

Search Goal Tunes Visual Features Optimally

Vidhya Navalpakkam^{1,*} and Laurent Itti^{2,*}¹ Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA² Computer Science Department, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA*Correspondence: vidhya@klab.caltech.edu (V.N.), itti@usc.edu (L.I.)

DOI 10.1016/j.neuron.2007.01.018

SUMMARY

How does a visual search goal modulate the activity of neurons encoding different visual features (e.g., color, direction of motion)? Previous research suggests that goal-driven attention enhances the gain of neurons representing the target's visual features. Here, we present mathematical and behavioral evidence that this strategy is suboptimal and that humans do not deploy it. We formally derive the optimal feature gain modulation theory, which combines information from both the target and distracting clutter to maximize the relative salience of the target. We qualitatively validate the theory against existing electrophysiological and psychophysical literature. A surprising prediction is that it is sometimes optimal to enhance nontarget features. We provide experimental evidence toward this through psychophysics experiments on human subjects, thus suggesting that humans deploy the optimal gain modulation strategy.

INTRODUCTION

It is well known that attention is guided to both stimulus-driven (bottom-up salient [Itti and Koch, 2001a]) and goal-driven (top-down relevant [Hopfinger et al., 2000]) locations and features (Moran and Desimone, 1985; Motter, 1994; Treue and Martinez Trujillo, 1999). Yet, the mechanisms by which top-down relevance of features are determined and combined with bottom-up salience are relatively unknown. Below, we address one such outstanding question in the context of visual search.

Imagine that you are on a safari. The guide cautions you to beware of tigers hiding in the grasslands. Which visual features will you enhance or suppress in order to quickly detect a tiger? Enhancing the typical yellow color of a tiger's skin might seem like a good strategy. Indeed, previous research (Treue and Martinez Trujillo, 1999; Motter, 1994; Chelazzi et al., 1993; Martinez-Trujillo and Treue, 2004; Wolfe et al., 2004; Vickery et al., 2005) in top-down attention suggests that attention enhances the neural representation of the target-defining features. For

instance, the feature similarity gain model (Treue and Martinez Trujillo, 1999) suggests that gains increase as a function of similarity between the neuron's preferred feature and the target feature. While this may be true in simple scenes where there is no background clutter or the target and distractor features are very different, it may not apply to more complex scenes where the distractor features are similar to the target. Here, we investigate the optimal gain modulation strategy and ask whether humans deploy it. Understanding human feature selection strategies is not only crucial for further progress in understanding top-down attention, but may help in designing better robots and machines for active vision.

Related Work

In this section, we present a brief overview of the relevant visual search literature. The "biased competition" hypothesis suggests that multiple stimuli compete in a mutually suppressive manner to gain access to the limited resources (such as representation, analysis, control) and attention biases this competition toward the salient and behaviorally relevant locations or features. Although the details of the amount of top-down feature bias are not formally specified, the general idea is that visual inputs that match the target description (or "attentional template" [Duncan and Humphreys, 1989]) are favored in the visual cortex (Bundesen, 1990). In other words, the top-down competitive bias toward a stimulus depends on its similarity to the "attentional template," thereby yielding a stronger competitive bias toward the target than distractors that resemble it or distractors that are dissimilar (Desimone and Duncan, 1995). This theory has received much support from the neurophysiology of spatial (Luck et al., 1997; Reynolds et al., 1999; Kastner et al., 1999) and object-based attention (Chelazzi et al., 1993). Several neurodynamic implementations of the biased competition hypothesis have also been proposed (Deco and Rolls, 2002; Hamker, 2004).

In addition to a spatial bias, recent studies have shown strong feature-based attentional modulation effects that are spatially global and occur throughout the visual field (Treue and Martinez Trujillo, 1999; Saenz et al., 2002). These observations led to an elegant "feature similarity gain" model, where attention causes a multiplicative change in the response gain of a neuron that depends on the similarity between its preferred feature (or location) and the attended feature (or location). This theory has

recently received more experimental support (Martinez-Trujillo and Treue, 2004; Bichot et al., 2005).

Cave (1999) proposed a neural network implementation of the guided search model (Wolfe, 1994) that combines both bottom-up and top-down influences. It consists of a hierarchy of spatial feature maps and the flow of information is selectively gated from lower to higher levels of the visual hierarchy. The top-down bias is applied by opening (or closing) gates at each level, depending on the similarity (or dissimilarity) between the target features and the features at that location. Thus, the top-down component of this model enhances locations whose features are similar to the target.

Tsotsos et al. (1995) suggest that attention to a stimulus (location or feature) causes selective tuning by triggering a cascade of top-down winner-take-all selection processes along the visual hierarchy. The attended stimulus (or most salient or task-relevant stimulus) is selected at the top and at the subsequent WTA selection at the lower stages, the neural input that contributes most to the attended stimulus is selected, and irrelevant signals that interfere are eliminated. Thus, attention causes selective tuning to the attended stimulus. The model includes a task-specific executive controller that selects the task-relevant feature at the top. While the details of the task-specific feature bias are not specified, they suggest that the working memory may store a target template and the WTA selection may activate stimuli that resemble the target.

Several other models have been proposed. Hamker (2004) suggests that prefrontal areas might store a target template. Feedback connections from prefrontal to IT (and from IT to V4) may enhance the activity of neurons whose visual input matches the target template. As a result of the re-entry signals, locations whose features are similar to the target are enhanced, while others are suppressed. Rao et al. (2002) proposed a saliency model to explain eye movements during visual search. In their model, saliency was computed as the euclidean distance between a target template (memorized vector of responses to the target stimulus) and responses at each location.

Several models of top-down attention have been proposed earlier, and all of them include a top-down biasing or feature selection process that enhances features that are similar to the target. In the rest of this paper, we investigate whether this target-similarity-based feature selection strategy is optimal. We formally derive the optimal top-down feature biasing strategy and contrast it to the above target-similarity-based approaches.

Model

We formally derive a theory of how prior statistical knowledge of the target and distractor features optimally influences feature gains. From a theoretical standpoint, gains must be modulated in order to maximize search speed, which is a function of at least two critical variables: $S_T(A)$, the mean perceived saliency of target instances in the display A (formed as a result of combined top-down

and bottom-up influences); and $S_D(A)$, the mean perceived saliency of distractor instances. The relative values of $S_T(A)$ and $S_D(A)$ determine visual search efficiency (Itti and Koch, 2000; Wolfe et al., 2003). Hence, the relevant goal for optimizing top-down gains is to maximize the signal-to-noise ratio (SNR), i.e., to maximize the ratio between signal strength (target saliency) and noise strength (distractor saliency). Such optimization renders the target more salient than distractors in the display, thereby attracting attention (Koch and Ullman, 1985) and decreasing the search time (Wolfe et al., 2003).

Later, we compare the results obtained by setting gains according to different objective functions, such as maximizing discriminability between saliency of the target and distractor versus maximizing SNR.

A Theory of Optimal Feature Gain Modulation

$S_T(A)$ and $S_D(A)$ are random variables that depend on the top-down gains as well as the following bottom-up factors: (1) values of target and distractor features $\Theta|T$ and $\Theta|D$ in the display [sampled from probability density functions $p(\Theta|T)$ and $p(\Theta|D)$ and possibly corrupted by external noise], (2) spatial configuration C of target and distractor items in the display, and (3) internal noise in neural response, η . Thus, $SNR = E_{\Theta|T,C,\eta}[S_T(A)]/E_{\Theta|D,C,\eta}[S_D(A)]$.

We formulate the optimal theory within the framework of a “consensus model” based on current evidence in neurobiology and psychophysics (Treisman and Gelade, 1980; Koch and Ullman, 1985; Wolfe, 1994; Treue and Martinez Trujillo, 1999; Saenz et al., 2002) (Figure 1). The visual input is analyzed in different feature dimensions (e.g., color, orientation, direction of motion). For clarity, we focus on one dimension at a time. The results can be generalized across multiple dimensions. We assume that each dimension is encoded by a population of n neurons with overlapping tuning curves tuned to different feature values (Deneve et al., 1999). The i th neuron ($i \in \{1 \dots n\}$) is tuned to feature value μ_i , and its output is used to compute the bottom-up saliency (Itti and Koch, 2001b) $s_i(x, y, A)$ at location (x, y) in search array A . The overall perceived saliency, S for a feature dimension is then computed as a function of the saliencies s_i for feature values within that dimension. While many functions are possible, one of the simplest functions consistent with existing data is a linear combination of s_i (Itti and Koch, 2001b), weighted in a top-down manner by multiplicative gains g_i (Hillyard et al., 1998):

$$S(x, y, A) = \sum_{i=1}^n g_i s_i(x, y, A) \quad (1)$$

Thus, the saliency map for a dimension is computed as a weighted sum of saliency maps from all feature values and is used to guide attention. The saliency of the target (S_T) can be computed as follows:

$$E[S_T(A)] = E_{\Theta|T,C,\eta} \left[\sum_{i=1}^n g_i s_{iT}(A) \right] \quad (2)$$

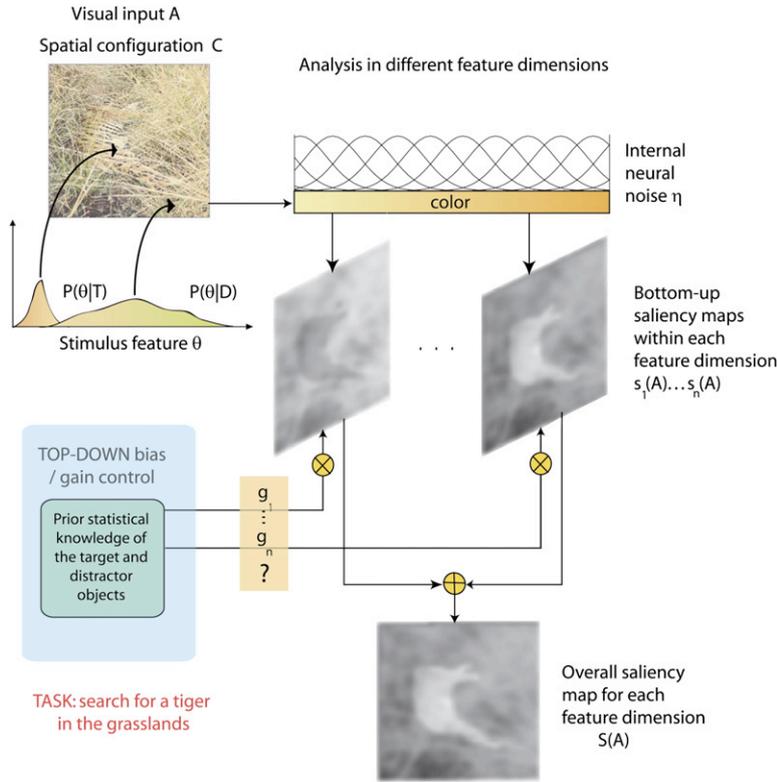


Figure 1. Overview of Our Model

The incoming visual scene A is analyzed in several feature dimensions (e.g., color and orientation) by populations of neurons with bell-shaped tuning curves. For clarity, we show just one dimension here. Within each dimension, bottom-up saliency maps ($s_1(A) \dots s_n(A)$) are computed for different feature values and combined in a weighted linear manner to form the overall saliency map ($S(A)$) for that dimension. Given this model, how do we choose the optimal set of top-down gains ($g_1 \dots g_n$) such that the target tiger becomes most salient among distracting clutter? Our theory shows that the intuitive choice of looking for the tiger's yellow feature would actually be suboptimal, because this would activate the distracting grassland more than the tiger. Instead, the optimal strategy would be to look for orange, which is mildly present in the tiger, but hardly present in the grasslands, and hence best differentiates between the target and the distracting background.

$$= \sum_{i=1}^n g_i E_{\Theta|T} [E_C [E_{\eta} [s_{iT}(A)]]] \quad (3)$$

(since η , C , and Θ are independent random variables) (4)

$$E[S_{iD}(A)] = \sum_{i=1}^n g_i E_{\Theta|D} [E_C [E_{\eta} [s_{iD}(A)]]] \quad (\text{similarly}) \quad (5)$$

Thus, we have,

$$SNR = \frac{\sum_{i=1}^n g_i E_{\Theta|T} [E_C [E_{\eta} [s_{iT}(A)]]]}{\sum_{i=1}^n g_i E_{\Theta|D} [E_C [E_{\eta} [s_{iD}(A)]]]} \quad (6)$$

To maximize SNR , we differentiate it with regard to g_j .

$$\frac{\partial}{\partial g_j} SNR = \frac{\frac{E_{\Theta|T} [E_C [E_{\eta} [s_{iT}(A)]]]}{E_{\Theta|D} [E_C [E_{\eta} [s_{iD}(A)]]]} - \frac{\sum_{i=1}^n g_i E_{\Theta|T} [E_C [E_{\eta} [s_{iT}(A)]]]}{\sum_{i=1}^n g_i E_{\Theta|D} [E_C [E_{\eta} [s_{iD}(A)]]]}}{\frac{\sum_{i=1}^n g_i E_{\Theta|D} [E_C [E_{\eta} [s_{iD}(A)]]]}{E_{\Theta|D} [E_C [E_{\eta} [s_{iD}(A)]]]}}} \quad (7)$$

$$= \frac{SNR_i}{SNR} - 1 \quad (8)$$

$$= \alpha_i$$

where α_i is a normalization term and $SNR_i = E_{\Theta|T} [E_C [E_{\eta} [s_{iT}(A)]]] / E_{\Theta|D} [E_C [E_{\eta} [s_{iD}(A)]]]$. It is easy to show that g_i/g_{i0} (where $g_{i0} = 1$ is the default baseline gain) increases as SNR_i/SNR increases. With an added constraint that the gains must sum to a constant,

$$\sum_{i=1}^n g_i = n,$$

the simplest solution is

$$g_i = \frac{SNR_i}{\frac{1}{n} \sum_{j=1}^n SNR_j} \quad (9)$$

Thus, the top-down gain on a visual feature depends on its signal-to-noise ratio (SNR_i).

The above theory assumes an ideal observer who knows the true distribution of target and distractor features $[p(\Theta|T), p(\Theta|D)]$. Instead, a real observer may possess incomplete knowledge or a belief in the likely target and distractor features $[p(\Theta^b|T), p(\Theta^b|D)]$. This belief may be learned from a preview of picture cues (Wolfe et al., 2004; Vickery et al., 2005), verbal instructions (e.g., search for a "red" item) (Wolfe et al., 2004), or from observations of past trials (Maljkovic and Nakayama, 1994) (see Figure 2). In such cases, we assume that the observer can use an internal model to translate his/her belief in features into a belief in salience of the target and distractors S_T^b, S_D^b . In this extended framework, it is easy to show that the other derivations remain identical, i.e., gains can be chosen so as to maximize SNR^b (SNR derived from top-down belief). The overall framework that integrates bottom-up salience with top-down beliefs is shown in Figure 2.

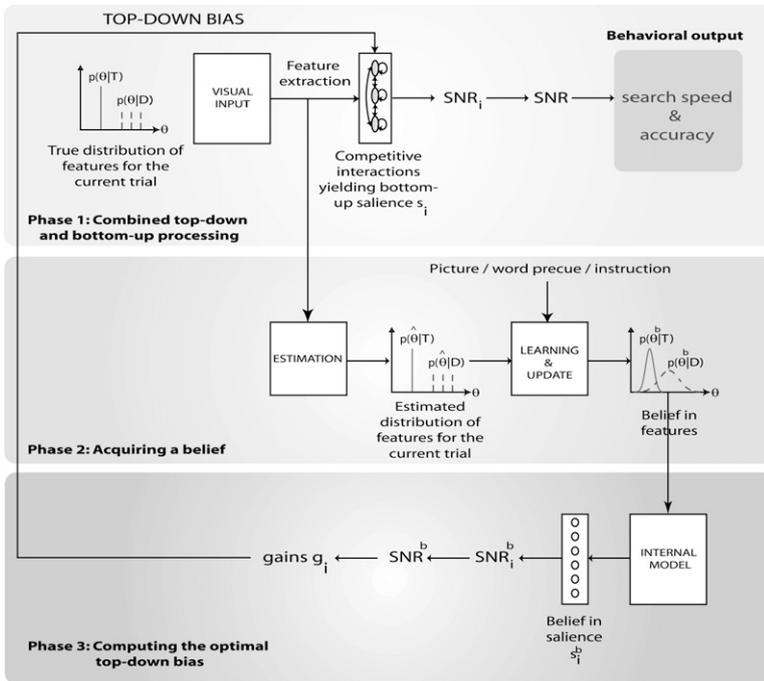


Figure 2. Three Phases of Visual Search

Phase 1: Combined bottom-up and top-down processing of the visual input. The top-down gains (phase 3) derived from the observer's beliefs (phase 2) are combined with bottom-up saliency computations to yield the overall saliency of the target and distractors. This determines search performance, measured by SNR . Phase 2: Acquiring a belief. The distributions of target and distractor features may be learned through estimation from past trials, preview of picture cues, verbal instructions, or other means. Phase 3: Generating the optimal top-down gains. The learned belief in target and distractor features is translated into a belief in saliency of the target and distractors, thus yielding SNR^b . The top-down gains are chosen so as to maximize SNR^b .

RESULTS

In this section, we report the theory's predictions on various search conditions through numerical simulations on networks of neurons encoding features of the target and distractors. Subsequently, we test novel predictions of the theory through psychophysics experiments on human participants.

Simulating Visual Search Conditions

To test the optimal feature gain modulation strategy, we perform detailed numerical simulations. For different search conditions and displays, we compute the bottom-up saliency of the target and distractors S_{IT} , S_{ID} as a function of the true distribution of the target and distractor features $p(\theta|T)$, $p(\theta|D)$ using the saliency computations proposed by Itti and Koch (2001b). Next, we apply the optimal top-down gains g_i derived from the observer's belief $p(\theta^b|T)$, $p(\theta^b|D)$ on the bottom-up saliency maps (s_i). Then we compute the overall saliency, S_T , S_D , and the overall signal-to-noise ratio, SNR (Figure 2). The resulting SNR may be high, and search may be efficient due to high bottom-up saliency of the target relative to the distractors (e.g., a red target pops out among green distractors (Treisman and Gelade, 1980) as $s_{IT} \gg s_{ID}$ in the saliency map tuned to the red feature) or due to efficient top-down guidance to the target (e.g., a red target among randomly colored distractors becomes easy to find once subjects know that the target is red (Duncan, 1989) since $g_i \gg 1$ on the red feature) or both.

Figure 3 shows the results of our simulations for different search conditions. Figures 3A–3C together show that

for a given target and distractor stimulus better prior knowledge of their features (or decreased uncertainty) allows the relevant features to be primed, thus leading to higher SNR and faster search. These results are in qualitative agreement with existing psychophysics literature on the role of uncertainty in target features (Wolfe et al., 2004; Vickery et al., 2005) and the role of feature priming (Shiffrin and Schneider, 1977; Maljkovic and Nakayama, 1994; Wolfe et al., 2003). Figure 3D shows that knowledge of the target (only) improves SNR by enhancing target features. Evidence for such target-based enhancement has been observed in single-unit recordings in MT and is consistent with the feature similarity gain model (Treue and Martinez Trujillo, 1999). In addition, psychophysics studies provide evidence that knowledge of the target accelerates search performance (Vickery et al., 2005). Figure 3E predicts that knowledge of the distractor also improves search by suppressing the distractor features. Partial experimental evidence comes from studies that show decreased responses to the distractor feature (in MT [Martinez-Trujillo and Treue, 2004], in FEF [Bichot and Schall, 2002]) and from psychophysics studies that show a benefit in search performance due to knowledge of distractors (Maljkovic and Nakayama, 1994; Braithwaite and Humphreys, 2003).

Figures 3C and 3F together demonstrate the effect of distractor heterogeneity (Duncan and Humphreys, 1989), i.e., search efficiency decreases as the number of types of distractors increases (e.g., searching for a red target among blue, green, yellow, and white distractors is harder than searching for a red target among green distractors). Consistent with this effect, our simulations show that

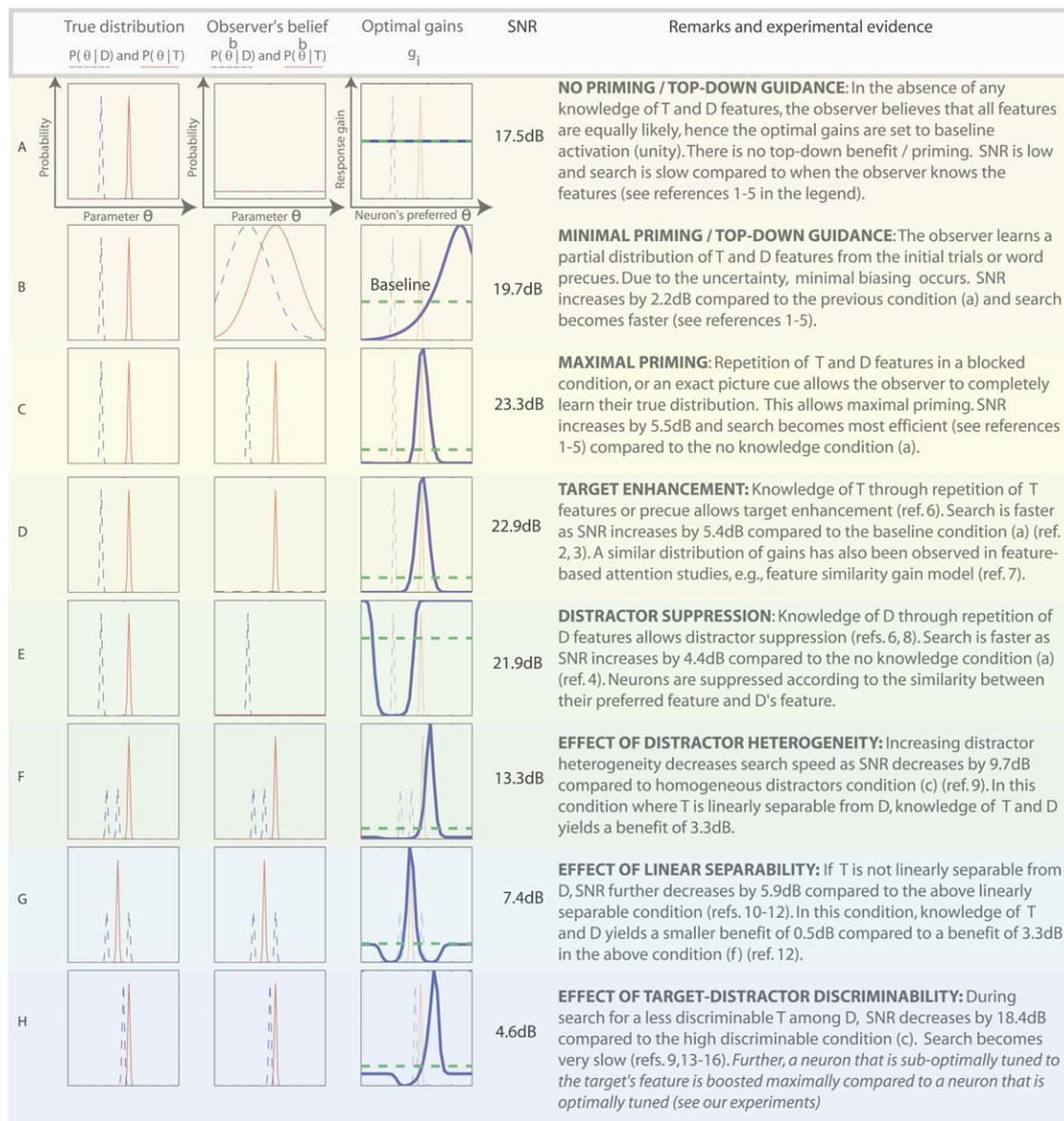


Figure 3. Simulation Results for a Variety of Search Conditions Shown in Different Rows

The first column shows the true distribution of the target (T) features $p(\theta|T)$, solid red] and distractor (D) features $p(\theta|D)$, dashed blue], and the second column shows the observer's belief $[p(\theta^b|T), p(\theta^b|D)]$. The third column shows the optimal distribution of neural response gains superimposed over $p(\theta|T), p(\theta|D)$. The fourth column shows SNR followed by the implications of our results, along with experimental evidence. For example, row (A) illustrates how lack of prior knowledge prevents any top-down guidance of search. Let the true distributions $p(\theta|T)$ and $p(\theta|D)$ peak at different values, e.g., red target among green distractors. When T and D are unknown, the beliefs $p(\theta^b|T), p(\theta^b|D)$ are a uniform distribution with all features being equally likely. Hence, the optimal gains are set to baseline ($g_i = 1, i \in \{1 \dots n\}$). Remarks and supporting experimental evidence for the remaining search conditions (A-H) are shown in the fifth column in this figure. Our theory is able to formally predict several effects in visual search behavior which have been previously studied empirically. References: 1, Wolfe et al., 2004; 2, Vickery et al., 2005; 3, Wolfe et al., 2003; 4, Maljkovic and Nakayama, 1994; 5, Shiffrin and Schneider, 1977; 6, Bichot and Schall, 2002; 7, Treue and Martinez Trujillo, 1999; 8, Braithwaite and Humphreys, 2003; 9, Duncan and Humphreys, 1989; 10, D'Zmura, 1991; 11, Bauer et al., 1996; 12, Hodsoll and Humphreys, 2001; 13, Wolfe, 1994; 14, Pashler, 1987; 15, Nagy and Sanchez, 1990; 16, Treisman, 1991.

SNR decreases from 23.0 dB (Figure 3C, homogeneous distractors) to 13.3 dB (Figure 3F, heterogeneous distractors), resulting in slower search due to increased distractor heterogeneity.

A comparison of Figures 3F and 3G reveals the linear separability effect, i.e., search for a target flanked by distractor features (as shown in Figure 3G) is harder than search for a target that is linearly separable from

distractors in feature space (as shown in Figure 3F). This effect has been demonstrated in features such as size, chromaticity, and luminance (Hodsoll and Humphreys, 2001; D’Zmura, 1991; Bauer et al., 1996). For example, search for a medium sized target among small and big distractors is known to be harder than search for a big target among small and medium sized distractors (Hodsoll and Humphreys, 2001). Our simulation results are consistent with this effect and show a decline in SNR from 13.3 dB (Figure 3F, linearly separable target) to 7.4 dB (Figure 3G, target that is not linearly separable). Furthermore, in agreement with psychophysics (Hodsoll and Humphreys, 2001), our simulations reveal a greater top-down benefit of knowing the target and distractors in the linearly separable condition (3.3 dB in Figure 3F) than otherwise (0.5 dB in Figure 3G).

One of the classic effects in visual search behavior is that search efficiency decreases as target-distractor discriminability decreases (Pashler, 1987; Duncan and Humphreys, 1989; Nagy and Sanchez, 1990; Treisman, 1991; Wolfe, 1994). Figures 3C and 3H demonstrate this effect. While SNR is high (23.0 dB) when the target and distractor features are very different (e.g., 55° oriented target among 25° oriented distractors, as shown in Figure 3C), SNR drops to as low as 4.6 dB when the target and distractor features are similar (e.g., 55° oriented target among 50° oriented distractors, as shown in Figure 3H).

Psychophysics Experiments

Notably, our theory makes a new prediction that, during search for a less discriminable target among distractors, an exaggerated target feature is promoted more than the exact target feature (see Figure 3H). Though seemingly counterintuitive, this occurs because a neuron that is tuned to an exaggerated target feature provides higher SNR_i (as it responds much more to the target than the distractor), whereas a neuron that is tuned to the exact target feature provides lower SNR_i (as it responds similarly to the target and distractor). This is shown in Figure 4. To validate this claim, we conducted new psychophysics experiments that were designed in two phases: (1) to set up the top-down bias and (2) to measure the bias.

To set up the top-down gains, we asked subjects to perform the primary task T₁, which is a hard visual search for the target (55° tilted line) among several distractors (50° tilted lines). A typical T₁ trial is shown in Figure 5A: it starts with a fixation, followed by the search array. Upon finding the target among distractors, subjects press any key. To ensure that subjects would bias for the target among distractors in each and every trial, we introduce a *No Cheat* scheme (see legend of Figure 5A). Subjects are trained on T₁ trials until their performance stabilizes with at least 80% accuracy. Thus, the top-down bias is set up by performing T₁ trials.

To measure the top-down gains generated by the above task, we randomly insert T₂ trials in between T₁ trials (Figure 5A). Our theory predicts that during search for the target (55°) among distractors (50°), the most relevant

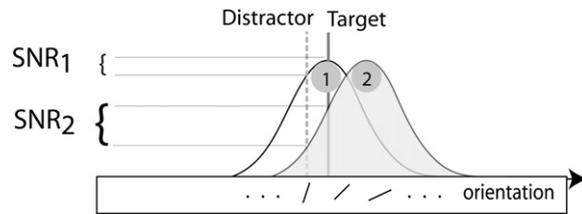


Figure 4. Boosting a Neuron Tuned to an Exaggerated Target Feature Helps in a Difficult Search Task

When the target feature (shown by a solid vertical line) is similar to the distractor feature (shown by a dotted vertical line), neuron 2, which is tuned to an exaggerated feature, provides higher SNR_i than neuron 1, which is tuned to the exact target feature.

feature will be around 60° and not 55°. To test this, we ask subjects to “find the target” in a brief display (300 ms) of five items representing five different features: steepest (80°), relevant as predicted by our theory (R, 60°), target (T, 55°), distractor (D, 50°), and shallowest (30°). The display is brief, and its occurrence is unpredictable in order to minimize any alteration in the top-down gains set up by the T₁ trials. If the top-down gain on a feature is higher than other features, then it should appear more salient, draw attention, and hence be reported. Thus, although subjects search for the target, our theory predicts a higher number of reports on the relevant feature R than on the target feature T (since R has a higher top-down bias than T).

Experimental results across all subjects indicate significantly ($p < 0.05$) higher number of reports on R than on T (Figure 5B). As predicted by our theory, subjects could not help but be attracted toward R, although the task was to search for T. In additional controls, when the distractor feature was reversed (60°) while the target remained the same (55°), the same subjects showed a reversal in the trend of biasing (described in Figure 5C). Similar results were obtained in the color dimension as well (see Figure 6). Our results provide experimental evidence that humans may deploy optimal top-down feature gain modulation strategies.

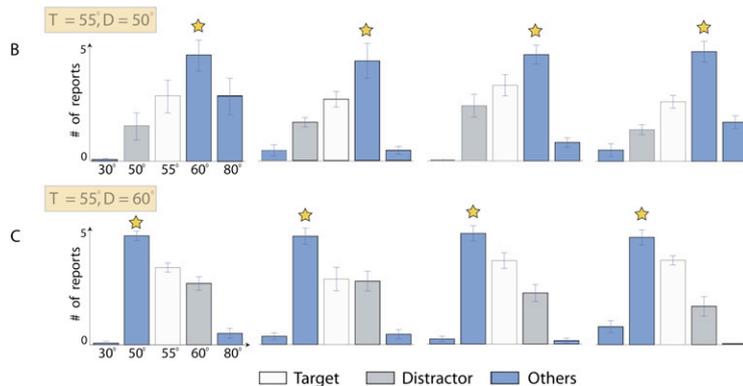
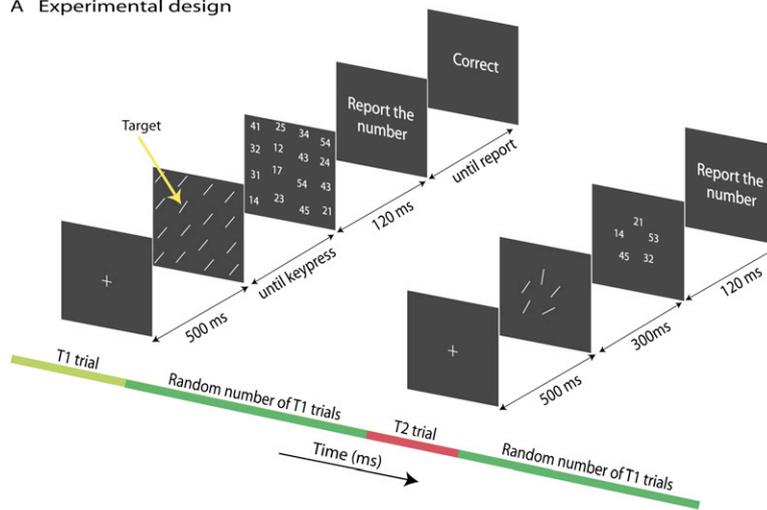
Alternative Objective Functions

We have shown that a simple function such as the ratio of expected salience of the target over the distractors is sufficient to account for most visual search data. For a fixed ratio of means, when the target and distractor feature distributions are narrow, as shown in Figures 3B and 3C, SNR increases compared to when the feature distributions are wide. Thus, variance in target and distractor features is implicitly encoded in the population code of SNR. In Figure 7, we compare our SNR measure against D' , which is the discriminability between the salience of the target and distractor, defined as follows:

$$D' = \frac{E[S_T(A)] - E[S_D(A)]}{\sqrt{0.5(V[S_T(A)] + V[S_D(A)])}} \quad (10)$$

where $V[\cdot]$ refers to the variance. The gains that maximize D' are derived in the section on [Experimental Procedures](#).

A Experimental design



As shown in Figure 7, given our assumption of normalizing gains, our *SNR* measure effectively captures psychophysical behavior in several search conditions, while *D'* fails in some cases (see the Supplemental Data available with this article online). This suggests that *SNR* is the relevant objective function to be optimized for improving visual search behavior.

DISCUSSION

Several theories of visual search have been proposed in the past—while some attempt to explain the behavior of the organism (e.g., feature integration theory, guided search theory), others attempt to account for the single-unit responses (e.g., feature similarity gain model, feature matching hypothesis). Here, by modulating the gains such that behavioral performance (quantified in terms of *SNR*) is optimized, we provide a simultaneous account of the search behavior of the organism as well as neural gains at the single-unit level. Specifically, we suggest that gains

Figure 5. Psychophysics Experiment to Test Optimal Biasing in the Orientation Dimension

(A) Experimental design. We test the theory's prediction of top-down bias during search for a low-discriminability target among distractors (Figure 3H). The top-down bias is set when subjects perform T_1 trials. After a random number of T_1 trials, the top-down bias is measured in a T_2 trial. A T_1 trial consists of a fixation followed by a search array containing one target (55°) among several distractors (50°). Subjects are instructed to report the target as soon as possible. Subjects' responses are validated on a per-trial basis through a novel *No Cheat* scheme that is described in the main text. A T_2 trial consists of a fixation, followed by a brief display of five items representing five features, and by five fingerprint random numbers. Subjects are asked to report the number at the target location.

(B) Experimental results. We ran four subjects (three naive), aged 22–30, normal or corrected vision, with IRB approval. The T_2 trials were analyzed to find the number of reports (mean \pm SD) on 30° , 50° , 55° , 60° , and 80° features. The number of reports on the relevant feature (60° , marked by a golden star) is significantly higher (paired *t* test, $p < 0.05$) than the number of reports on the target feature (55°).

(C) Controls. In a control experiment, we maintained the same target feature, but reversed the distractor feature. In the T_1 trials, the same subjects now searched for the 55° oriented target among 60° oriented distractors. Everything else, including the T_2 trials, instructions, and analysis remained the same. Statistical analysis of number of reports showed a reversal in trend compared to (B), with significantly higher number of reports on the currently relevant feature (50° , marked by a golden star) than the target feature (55°).

are modulated so as to optimize the salience of the target relative to the distractors (which we refer to as the signal-to-noise ratio, *SNR*). Such optimization of *SNR* increases both search accuracy and speed. The theory makes a number of testable predictions at the single-unit and behavioral level and bears implications for electrophysiology, brain imaging, and psychophysics of visual search.

While several models of attention have been proposed in the past, most of them include a top-down component that biases features according to their similarity to the target (Desimone and Duncan, 1995; Deco and Rolls, 2002; Hamker, 2004; Treue and Martinez Trujillo, 1999; Boynton, 2005; Cave, 1999; Tsotsos et al., 1995; Rao et al., 2002). For instance, one of the prominent models, “the feature similarity gain model,” suggests that the gain on a neuron encoding a visual feature depends on the similarity between the neuron's preferred feature and the target feature. We show that this is a special case of our general theory, which occurs whenever the target feature differs substantially from the distractor feature. Thus, previous

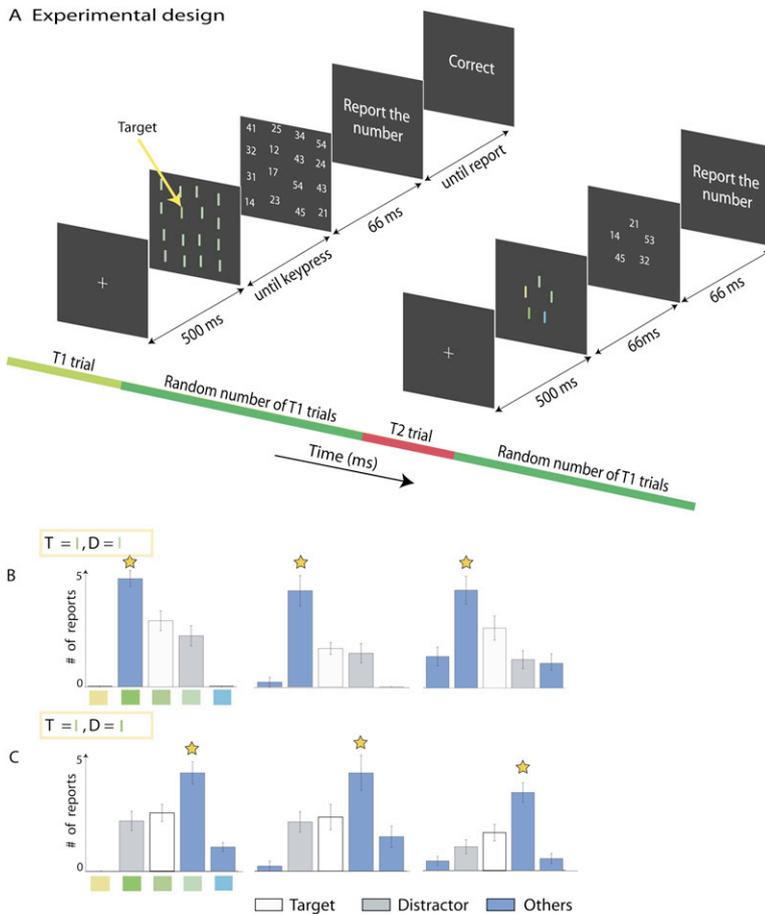


Figure 6. Psychophysics Experiments to Test Optimal Biasing in the Color Dimension

(A) Experimental design. We test the theory's prediction of top-down bias in the color dimension. The experimental design is similar to Figure 5. The target has medium green hue (CIE $x = 0.24$, $y = 0.42$), while the distractor is either more green ($x = 0.25$, $y = 0.45$, Figure 6B) or less green ($x = 0.23$, $y = 0.38$, Figure 6C), and the irrelevant controls are yellow ($x = 0.42$, $y = 0.50$) and blue ($x = 0.21$, $y = 0.27$). The presentation time of the T_2 probe trials is brief (66 ms). (B) Experimental results. We ran three subjects (naive), aged 22–30, normal or corrected vision, with IRB approval. The T_2 trials were analyzed to find the number of reports (mean \pm SD) on the yellow, more green, medium green, less green, and blue features. When subjects searched for a medium green target among less green distractors, as predicted by the theory, there were significantly more reports (paired t test, $p < 0.05$) on the more green feature than the target feature.

(C) Controls. In a control experiment, we maintained the same target feature, but reversed the distractor feature. Now, subjects searched for a medium green target among more green distractors. Statistical analysis of the number of reports showed a reversal in trend compared to (B), with a significantly higher number of reports on the less green feature than the target feature. These results in the color and orientation dimensions support optimal feature biasing as suggested by our theory.

experiments with different target and distractor features or absence of distractor features (e.g., experiments by Bichot et al. [2005] in the color dimension in FEF, Treue and Martinez Trujillo [1999] in direction of motion in MT) that provide evidence for the feature-similarity gain model also provide evidence for our theory. In addition, we show examples of search conditions when the former strategy of enhancing target features is suboptimal. For instance, when the target and distractor features are similar (e.g., 5° difference in orientation), neurons tuned to the target respond to the distractor as well (providing lower SNR_i), hence enhancing such neurons increases the response to the distractor, which is undesirable for performance. On the other hand, a neuron that is tuned to an exaggerated target feature responds much more to the target relative to the distractor and provides higher SNR_i than a neuron that is tuned to the exact target feature. Hence, the optimal strategy is to boost a neuron tuned to the exaggerated target feature and not the exact target feature. This effect has also been reported in discrimination tasks where a neuron tuned to an exaggerated stimulus feature contains higher fisher information than a neuron that is tuned to the exact stimulus feature (Lee et al., 1999). To the best of our knowledge, this is the first study to demonstrate a similar effect during visual search.

Here, we summarize the differences between our model and previous models. (1) Most previous models ignore the role of the distractor in determining gain modulation. They enhance features that are similar to the target. On the other hand, we predict that the distractor plays a critical role and determines whether the target feature will be enhanced or not. (2) In several earlier models (e.g., Feature-Gate, feature similarity gain, Hamker's model, Rao's model), the top-down bias only works when the target features are known. They cannot predict the top-down bias when distractor features are known but the target is unknown (e.g., when the distractor feature does not change across search trials but the target feature changes). Our model predicts the top-down bias for all combinations of knowledge of target and distractor features, including when the target is unknown but the distractor is known (or trivially, when both the target and distractor are unknown, in which case the gains remain at default values). (3) While most previous models are either purely top-down (Rao et al., 2002) or bottom-up driven (Li, 2002), a key distinguishing aspect of our model is that it integrates both bottom-up salience and top-down feature bias.

By applying optimal top-down gains on bottom-up salience responses, our theory integrates both goal-driven, top-down and stimulus-driven, bottom-up factors to

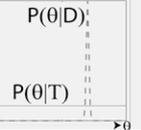
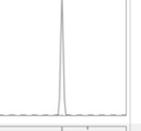
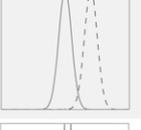
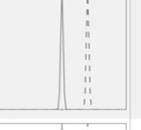
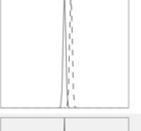
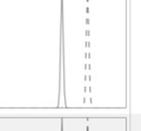
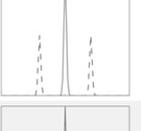
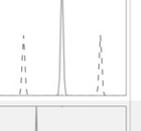
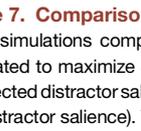
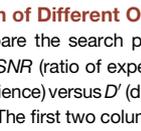
	Difficult vs. Easier search		$\frac{SNR_{easy}}{SNR_{hard}}$	$\frac{D'_{easy}}{D'_{hard}}$
				
A			1.25 ✓	14.01 ✓
B			1.05 ✓	3.56 ✓
C			1.18 ✓	4.18 ✓
D			4.87 ✓	3.65 ✓
E			1.58 ✓	0.57 ✗
F			3.07 ✓	0.93 ✗
G			1.86 ✓	0.65 ✗

Figure 7. Comparison of Different Objective Functions

These simulations compare the search performance when gains are modulated to maximize SNR (ratio of expected target saliency relative to expected distractor saliency) versus D' (discriminability between target and distractor saliency). The first two columns illustrate different search conditions [each denoted by a particular distribution of target feature $P(\theta|T)$ shown in solid red, and distractor feature $P(\theta|D)$ shown in dotted blue]. According to previous psychophysics studies, the search condition illustrated in the first column is known to be more difficult than its counterpart in the second column. While maximizing SNR successfully accounts for this difference (as shown in the third column, ratio of SNR values in easier versus difficult conditions >1), maximizing D' fails in some cases [as shown in the fourth column, in (E)–(G), ratio of $D' < 1$]. This validates our choice of SNR as the relevant objective function.

guide visual attention. It successfully accounts for a large body of available visual search literature. For instance, it accounts for several reported knowledge-based effects such as the role of uncertainty in target features (Wolfe et al., 2004; Vickery et al., 2005), role of feature priming (Shiffrin and Schneider, 1977; Maljkovic and Nakayama, 1994; Wolfe et al., 2003), target enhancement and distractor suppression (Bichot and Schall, 2002; Braithwaite and Humphreys, 2003), and top-down effects on linear separability (Hodsoll and Humphreys, 2001).

It also demonstrates other well known bottom-up effects such as the role of target-distractor discriminability (Pashler, 1987; Duncan and Humphreys, 1989; Nagy and Sanchez, 1990; Treisman, 1991; Wolfe, 1994), distractor heterogeneity (Duncan and Humphreys, 1989), and linear separability (D'Zmura, 1991; Bauer et al., 1996). Thus, the theory, despite being simple, yields good predictive power. It is general and applicable to top-down selection of relevant information in biological as well as artificial systems, in visual and other modalities, including auditory, somatosensory, and cognitive.

Could the observed behavioral response of subjects (in Figures 5 and 6) reflect higher decision processes rather than attentional biasing? Indeed, subjects' responses in psychophysics studies such as ours is the outcome of several visuo-motor transformations from the early and intermediate visual areas to higher decision areas. However, it is unlikely that our results reflect decision-making processes for the following reasons. The presentation time of our probe trials is brief (66 ms in the experiments on color, 300 ms for orientation) and prevents scanning of all five items before reporting the target. The briefness of probe trials minimizes the contribution of covert serial recognition or decision processes, so that the subjects' responses may reflect fast attentional biasing processes rather than slow recognition or decision processes. Further validation of attentional biasing and the theory's predictions on gain modulation calls for more studies in electrophysiology.

So far, gain modulation has been studied systematically only for one configuration: when the target feature is known (Treue and Martinez Trujillo, 1999; Martinez-Trujillo and Treue, 2004). A feature similarity gain model was proposed to account for the observations. Here, we show that the feature similarity gain model can be explained as a special case of our general theory. Our theory agrees with the predictions of the feature similarity gain model under the condition that the target and distractor features are very different. In addition, we predict that the distribution of gains will be skewed away from the target and distractor feature when they are very similar. Indeed, natural scenes are full of clutter, and it is common for targets of interest (e.g., prey, predators, suspects, etc.) to be camouflaged or embedded in distracting backgrounds. We predict that in such cases the distractor feature (and not just the target feature) will play a critical role in gain modulation. We have empirically verified this on natural scenes (Navalpakam and Itti, 2006), where the optimal gain modulation strategy based on the target and distractor features performs better than one which considers target features only. This prediction remains to be tested neurally.

To summarize, we have proposed a theory of neural function that suggests that the "end result" of feature-based attention, possibly mediated through complex neural interactions and feature processing, is to modulate neural response gains according to their signal-to-noise ratio. The details of the neural mechanisms in the

intermediate steps are not yet addressed by the theory. The functional role of attention suggested by the theory is general and applicable to any population of neurons that encode a continuous feature dimension in a distributed manner, e.g., neurons in MT that are tuned to direction of motion, V4 neurons that are tuned to orientation. For simple feature dimensions such as orientation that we have currently tested in our psychophysics experiments, we suggest that the attentional modulation may occur as early as in a V1 hypercolumn (Motter, 1993; Roelfsema et al., 1998; McAdams and Maunsell, 1999; Watanabe et al., 1998; Somers et al., 1999; Gandhi et al., 1999; Martinez et al., 1999; W.A. Press and D.C. van Essen, 1997, Soc. Neurosci, abstract).

The current report primarily focuses on gain modulation within a single feature dimension. This provides a theoretical foundation for further research on integrating multiple feature dimensions. As shown elsewhere (Navalpakkam and Itti, 2006), this theory can be easily extended to multiple dimensions if they are combined linearly as suggested by the guided search theory (Wolfe, 1994).

By focusing on visual features as opposed to locations in space, our study on optimal feature gain modulation complements recent studies on optimal eye position strategies (Najemnik and Geisler, 2005). While the latter suggests that humans can select relevant locations optimally, here, we show that humans select visual features optimally as well. Together, these studies suggest that human visual search behavior is optimal.

EXPERIMENTAL PROCEDURES

Special Cases

Here, we derive analytical expressions for gains in some common visual search conditions. To simplify the expressions, we assume that the feature dimension is encoded by neurons with Gaussian tuning curves (f_i) whose preferred features (μ_i) vary continuously along the dimension. In the following equation, σ is the tuning width and a is the amplitude of firing rate, and b is the background firing rate.

$$f_i(\theta) = \frac{a}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\theta - \mu_i)^2}{2\sigma^2}\right\} + b \quad (11)$$

We further approximate saliency (s_i) by the raw neural response (r_i), which is a poisson random variable with mean response f_i .

$$E_{\Theta|T}[E_C[E_{\eta}[s_{iT}(A)]]] = E_{\Theta|T}[E_C[E_{\eta}[r_{iT}(A)]]] \quad (12)$$

$$= E_{\Theta|T}[E_C[f_{iT}(A)]] \quad (13)$$

$$= E_{\Theta|T}[f_{iT}(A)] \quad (14)$$

$$E_{\Theta|D}[E_C[E_{\eta}[s_{iD}(A)]]] = E_{\Theta|D}[f_{iD}(A)] \quad (\text{similarly}) \quad (15)$$

$$SNR_i = \frac{E_{\Theta|T}[f_{iT}(A)]}{E_{\Theta|D}[f_{iD}(A)]} \quad (16)$$

$$g_i = \frac{SNR_i}{\frac{1}{n} \sum_j SNR_j} \quad (17)$$

We derive the optimal gains when the target is known and consists of a single feature [$P(\Theta|T)$ is a Dirac Delta function], while the distractor

is unknown and may assume any feature with equal probability [$P(\Theta|D)$ is a uniform distribution].

$$P(\Theta|T) = \delta(\theta_t) \quad (18)$$

$$P(\Theta|D) = \frac{1}{\pi} \quad (19)$$

$$E_{\Theta|T}[f_{iT}(A)] = \int_{\Theta|T} f_i(\theta) P(\theta) d\theta \quad (20)$$

$$= \frac{a}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\theta_t - \mu_i)^2}{2\sigma^2}\right\} + b \quad (21)$$

$$E_{\Theta|D}[f_{iD}(A)] = \frac{a}{\pi} + b \quad (22)$$

$$SNR_i = \left(\frac{a}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\theta_t - \mu_i)^2}{2\sigma^2}\right\} + b \right) / \left(\frac{a}{\pi} + b \right) \quad (23)$$

$$\text{Let } C_1 = \frac{a+b}{n} \sum_j SNR_j \quad (24)$$

$$g_i = C_1 \left(\frac{a}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\theta_t - \mu_i)^2}{2\sigma^2}\right\} + b \right) \quad (\text{from Equation 9}) \quad (25)$$

where C_1 is a normalization constant. Equation 25 shows that the gain on a neuron depends on the similarity between its preferred feature and the target feature. Thus, the expression for optimal gains reduces the "feature similarity gain model" (Treue and Martinez Trujillo, 1999).

In the opposite case where the distractor feature is known and the target is unknown, we have the following expression for gains:

$$P(\Theta|T) = \frac{1}{\pi} \quad (26)$$

$$P(\Theta|D) = \delta(\theta_d) \quad (27)$$

$$SNR_i = \left(\frac{a}{\pi} + b \right) / \left(\frac{a}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\theta_d - \mu_i)^2}{2\sigma^2}\right\} + b \right) \quad (28)$$

$$\text{Let } C_2 = \frac{1}{\left(\frac{a}{\pi} + b\right)n} \sum_j SNR_j \quad (29)$$

$$g_i = \frac{C_2}{\frac{a}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\theta_d - \mu_i)^2}{2\sigma^2}\right\} + b} \quad (\text{from Equation 9}) \quad (30)$$

where C_2 is a normalization constant. Thus, the gain of a neuron decreases as similarity between its preferred feature and the distractor feature increases.

How do target enhancement and distractor suppression combine when both the target and distractor features are known? Below, we consider the simplest case where both the target and distractor consist of a single feature.

$$P(\Theta|T) = \delta(\theta_t) \quad (\delta() \text{ is the Dirac Delta function}) \quad (31)$$

$$P(\Theta|D) = \delta(\theta_d) \quad (32)$$

$$SNR_i = \left(\frac{a}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\theta_t - \mu_i)^2}{2\sigma^2}\right\} + b \right) / \left(\frac{a}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\theta_d - \mu_i)^2}{2\sigma^2}\right\} + b \right) \quad (33)$$

$$\text{Let } \Delta_i = \frac{\theta_t - \mu_i}{\sigma} \quad (34)$$

$$\text{Let } d' = \frac{\theta_d - \theta_t}{\sigma} \quad (35)$$

$$\text{Let } C_3 = \frac{b\sigma\sqrt{2\pi}}{a} \quad (36)$$

$$\text{Let } C_4 = \frac{1}{n} \sum_j SNR_j \quad (37)$$

$$g_i = C_4 (\exp\{-\Delta_i^2/2\} + C_3) / (\exp\{-(\Delta_i + d')^2/2\} + C_3) \quad (38)$$

Thus, we obtain an expression for optimal gains as a function of d' (discriminability between the target and distractor features) and Δ_i (distance between target feature and neuron's preferred feature in units of tuning width) (Figure 8). For a given neuron, as d' increases, SNR_i increases and g_i increases. When d' is very high, we have:

$$d' \gg \Delta_i \Rightarrow \Delta_i + d' \approx d' \quad (39)$$

$$\Rightarrow g_i \approx C_4 (\exp\{-\Delta_i^2/2\} + C_3) / (\exp\{-d'^2/2\} + C_3) \quad (40)$$

$$\approx \exp\{-\Delta_i^2/2\} + C_3 \quad (41)$$

Thus, when d' is very high, the gain of a neuron decreases as Δ_i (distance between target feature and neuron's preferred feature) increases. In other words, the gains vary according to the feature similarity gain model. The neuron that is best tuned to the target ($\Delta_i = 0$) contributes maximum SNR_i and consequently has maximum gain.

To summarize, when the distractor is unknown or when the distractor is very different from the target (d' is high), then gains follow the feature similarity gain model, which is to our knowledge the situation in which this model has been tested to date. However, when the distractor is similar to the target (d' is low), gains do not follow the feature similarity gain model. Instead, a neuron whose preferred feature is shifted away from the target and distractor feature has higher gain than a neuron that is most similar to the target.

Model Simulations

Additional details of the simulations are given below. We simulate a simple model of early visual cortex as follows: Let f_i represent the bell-shaped tuning curve of the i th neuron (with preferred feature value μ_i) in a population of n neurons with broad, overlapping tuning curves. Let the tuning width σ and amplitude a be the same for all neurons. Let $r_i(\theta)$ be the neural response to stimulus feature θ . $r_i(\theta)$ may be considered a Poisson random variable with mean $f_i(\theta)$ (Softky and Koch, 1993). For simulation purposes, we compute bottom-up salience s_i using the "classic" approach of weighting the local neural response r_i with the square of the difference between the maximum MAX_i and mean responses $MEAN_i$ in that map (for details, see section 2.3 in Itti and Koch, 2001b). Thus, bottom-up salience is low if a feature map has several active locations [i.e., $(MAX_i - MEAN_i)^2 \approx 0$] and is high if a feature map has few active locations [i.e., $(MAX_i - MEAN_i)^2 > 0$]. We chose the following values for our simulation parameters: $n = 100$ (number of neurons in the population), $\sigma = 5$ (width of Gaussian tuning curves), $gap = 0.6\sigma$ (interneural spacing in units of σ), $a = 100$ Hz (amplitude of tuning curve), $\mu_i \in \{0..300\}$ (preferred feature of the i th neuron), $N = 3$ (i.e., 1 target and $N^2 - 1 = 8$ distractors in the display).

Psychophysics Experiments

Additional details of the psychophysics experiments are given below. Subjects were naive to the purpose of the experiment (except one) and were USC students (2 females, 2 males, mixed ethnicities, ages 22–30,

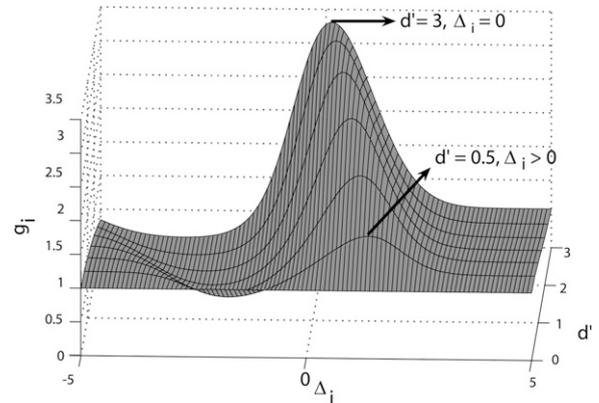


Figure 8. Optimal Gains as a Function of d' and Δ_i , Computed According to Equation 38

When d' is high (e.g., $d' \geq 3$), the maximum gain occurs at $\Delta_i = 0$, i.e., when the target-distractor discriminability is high, a neuron that is tuned to the target feature is promoted maximally. However, when d' is low (e.g., $d' = 0.5$), the maximum gain occurs at $\Delta_i > 0$, i.e., when the target-distractor discriminability is low, a neuron that is tuned to a nontarget feature is promoted more than a neuron tuned to the target feature.

normal corrected or uncorrected vision). Informed written consent was obtained from all the subjects, and they either volunteered or participated for course credit. All experiments received IRB approval. Stimuli were presented on a 22 inch computer monitor (LaCie Corp; 640 × 480, 60.27 Hz double-scan, mean screen luminance 30 cd/m², room 4 cd/m²). Subjects were seated at a viewing distance of 80 cm (52.5° × 40.5° usable field of view) and rested on a chin-rest. Stimuli were presented on a Linux computer under SCHED_FIFO scheduling, which ensured microsecond-accurate timing.

In the experiment shown in Figure 5, the top-down bias is set when subjects perform T_1 trials. After a random number of T_1 trials, the top-down bias is measured in a T_2 trial. A T_1 trial consists of a fixation for 500 ms followed by a search array containing one target (55°) among 25 distractors (50°). Subjects are instructed to find the target as soon as possible and press any key. The time until keypress varied anywhere between 500 and 7000 ms. To verify that subjects indeed find the target on every trial, we introduce a novel *No Cheat* scheme: Following the key press when the subject finds the target, we flash a grid of fingerprint random numbers briefly (120 ms) and ask subjects to report the number at the target's location. The brevity of the display ensures that subjects find the target and fixate it in order to report the number correctly. Online feedback on accuracy of report is provided. Unlike conventional use of target absent trials, which cannot isolate individual trials with invalid responses, our *No Cheat* scheme allows validation of the subject's response on a trial-by-trial basis. Subjects receive training on this experiment until they achieve at least 80% accuracy. During testing, a block is rejected if the accuracy falls below 80%. A T_2 trial consists of a fixation for 500 ms, followed by a brief display of five items representing five features (300 ms), and by five fingerprint random numbers. The task is the same as in the T_1 trials. Subjects are asked to report the number at the target location. Each subject performed ten blocks of 50 trials each, with 160 T_2 trials randomly inserted in between 340 T_1 trials. For each of the four subjects, the reports on the 160 T_2 trials were analyzed using a paired t test ($p < 0.05$) to compare the number of reports on 30°, 50°, 55°, 60°, and 80° features.

Alternative Objective Functions

Here, we explore another objective function, D' , discriminability between the salience of the target and distractor (as defined in

Equation 10). Using the additive hypothesis in Equation 1 (i.e., assuming that saliency adds across the different saliency maps), we get the following:

$$D' = \frac{E[\sum_i g_i s_{IT}(A)] - E[\sum_i g_i s_{ID}(A)]}{\sqrt{0.5(V[\sum_i g_i s_{IT}(A)] + V[\sum_i g_i s_{ID}(A)])}} \quad (42)$$

$$= \frac{\sum_i g_i (E[s_{IT}(A)] - E[s_{ID}(A)])}{\sqrt{0.5(\sum_i g_i^2 (V[s_{IT}(A)] + V[s_{ID}(A)]))}} \quad (43)$$

(assuming s_{IT} , s_{IT} ; s_{ID} , s_{ID} are independent r.v.) (44)

Differentiating D' with regard to g_i yields the following:

$$\left\{ \frac{\partial D'}{\partial g_i} \right\}_{g_i=1} = \frac{t_i - 1}{\alpha_i} \quad (45)$$

where $t_i = \frac{(E[s_{IT}(A)] - E[s_{ID}(A)])}{(V[s_{IT}(A)] + V[s_{ID}(A)])}$ (46)

where $T = \frac{\sum_i E[s_{IT}(A)] - E[s_{ID}(A)]}{\sum_i V[s_{IT}(A)] + V[s_{ID}(A)]}$ (47)

where $\alpha_i = \frac{\sqrt{\frac{1}{2} \sum_i V[s_{IT}(A)] + V[s_{ID}(A)]}}{T \times (V[s_{IT}(A)] + V[s_{ID}(A)])}$ (48)

From Equation 45, it is easy to show that g_i/g_{i0} (where $g_{i0} = 1$ is the default baseline gain) increases as t_i/T increases. Assuming the monotonic relationship to be linear, and with an added constraint that the gains must sum to a constant,

$$\sum_{i=1}^n g_i = n,$$

the simplest solution is:

$$g_i = \frac{t_i}{\frac{1}{n} \sum_{j=1}^n t_j} \quad (49)$$

To compare SNR and D' , we ran simulations and compared the predictions on search performance for different target and distractor feature distributions (see Figure 7). For computing the top-down gains in these simulations, we assumed that saliency s_i could be approximated by the raw neural response r_i . While computing D' , we further assumed that the neural firing rate followed a poisson distribution, hence variance $V[\cdot]$ equals the expectation $E[\cdot]$. The top-down gains were combined with bottom-up saliency (as computed in section 2.8 in Itti and Koch, 2001b) to compute the overall saliency.

Supplemental Data

The Supplemental Data for this article can be found online at <http://www.neuron.org/cgi/content/full/53/4/605/DC1/>.

ACKNOWLEDGMENTS

This work was supported by the Defense Advanced Research Projects Agency (DARPA), Human Frontier Science Program (HFSP), National Geospatial-Intelligence Agency (NGA), National Science Foundation (NSF CRCNS), and office of Naval Research (ONR).

Received: June 28, 2006
 Revised: October 30, 2006
 Accepted: January 8, 2007
 Published: February 14, 2007

REFERENCES

Bauer, B., Jolicoeur, P., and Cowan, W.B. (1996). Visual search for colour targets that are or are not linearly separable from distractors. *Vision Res.* 36, 1439–1465.

Bichot, N.P., and Schall, J.D. (2002). Priming in macaque frontal cortex during popout visual search: feature-based facilitation and location-based inhibition of return. *J. Neurosci.* 22, 4675–4685.

Bichot, N.P., Rossi, A.F., and Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science* 308, 529–534.

Boynton, G.M. (2005). Attention and visual perception. *Curr. Opin. Neurobiol.* 15, 465–469.

Braithwaite, J.J., and Humphreys, G.W. (2003). Inhibition and anticipation in visual search: evidence from effects of color foreknowledge on preview search. *Percept. Psychophys.* 65, 213–237.

Bundesen, C. (1990). A theory of visual attention. *Psychol. Rev.* 97, 523–547.

Cave, K.R. (1999). The FeatureGate model of visual selection. *Psychol. Res.* 62, 182–194.

Chelazzi, L., Miller, E.K., Duncan, J., and Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature* 363, 345–347.

Deco, G., and Rolls, E.T. (2002). *Computational Neuroscience of Vision* (New York: Oxford University Press).

Deneve, S., Latham, P.E., and Pouget, A. (1999). Reading population codes: a neural implementation of ideal observers. *Nat. Neurosci.* 2, 740–745.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.

Duncan, J. (1989). Boundary conditions on parallel processing in human vision. *Perception* 18, 457–469.

Duncan, J., and Humphreys, G.W. (1989). Visual search and stimulus similarity. *Psychol. Rev.* 96, 433–458.

D’Zmura, M. (1991). Color in visual search. *Vision Res.* 31, 951–966.

Gandhi, S.P., Heeger, D.J., and Boynton, G.M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proc. Natl. Acad. Sci. USA* 96, 3314–3319.

Hamker, F.H. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Res.* 44, 501–521.

Hillyard, S.A., Vogel, E.K., and Luck, S.J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1257–1270.

Hodsoll, J., and Humphreys, G.W. (2001). Driving attention with the top down: the relative contribution of target templates to the linear separability effect in the size dimension. *Percept. Psychophys.* 63, 918–926.

Hopfinger, J.B., Buonocore, M.H., and Mangun, G.R. (2000). The neural mechanisms of top-down attentional control. *Nat. Neurosci.* 3, 284–291.

Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40, 1489–1506.

Itti, L., and Koch, C. (2001a). Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203.

Itti, L., and Koch, C. (2001b). Feature combination strategies for saliency-based visual attention systems. *J. Electron. Imaging* 10, 161–169.

Kastner, S., Pinsk, M.A., De Weerd, P., Desimone, R., and Ungerleider, L.G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* 22, 751–761.

Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4, 219–227.

- Lee, D.K., Itti, L., Koch, C., and Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nat. Neurosci.* *2*, 375–381.
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends Cogn. Sci.* *6*, 9–16.
- Luck, S.J., Chelazzi, L., Hillyard, S.A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* *77*, 24–42.
- Maljkovic, V., and Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Mem. Cognit.* *22*, 657–672.
- Martinez, A., Anllo-Vento, L., Sereno, M.I., Frank, L.R., Buxton, R.B., Dubowitz, D.J., Wong, E.C., Hinrichs, H., Heinze, H.J., and Hillyard, S.A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat. Neurosci.* *2*, 364–369.
- Martinez-Trujillo, J.C., and Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.* *14*, 744–751.
- McAdams, C.J., and Maunsell, J.H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* *19*, 431–441.
- Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science* *229*, 782–784.
- Motter, B.C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* *70*, 909–919.
- Motter, B.C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. Neurosci.* *14*, 2178–2189.
- Nagy, A.L., and Sanchez, R.R. (1990). Critical color differences determined with a visual search task. *J. Opt. Soc. Am. A* *7*, 1209–1217.
- Najemnik, J., and Geisler, W.S. (2005). Optimal eye movement strategies in visual search. *Nature* *434*, 387–391.
- Navalpakkam, V., and Itti, L. (2006). An integrated model of top-down and bottom-up attention for optimal object detection. *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–7.
- Pashler, H. (1987). Target-distractor discriminability in visual search. *Percept. Psychophys.* *41*, 385–392.
- Rao, R.P., Zelinsky, G., Hayhoe, M., and Ballard, D.H. (2002). Eye movements in iconic visual search. *Vision Res.* *42*, 1447–1463.
- Reynolds, J.H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* *19*, 1736–1753.
- Roelfsema, P.R., Lamme, V.A., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature* *395*, 376–381.
- Saenz, M., Buracas, G.T., and Boynton, G.M. (2002). Global effects of feature-based attention in human visual cortex. *Nat. Neurosci.* *5*, 631–632.
- Shiffrin, R.M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychol. Rev.* *84*, 127–190.
- Softky, W.R., and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.* *13*, 334–350.
- Somers, D.C., Dale, A.M., Seiffert, A.E., and Tootell, R.B. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proc. Natl. Acad. Sci. USA* *96*, 1663–1668.
- Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. *J. Exp. Psychol. Hum. Percept. Perform.* *17*, 652–676.
- Treisman, A., and Gelade, G. (1980). A feature integration theory of attention. *Cognit. Psychol.* *12*, 97–136.
- Treue, S., and Martinez Trujillo, J.C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* *399*, 575–579.
- Tsotsos, J.K., Culhane, S.M., Wai, W.Y.K., Lai, Y.H., Davis, N., and Nuflo, F. (1995). Modeling visual-attention via selective tuning. *Artif. Intell.* *78*, 507–545.
- Vickery, T.J., King, L.-W., and Jiang, Y. (2005). Setting up the target template in visual search. *J. Vis.* *5*, 81–92.
- Watanabe, T., Sasaki, Y., Miyauchi, S., Putz, B., Fujimaki, N., Nielsen, M., Takino, R., and Miyakawa, S. (1998). Attention-regulated activity in human primary visual cortex. *J. Neurophysiol.* *79*, 2218–2221.
- Wolfe, J.M. (1994). Guided search 2.0: a revised model of visual search. *Psychonomic Bulletin and Review* *1*, 202–238.
- Wolfe, J.M., Butcher, S.J., and Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *J. Exp. Psychol. Hum. Percept. Perform.* *29*, 483–502.
- Wolfe, J.M., Horowitz, T.S., Kenner, N., Hyle, M., and Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Res.* *44*, 1411–1426.