) process will be most efficient.

4 Discussion

This conception of search processes, based on an entirely parallel target location process, has converged with the Guided Search model (Wolfe, Cave, & Franzel, 1989; Wolfe, 1994), in which a serial search is guided by a parallel “saliency map” operation. If we assumed that the process retained all of its information instead of 1/2 as in figure 4b, we would have exactly reproduced Wolfe’s guided search model (We do not make this assumption because Wolfe’s own recent work has shown that location information retention is not nearly perfect, (e.g. Horowitz & Wolfe, 2001).

In our model, the parameters that lead to the fastest search depend on how long an identification process will take, the amount of noise in the system, and the search display size. The first two parameters can be expected to vary with the perceptual discriminability of target vs distractors, while the participants knowledge of the last can be varied experimentally. Our analysis predicts that subjects should be measurably more efficient for searches in which they know the display size before the trial.

According to this analysis, parallel neural network models lead directly to the conclusion that, under many conditions, search will have a small number of serial fixations, and that

the number of such fixations will depend on the exact parameters of such search. This conclusion corresponds well to the finding from eye tracking experiments that participants in visual search tasks that allow eye movements show a small number of fixations in searching relatively large displays (Williams & Reingold, 2001; Brown & Gilchrist, 2000)

The optimal parameters for a search process, including average number of attentional fixations, vary according to the task parameters. This finding raises the fascinating but seldom mentioned possibility that search process actually take on different characteristics, ranging from nearly completely parallel to nearly completely serial, depending on task demands. This idea can explain why physiological data seems to convincingly point to a parallel process in some cases (reviewed in Chelazzi, 1999), and to a serial process under different task demands (e.g. Woodman & Luck, 1999).

The parallel nature of neural network models makes it unsurprising that the first forays into modeling VS results with ANN methods have been fully parallel search models (Deco & Zihl, 2001). However, taken in a context of overall task demands, these models actually support a view of visual search in which the parallel vs. serial nature of search varies, depending on task demands. This view can potentially unify the many behavioral and neuroscientific studies that present potent but conflicting evidence for both types of search.

5 References

- Brown, J. M. F. V., & Gilchrist, I. D. (2000). Saccade target selection in visual search: The effect of information from the previous fixation. *Vision Research*, 41(1), 87–95.
- Chelazzi, L. (1999). Serial attention mechanisms in visual search: A critical look at the evidence. *Psychological Research*, 62, 195–219.
- Deco, G., & Zihl, J. (2001). Top-down selective visual attention: A neurodynamical approach. *Visual Cognition*, 8, 119–140.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3).
- Horowitz, T. S., & Wolfe, J. M. (2001). Search for multiple targets: Remember the targets, forget the search. *Perception & Psychophysics*, 63 (2), 272–285.
- O'Reilly, R. C. (1998). Six principles for biologically-based computational models of cortical cognition. *Trends in Cognitive Sciences*, 2(11), 455–462.
- O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.
- Salin, P. A., & Bullier, J. (1995). Corticocortical connections in the visual system: Structure and function. *Physiological Review*, 75, 107–154.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. *Psychonomic Bulletin & Review*, 8 (3), 476–488.
- Wolfe, J. M. (1994). Guided search 2.0- a revised model of visual search. *Psychonomic Bulletin and Review*, 1(2), 202–238.
- Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–73). Philadelphia: Psychology Press.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance*, 15(3), 419–433.
- Woodman, G., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, 400, 867–869.