Controlling Attention With Noise:
The Cue-Combination Model of Visual Search

David F. Baldwin
College of Computer Science
Northeastern University
dfb@ccs.neu.edu

Michael C. Mozer
Institute of Cognitive Science
University of Colorado at Boulder
mozer@colorado.edu

Abstract

Visual search is a ubiquitous human activity. Individuals can perform a remarkable range of tasks involving search for a target object in a cluttered environment with ease and efficiency. Wolfe (1994) proposed a model called Guided Search to explain how attention can be directed to locations containing task-relevant visual features. Despite its attractive qualities, the model is complex with many arbitrary assumptions, and heuristic mechanisms that have no formal justification. We propose a new variant of the Guided Search model that treats selection of task-relevant features for attentional guidance as a problem of cue combination: each visual feature serves as an unreliable cue to the location of the target, and cues from different features must be combined to direct attention to a target. Attentional control involves modulating the level of additive noise on individual feature maps, which affects their reliability as cues, which in turn affects their ability to draw attention. We show that our *Cue-Combination Guided Search* model obtains results commensurate with Wolfe's Guided Search. Through its leverage of probabilistic formulations of optimal cue combination, the model achieves a degree of mathematical elegance and parsimony, and makes a novel claim concerning the computational role of noise in attentional control.

Introduction

Visual search is a ubiquitous human activity. We search for our keys on a cluttered desk, a familiar face in a crowd, an exit sign on the highway, our favorite brand of cereal at the supermarket, and so forth. That the human visual system can perform such a diverse variety of tasks is truly remarkable.

The flexibility of the human visual system stems from the top-down control of attention, which allows for processing resources to be directed to task-relevant regions and objects in the visual field. How is attention directed based on an individual's goals? To what sort of features of the visual environment can attention be directed? These two questions are central to the study of how humans interact with their environment.

Visual search has traditionally been studied in the laboratory using cluttered stimulus displays containing artificial objects. The objects are defined by a set of *primitive visual features*, such as color, shape, and size. For example, an experimental task might be to search for a red vertical line segment—the *target*—among green verticals and red horizontals—the *distractors*. In a reaction-time paradigm, the difficulty of the task is assessed by measuring the response latency to detect the presence or absence of a target.

With a burgeoning experimental literature, models of visual search have been proposed to explain data within a mechanistic framework (e.g., Mozer, 1991; Sandon, 1990, Itti & Koch, 1998). Perhaps the most influential and thoroughly developed model is Guided Search 2.0 (Wolfe, 1994), which we'll refer to as *GS2.0*. Guided Search has been refined further since (Wolfe & Gancarz, 1996; Wolfe, 2001), but its essential claims have remained constant and have been used as a theoretical framework for explaining visual search data for over a decade. Figure 1 (left panel) shows a sketch of GS2.0.

GS2.0, like most models of early visual processing, supposes that the visual scene is analyzed by independent retinotopic *feature maps* that detect the presence of primitive visual features across the retina along dimensions such as color, orientation, and scale. The feature maps represent each dimension via a coarse coding, i.e., the maps for a particular dimension—referred to as the *channels*—are highly overlapping and broadly tuned. GS2.0 characterizes the channels as encoding categorical features. For example, color has four channels representing the salient color primitives—red, green, blue and yellow; orientation also has four channels representing left, right, steep, and shallow slopes.

The categorical channels are analyzed by a differencing mechanism that enhances local contrast, yielding a *bottom-up activation*. Top-down control of attention takes place by emphasizing task-relevant channels—the set of channels, one per feature, that best distinguishes the target from its distractors. For example, given a red vertical target among green horizontal distractors, the red and vertical channels should be enhanced, yielding a *top-down activation*.

The bottom-up and top-down activations from all channels are combined to form a *saliency map* in which activation at a location indicates the priority of that location for the task at hand. Attention is directed to locations in order from most salient to least, and the object at each location is identified. The model supposes that response time is monotonically related to the number of locations that need to be searched before a target is found. (The model includes rules for terminating search if no target is found after a reasonable amount of effort.)
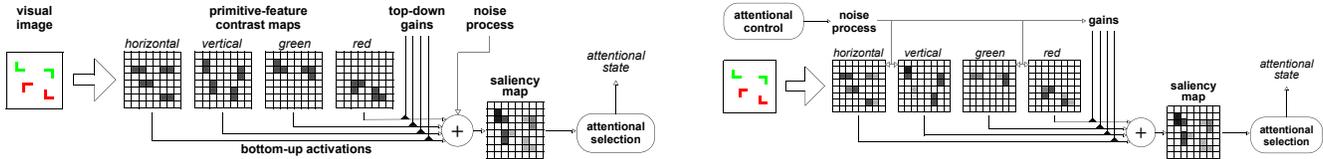
Figure 1: (left panel) Architecture of GS2.0; (right panel) architecture of CCGS

Critique of GS2.0

GS2.0 has explained key phenomena from the visual search literature (Wolfe, 1994), and ongoing elaboration of the model allows it to accommodate an even broader range of phenomena and functionality. GS2.0 makes one key theoretical claim—that task-relevant guidance of visual search is achieved by modulating the influence of different primitive visual feature types on the deployment of attention. Although we find this claim compelling, we have two specific concerns with its translation into the specific design and implementation of GS2.0.

(1) GS2.0 incorporates many arbitrary assumptions and mechanisms that have little formal rationalization. As a result, the model appears somewhat ad hoc, and it is difficult to determine which properties of the model give rise to its behavior in any task. As one example, GS2.0 includes a series of heuristics to determine the weighting of feature-map inputs into the saliency map. These heuristics are based on intuition, not a mathematical foundation. Because the model is quite complex relative to amount of data it can explain, it lacks parsimony and elegance. GS2.0 would clearly benefit from a principled computational theory that helped to constrain the choice of processing mechanisms.

(2) Like most models of human cognition, processing in GS2.0 is corrupted by noise, which limits the model's performance and ensures that it is comparable to that of human subjects. To explain why GS2.0 incorporates noise, we must provide a bit of background. GS2.0 allows attention to be guided to locations containing task-relevant features. For example, in the case of search for a red vertical, the saliency map receives a contribution from the red channel and from the vertical channel. Because the location of a red vertical contains two task-relevant features, whereas locations of distractors contain at best one task-relevant feature, the top-down saliency of the target location is greater than that of any distractor location, leading to rapid target detection. However, experimental results are inconsistent with this claim: latency to detect the target in a conjunction-search task is strongly dependent on the number of distractors in the display.

Wolfe explains the inefficiency of attentional guidance to a conjunction target by assuming that Gaussian noise is injected into the saliency map. To match human data, the noise must be of extremely high amplitude—typically with a standard deviation that is 12–25% of the activation level. Further, the model has the unusual property that the noise decreases as the signal (activation into the saliency map) increases; i.e., the noise and signal are not independent. The rationale for this as-

sumption is unclear, but presumably it is essential for explaining data.

This treatment of noise also seems unrealistic in that early stages of processing in the model—i.e., feature extraction, contrast enhancement, weighting of feature maps—are entirely noise free and rely on a veridical representation of the environment. Noise enters only at the stage of attentional selection. Noise lacks a purposeful treatment in GS2.0, and appears to be incorporated as an afterthought to replicate human data.

Reconceptualizing the role of noise

In this paper, we suggest a novel perspective on Guided Search, which is obtained by reconceptualizing the role of noise. Rather than viewing noise as an ad hoc means of "dumbing down" the model, our perspective views noise as being intimately linked to the operation of attentional control processes. According to our perspective, each primitive-feature channel conveys information that can potentially provide *cues* as to the location of the target. The cues provided by any one channel may be intrinsically unreliable. For example, in the case of search for a red vertical target, activation in the red channel will indicate both the target and some distractors, and likewise for the vertical channel. Nonetheless, the red and vertical channel activations have higher *cue reliability* than the green and horizontal channel activations, because activation in the green and horizontal channels never indicates the target location.

In GS2.0, each top-down feature map (cue) has an associated weight that specifies the contribution of that cue to the saliency map. The set of weights is assumed to be determined by a control processes that attempts to assign larger weights to cues that have higher reliability.

However, as we will argue, the optimal cue-combination weights—the weights that will yield the best performance—can be determined analytically if the cue reliability is known. Therefore, one might consider the following hypothesis concerning the operation of attentional control: *Because the optimal combination weights can be derived from cue reliability, attentional control processes should not adjust weights but rather should modulate the cue reliability directly, via the injection of noise.* In essence, the hypothesis suggests a mechanism that can attenuate the level of additive noise to each feature map individually. The noise level affects cue reliability, and contributions to the saliency map are weighted—in a manner we'll explain shortly—by cue reliability. Because our perspective conceptualizes attentional control as fundamentally a *cue combination* problem, we refer to this perspective as *Cue-Combination*

Guided Search or *CCGS*. According to this perspective, noise is used to guide visual attention. CCGS is sketched in the right panel of Figure 1.

CCGS has four virtues relative to GS2.0. (1) In CCGS, the earliest stage of feature registration is corrupted by noise, and this noise then propagates along the processing pathway. As a result, noise affects every stage of the model, in contrast to GS2.0 which makes the questionable claim that noise influences only late stages of processing. (2) Because noise cascades through the model, the magnitude of noise necessary to model human performance is significantly smaller, with a standard deviation of roughly 8% of the activation level, versus 20% for GS2.0. Further, noise in CCGS is independent of signal strength, in contrast to GS2.0. (3) By viewing attentional control as cue combination, CCGS is able to leverage principled probabilistic approaches to optimal cue combination, leading to a variant of Guided Search within a mathematically elegant framework. These probabilistic approaches provide strong constraints on processing mechanisms, yielding a model that is simpler, requires fewer assumptions, arbitrary parameters, and ad hoc mechanisms than GS2.0. (4) Finally, CCGS offers a novel view of attentional control in terms of the modulation of noise. CCGS therefore treats noise not as a deleterious side effect of neural hardware, but as serving a computational role in cognition.

Cue combination

Cue combination is a problem that pervades perception. For example, in determining the three-dimensional structure of the world from visual images, depth information can be provided by motion, texture, and binocular cues. Each cue in isolation may be unreliable, and the challenge of cue combination is to reconstruct an accurate measure from the unreliable cues.

In many cases, the human visual system combines cues optimally (Jacobs, 2002). Probabilistic methods can be used to determine the formula for optimal cue combination. Formally, the task is to combine two (or more) unreliable cues, c_1 and c_2, to estimate some unknown signal, S. In a probabilistic framework, the reliability of the individual cues corresponds to having an estimate of the cue-conditional probability distributions $P(S|c_1)$ and $P(S|c_2)$, and the result of cue combination is to use these distributions to estimate $P(S|c_1, c_2)$. More importantly, one wishes to obtain the optimal estimate of the signal, s_{12}^*, which is the value of S that maximizes $P(S|c_1, c_2)$. Under certain assumptions, including the assumption that the cue-conditional distributions are Gaussian, s_{12}^* can be estimated from a linear combination of s_1^* and s_2^*, the optimal estimates of the signal based on cue 1 alone and cue 2 alone, respectively, where the weight on each cue is the normalized inverse of its variance (Yuille & Bülthoff, 1996). The formula is presented in the "Implementation" section that follows.

What this result boils down to in the case of CCGS and attentional control is the following. Mean-zero Gaussian noise is added to the activation in each feature map. Control processes can attenuate the noise on individual maps, although there are assumed limits on the range over which the noise variance can be modulated. Noise cannot be suppressed entirely, reflecting either an intrinsic noise level in the system or the influence of unmodeled and task-irrelevant processes. Noise variance also has an upper limit, reflecting the fact that neural activities have a limited range. By adopting the ideal observer assumption that cue combination is optimal, one obtains a formula for combining activations from the noisy feature maps into the saliency map. The formula is a weighted summation, with the weightings determined indirectly by control processes via the noise levels. CCGS is related to a mechanism used by Triesch and von der Malsburg (2001) for determining feature relevance in a face tracking task, although their focus was not explicitly on attentional control.

We incorporate in CCGS a simple learning algorithm by which control processes tune noise levels so as to maximize the efficiency of visual search i.e., to obtain the maximum level of activation for the target in the saliency map, relative to the activation level of distractors. Effectively, over a series of experimental trials, this learning algorithm will maximize the noise level for task-irrelevant channels, and will minimize the noise level for channels useful for locating the target.

CCGS Implementation

This section provides an algorithmic level description of CCGS. To clearly delineate where CCGS differs from GS2.0, we mark sections that differ with an asterisk (*). Unless otherwise noted, the algorithm we describe is identical to GS2.0 and the interested reader can refer to Wolfe (1994) for details we have omitted.

Activating Channels

As in GS2.0, our implementation includes two visual feature dimensions—color and orientation—each with four channels. The stimulus objects we use in our simulation are therefore colored, oriented line segments. We model stimulus displays that consist of an 8x8 array of cells, each of which contains a single object or is blank.

The visual scene representation consists of values for the color and orientation of objects (or lack thereof) at every location in the 8x8 array. We use the notation $V_f[m, n]$ to refer to the value of feature dimension f (color or orientation) at row m and column n of the array. The value of the orientation feature is specified in degrees, and the value of the color feature is specified in units monotonically related to wavelength.

The activation of a channel c of feature f at location $[m, n]$ is denoted $A_{f,c}[m, n]$. Each channel has a broadly tuned receptive field function, g_c, such that $A_{f,c}[m, n] = g_c(V_f[m, n])$. As is typical of neural receptive fields, g_c is tuned to a particular feature value, meaning its maximal response is to this value, and the response falls off monotonically with increasing distance from this value. In CCGS as well as GS2.0, the activation pattern corresponding to a stimulus display consists of 2x4x8x8=512 activity levels (feature dimensions x channels x array size), where each activity level is in

the range [0, 1]. Blank cells in the stimulus array produce an activation level of zero for all channels.

Injecting Noise*

Unlike GS2.0, which assumes large-amplitude additive noise to the saliency map, CCGS introduces smaller magnitude noise directly into the channel activations, $A_{f,c}[m,n]$. The noise is drawn from a mean-zero Gaussian distribution with standard deviation $\sigma_{f,c}$. The determination of the $\sigma_{f,c}$ parameters, which constitutes attentional control, is described shortly.

Enhancing Feature Contrast

We summarize the contrast-enhancement mechanism, unchanged from GS2.0. For each cell $[m,n]$ of feature f and channel c, a local contrast is computed between that cell's activation and the activity in its local neighborhood. The contrast representation, denoted $C_{f,c}[m,n]$, is defined as:

$$C_{f,c}[m,n] = A_{f,c}[m,n]/25 \sum_{-2 \le i,j \le 2} q_{f,c}(m,n,i,j), \quad (1)$$

where N is the number of neighbors in the 5x5 region centered on $[m,n]$, the function $q(.)$ computes a Euclidean-distance weighted contrast between cells $[m,n]$ and $[i,j]$, given a thresholding function $\theta(.)$:

$$q_{f,c}(m,n,i,j) = \frac{\theta(|A_{f,c}[m,n] - A_{f,c}[m+i,n+j]|)}{||[m,n] - [i,j]||}. \quad (2)$$

Determining Channel Weights*

In GS2.0, each feature map is assigned a weight that determines the contribution of that map to the guidance of attention. GS2.0 claims that top-down control processes modulate the weights. In contrast, CCGS supposes that control processes modulate the noise level on each feature map, and that the weights are determined automatically by the optimal cue-combination rule. By the cue-combination rule, the weight $w_{f,c}$ for channel c of feature f is the normalized reciprocal of the noise variance:

$$w_{f,c} = \sigma_{f,c}^{-2}/\sum_{f',c'} \sigma_{f',c'}^{-2}, \quad (3)$$

where $\sigma_{f,c}$ is the standard deviation of the mean-zero Gaussian noise added to the feature map activity.

Effectively, as the noise level of a map increases relative to the noise levels of other maps, CCGS gives less credence to the map in guiding attention. Note that the normalization is over all feature maps. Therefore, the extent that CCGS is able to suppress noise on, say, the red channel, it will result in a lower cue weight on, say, the steep channel. This natural competition for the guidance of attention helps explain why search for a conjunction of features is typically less efficient than search for a single feature.

The cue-combination rule is optimal if noise is Gaussian. Although the noise injected into the feature maps is Gaussian, it undergoes a nonlinear transformation by

the contrast-enhancement mechanism. Nonetheless, we have empirically found that the resulting noise is very close to Gaussian: over a sample of activations of an item after training the model, the resulting empirical distribution had a sample skew approaching zero and a sample kurtosis within a small deviation of that of a Gaussian distribution. This finding might be attributed to the central limit theorem, because the salience map involves combination from multiple noise sources.

Guiding Attention

The saliency map is a weighted linear combination of the feature map activations. Both GS2.0 and CCGS assume that the saliency map determines attentional priority. Locations are searched from the most salient location to the least, and reaction time, RT, measured in milliseconds, is assumed to be monotonically related to the number of locations, n, inspected before the target is found. Specifically, $RT = 50n + 400 + \eta$, where η introduces small-magnitude uniformly distributed noise.

Optimizing Noise Levels*

In CCGS, top-down control of attention is achieved by modulating the standard deviation of the Gaussian noise added to each feature map. The noise level for channel c of feature f, $\sigma_{f,c}$, is incrementally updated after each trial. By adjusting noise levels, CCGS attempts to satisfy the performance optimization criterion that the ranking of the target in the saliency map is higher than that of all distractors. If S_t is the saliency of the target location S_d is the saliency of a distractor location, then the criterion is achieved if $S_t > S_d$ for all distractors d in the set of distractor locations, D.

CCGS attains this criterion via a gradient ascent search, with one update of the $\{\sigma_{f,c}\}$ following each trial. Because the optimization criterion involves a discrete measure—the ranking of locations—and gradient ascent requires a continuous search space, we define a continuous performance measure P using the two-alternative softmax (logistic) function:

$$P = \sum_{d \in D} 1/(1 + exp(S_d - S_t)), \quad (4)$$

where the term in the summation is greater than 0.5 only if $S_t > S_d$, and approaches 1.0 as $S_t >> S_d$.

Gradient ascent implies the update $\Delta\sigma_{f,c} = \epsilon \partial P/\partial \sigma_{f,c}$, where ϵ is the step size. The $\{\sigma_{f,c}\}$ affect P in two ways: via the combination weights, and via the noise added to $A_{f,c}$, the feature maps. However, we treat the latter effect as a constant, thereby performing a type of stochastic gradient ascent. In other words, the $A_{f,c}$—and consequently the $C_{f,c}$—are treated as constants, and gradient descent maximizes P only for the current stimulus display and current sample of noise.

As we have described the learning procedure thus far, there are no limits on $\sigma_{f,c}$. However, clearly the noise level cannot be negative, and it seems sensible to place limits on the degree to which control processes can modulate the noise level. Consequently, we define $\sigma_{f,c} = bl_{f,c}$, where b is a baseline noise level and $l_{f,c}$

is a multiplicative modulation of the noise level, and we impose the restriction $||\mathbf{l}-\mathbf{1}|| \leq m$, where m is the maximum deviation by which control processes can modulate the noise level. In our simulations, we chose $b = 0.075$ and $m = 5$. These parameters were chosen via a coarse grid search over these two dimensions to find values that resulted in trends similar to GS2.0. We explored other techniques for limiting the modulation of noise by control processes (e.g., using an L1 norm instead of L2, placing limits on the range of individual $l_{f,c}$, etc.), but the choice of technique does not appear to significantly affect the model's qualitative behavior.

Figure 2: Simulation results from GS2.0 (left column) and CCGS (right column) for target-present trials of six experiments described in the text.

Results

Wolfe (1994) simulated GS2.0 on a series of experiments and showed that GS2.0 matched human performance. To verify the accuracy of our implementation of GS2.0 (the exact algorithm was not crisply specified in Wolfe, 1994), we reimplemented the model and ran it on the series of experiments. Each simulation began with a sequence of 100 practice trials, to set adaptive parameters of GS2.0, and was followed by a sequence of trials that included 1000 trials for each condition, with randomly chosen instances of displays for each condition (e.g., varying the locations of display elements, distractor identities).

With one exception to be described, our replication of GS2.0 matched Wolfe's (1994) simulations. The left column of Figure 2 shows each of six simulation results in a separate graph. We explain the graphs shortly. The right column of Figure 2 shows the corresponding simulations of CCGS. The key point to note is that CCGS obtains the same general pattern of results as GS2.0.

The original GS2.0 accurately models human data and, as such, it is not necessary to include the human data graphs. Further, only target-present trials are shown in our graphs. Target-absent trials were also simulated, but because we incorporated the GS2.0 mechanism for handling target-absent trials into CCGS, the two models produce the same result: approximately a 2:1 search-slope ratio for target-absent to target-present trials.

We briefly describe the six graphs. The first four involve displays of a homogeneous color, and search for a target orientation among homogeneous distractors of a different orientation. **Graph 1** (row 1 of Figure 2) explores an asymmetry in search for a target orientation as a function of the number of elements in the display (display size). Search for a 20° target among 0° distractors is efficient (dashed line)—i.e., search time does not depend on display size—but search for a 0° target among 20° distractors is inefficient (solid line)—i.e., search time increases linearly with display size. **Graph 2** describes search for a 0° target among distractors of a different orientation. The graph plots the *slope* of the function relating display size to response latency, as a function of the distractor orientation. Search slopes decrease monotonically as the target-distractor orientation difference increases. **Graphs 3 and 4** involve search for various target-distractor orientation combinations as a function of display size. **Graph 5** involves conjunction search. The curve with the shallower slope—but still indicative of inefficient search—corresponds to the condition in which the target is a red vertical among distractors that are green verticals and red horizontals. The steeper slope is obtained the target and distractor orientations are made more similar (40° and 90°). **Graph 6** examines search efficiency for a conjunction target (red vertical) among 55 distractors (red 60° and yellow vertical) as a function of the ratio of the two distractor types. The result shows that both models can guide search more effectively—i.e., response times are faster—if either target color or target orientation is relatively unique.

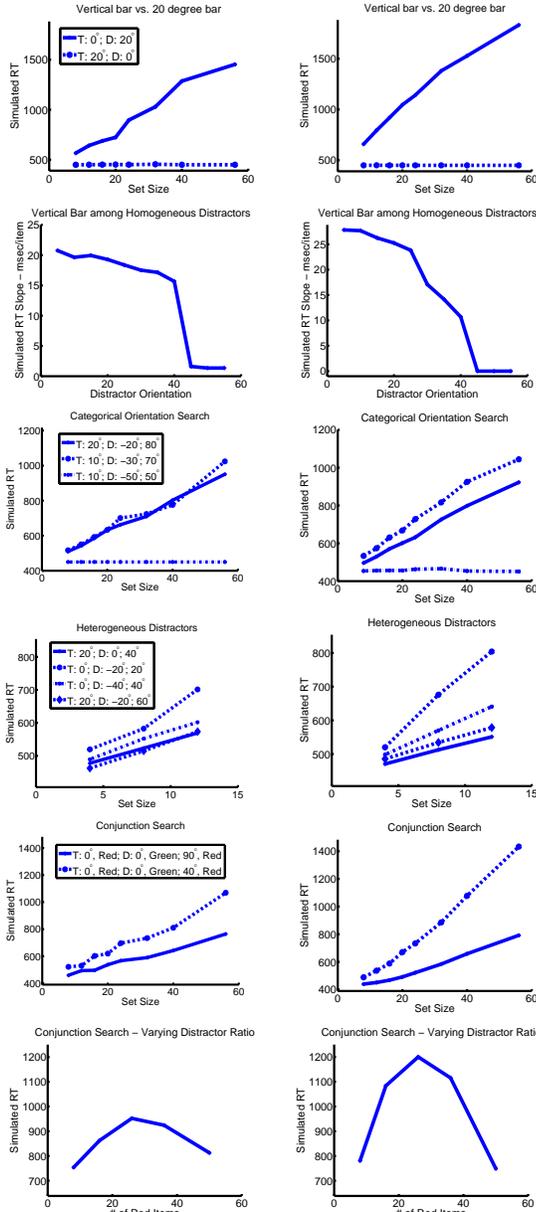CCGS and our replication of GS2.0 yield quite similar

results, all the more impressive considering that we did not tune model parameters carried over from GS2.0 (e.g., the rule that translates search ranking to RT). However, we note that our replication of GS2.0 did not match Wolfe's (1994) simulations in one regard: In graph 2, our replication of GS2.0 obtains a search slope for 0° targets among 20° distractors that is far steeper than Wolfe's original simulation. We were unable to determine how our implementation differs from Wolfe's description of GS2.0, but do not find the discrepancy of great concern, because our primary goal was to demonstrate that the cue-combination notion could be incorporated into GS2.0 without influencing its performance.

Discussion

We proposed a new model, CCGS, which, like GS2.0, is based on the principle that attention can be guided toward locations in the visual field likely to contain a target. Control processes modulate which primitive visual features drive attention via the saliency map. We have shown that CCGS explains the same data as GS2.0. Nonetheless, many advances in cognitive science have come about by taking intuitively sensible but heuristic models and reformulating them in terms of an underlying mathematical principle. As a result of this reformulation, CCGS is arguably simpler—requiring fewer arbitrary assumptions and mechanisms than GS2.0—and therefore makes stronger predictions and serves as a more parsimonious account of attentional selection.

Wolfe's Guided Search model has evolved since the 2.0 version in 1994, but the core of the model remains unchanged. Follow ups include mechanisms to simulate saccadic eye movements (Wolfe & Gancarz, 1996), and to model target selection more realistically (Wolfe, 2001). We believe these more recent incarnations of Guided Search are compatible with the primary insight of CCGS—viewing attentional control as a cue combination problem that is solved by the modulation of noise.

CCGS and GS2.0 differ fundamentally in their treatment of noise. GS2.0 makes some questionable assumptions concerning noise: noise contributes only to the late stages of processing, is of high amplitude, and is signal dependent. In contrast, CCGS assumes that noise affects all stages of processing, is relatively low amplitude, and is independent of the signal. However, the most interesting claim of CCGS with regard to noise concerns the pivotal role of noise in guiding attention. Although noise has a cost—it introduces task- and stimulus-irrelevant variability into the model's behavior—CCGS suggests that noise may have a purpose—that of suppressing the influence of task-irrelevant features on the guidance of attention. If noise were attenuated on all feature maps simultaneously, then according to CCGS, top-down control of attention would not be feasible. We find this sort of explanation to be more satisfying than one claiming that noise is simply thrown in weaken a model that would otherwise outperform humans. CCGS suggests a computational contribution of noise.

The notion that attentional control involves the modulation of noise is consistent with an intriguing empirical study. Lu and Dosher (1998) examined three possible mechanisms by which attention might affect visual information processing, and found support for the hypothesis that the allocation of attention corresponds to the suppression of early additive noise in the visual system.

Cue combination may not seem a natural way of conceptualizing integration of information across feature maps to form a saliency map, at least at the abstract mathematical level that we have described it. However, it has a very simple, elegant implementation in terms of neurobiology. Theoreticians in computational neuroscience have suggested that populations of neurons might be used to encode not only feature values but also uncertainty in the value—essentially a probability distribution. Simple neural mechanisms have been proposed that essentially perform cue combination using population codes (Pouget, Dayan, & Zemel, 2003).

CCGS also extends and improves on GS2.0 by proposing a principled account of attentional control as optimization. CCGS specifies a performance objective—to locate the target in the minimum time—and includes a learning rule that is guaranteed to move performance toward this objective. Although GS2.0 included a set of heuristics for setting control parameters (the top-down weights), CCGS offers a computational theory that characterizes attentional control as optimization. Because this optimization takes place over a series of trials, we are currently investigating the use CCGS to account for sequential effects in performance often found in attentional tasks.

Acknowledgments

References

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Patt. Anal. & Mach. Intell.*, *20*, 1254–1259.

Jacobs, R.A. (2002) What determines visual cue reliability? *Trends in Cognitive Science*, *6*, 345–350.

Lu, Z.-L., & Dosher, B. A. (1998). External noise distinguishes attentional mechanisms. *Vision Research*, *38*, 1183–1198. external noise at the target location. *Journal of Vision*, *2*, 312–323.

Mozer, M. C. (1991). The perception of multiple objects: A connectionist approach. Cambridge, MA: MIT Press.

Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Ann. Rev. Neurosci.*, *26*, 381–410.

Sandon, P. A. (1990). Simulating visual attention, *Cognitive Neuroscience*, *2*, 213–231.

Triesch, J., & von der Malsburg, C. (2001). Democratic Integration: Self-Organized Integration of Adaptive Cues, *Neural Computation*, *13*, 2049–2074.

Wolfe, J.M. (1994). Guided Search 2.0: A Revised Model of Visual Search. *Psychi. Bull. & Rev.*, *1*, 202–238.

Wolfe J.M., & Gancarz G. (1996). Guided Search 3.0: A model of visual search catches up with Jay Enoch 40 years later (pp. 189–192) In: Lakshminarayanan V, (Ed.) *Basic and Clinical Applications of Vision Science*, Dordrecht, Netherlands: Kluwer Academic.

Wolfe, J. (1998). Visual Search, In H. Pashler (Ed.), *Attention*, pp. 13–74. East Sussex: Psych. Press Ltd.

Yuille, A.L. and Bülthoff, H.H. (1996) Bayesian decision theory and psychophysics. (pp. 123–161) In *Perception as Bayesian Inference* Knill, D.C. and Richards, W., (Eds), Cambridge University Press.