Top–down attention selection is fine grained

Vidhya Navalpakkam
Department of Computer Science, University of Southern California, Los Angeles, CA, USA

Laurent Itti
Departments of Computer Science, Psychology, and Neuroscience Graduate Program, University of Southern California, Los Angeles, CA, USA

Although much is known about the sources and modulatory effects of top–down attentional signals, the information capacity of these signals is less known. Here, we investigate the granularity of top–down attentional signals. Previous theories in psychophysics have provided conflicting evidence on whether top–down guidance is coarse grained (i.e., one gain control term per feature dimension) or fine grained (i.e., multiple gain control terms per dimension). We resolve the conflict by designing new experiments that disentangle top–down from bottom–up contributions, thereby avoiding confounds existing in previous studies. The results of our eye-tracking experiments show that subjects can selectively saccade to items belonging to the relevant feature interval compared with irrelevant intervals within a dimension. This suggests that top–down signals can specify not only the relevant feature dimension but also the relevant feature interval within a dimension. We conclude that top–down signals are fine grained and can specify multiple gain control terms per dimension.

Keywords: top–down, visual attention, feature, gain, granularity, eye movements

Introduction

Attention is guided as a combination of top–down and bottom–up factors

Visual attention is guided as a combination of at least two factors: bottom–up factors based on spatiotemporal differences in visual input (Itti & Koch, 2001) and top–down factors based on prior knowledge of the stimuli (Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004). For instance, a ripe red fruit among green leaves is bottom–up salient and attracts attention due to the difference in color (Itti & Koch, 2000; Li, 1999; Tsotsos et al., 1995). Top–down factors such as prior knowledge that the fruit is red can further accelerate search speed by increasing the activity of neurons tuned to the red feature (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990; Motter, 1994; Saenz, Buracas, & Boynton, 2002; Treue & Martinez Trujillo, 1999). Thus, both bottom–up and top–down factors together guide attention to visually salient and relevant scene locations (Navalpakkam & Itti, 2005b).

Importance of studying granularity of top–down signals

The natural world contains prey and predators that are camouflaged and, hence, visually nonsalient. For instance, a lion camouflaged in the dry savannah is hard to detect because its golden fur has similar tint as the yellowish grasslands. In such situations where bottom–up guidance is minimal, the prey's survival depends on whether top–down can guide attention by selecting the fine-grained target feature (in this case, selecting the relevant shade of yellow among different shades). Hence, the granularity of top–down signals plays a critical role in determining visual search performance. Despite its importance, the granularity or information capacity of top–down signals has been less studied than their sources or modulatory effects on early sensory areas (Chawla, Rees, & Friston, 1999; Chelazzi, Miller, Duncan, & Desimone, 1993; Lee, Itti, Koch, & Braun, 1999; Moran & Desimone, 1985; Motter, 1994; Saenz et al., 2002; Treue & Maunsell, 1996). In an elegant electrophysiological study, Treue and Martinez Trujillo (1999) showed evidence for differential gains on neurons tuned to different directions of motion, thereby demonstrating high top–down granularity. Here, we investigate whether the same is true in other dimensions like intensity, color saturation, and size. We perform an additional test of granularity that was previously ignored in the electrophysiological studies, namely, whether attention can selectively enhance an intermediate feature in a dimension while suppressing flanking distractor features. Few psychophysics studies have tried to address these issues of granularity, and their results provide conflicting evidence. Some studies suggest that top–down signals are coarse grained (Figure 1a, e.g., one gain control term for the intensity dimension, thereby selecting all values or intervals of intensity; Found & Muller, 1996; Muller, Heller, & Ziegler, 1995), whereas others suggest that top–down signals are fine grained (Figure 1b, e.g., multiple gain control terms within the intensity dimension, allowing selection of a particular interval of intensity; Pomplun, 2006; see below for a detailed literature review). Investigating the

doi: 10.1167/6.11.4 Received February 24, 2006; published October 17, 2006 ISSN 1534-7362 © ARVO

granularity of top–down signals is therefore crucial for further progress in understanding top–down attention modulation. In the rest of this section, we present an overview of relevant literature.

Guided search theory

One of the most influential theories of visual search is the guided search theory (Wolfe, 1994). It successfully accounts for several observed phenomena in human visual search behavior, such as pop-out versus conjunction (Treisman & Gelade, 1980), target–distractor discriminability (Duncan & Humphreys, 1989; Nagy & Sanchez, 1990; Pashler, 1987; Treisman, 1991), distractor heterogeneity (Duncan & Humphreys, 1989), and feature priming (Maljkovic & Nakayama, 1994; Wolfe, Butcher, & Hyle, 2003). It suggests a two-stage model of visual processing. In the preattentive stage, feature maps are computed in parallel in several feature dimensions (e.g., red, blue, green, and yellow feature maps in color hue dimension; steep, shallow, left, and right maps in the orientation dimension). In the second stage, top–down multiplicative gains are applied on these bottom–up maps, and the weighted feature maps are combined additively to form an activation map that eventually guides visual attention in a sequential manner. Thus, during search for a red item, the theory suggests that the weight on the red feature may be increased, resulting in increased activity of all red items in the scene. Although the theory includes top–down guidance through a multiplicative gain control mechanism, it does not directly address the issue of granularity of top–down guidance. For some dimensions like orientation, it explicitly states that there may be multiple gain control terms for steep, shallow, left, and right features. However, for other dimensions like intensity, size, and color saturation, it does not comment on the granularity.

Linear separability effect

A popular effect observed in visual search behavior suggests that search is easier when the target can be separated from the distractors by a line in feature space. For instance, in the intensity dimension, search for the brightest item is easier than search for a medium-bright item among brighter and darker items. This effect has been reported in several dimensions such as color (D'Zmura, 1991), chromaticity (Bauer, Jolicoeur, & Cowan, 1996), luminance (Bauer et al., 1996; Hodsoll & Humphreys, 2001), and size (Hodsoll & Humphreys, 2001; Treisman, 1988; Wolfe & Bose, 1991, unpublished data). Inefficient search for a MID interval target seems to suggest that top–down cannot select the MID interval within a feature dimension. Hence, these results seem to support the hypothesis that top–down guidance is coarse grained. However, the above experiments varied both the target and the distractor stimuli across the search conditions, thereby varying both bottom–up and top–down guidance and making it difficult to tease apart the top–down contribution. For instance, in the HIGH search condition, their subjects searched for a single HIGH intensity target among many MID and LOW intensity distractors, whereas in the MID condition, subjects searched for a single MID intensity target among many LOW and HIGH intensity distractors. Thus, both the target and distractor stimuli varied across search conditions, leading to changes in both bottom–up and top–down effects. Indeed, bottom–up guidance alone suffices to account for the previous results—search for the MID intensity target is slower as the target is not bottom–up salient (due to its similarity to both LOW and HIGH intensity distractors), whereas search for the HIGH intensity target is faster as it is more bottom–up salient (due to the large difference from the LOW intensity distractors). Due to the entangling of top–down and bottom–up effects, the above experiments cannot reveal the role of top–down guidance. We overcome this confound by maintaining the background stimulus constant while varying only the target across the search conditions. Thus, bottom–up factors remain nearly constant, whereas the top–down factor varies, allowing us to infer its role unambiguously.

Subset search

Classical conjunction searches (e.g., search for a red vertical bar among green vertical and red horizontal bars) were known to be hard (Treisman & Gelade, 1980). However, later experiments revealed a distractor-ratio effect (Bacon & Egeth, 1997; Egeth, Virzi, & Garbart, 1984; Kaptein, Theeuwes, & van der Heijden, 1995; Shen, Reingold, & Pomplun, 2000; Zohary & Hochstein, 1989); that is, subjects tend to search in a dimension defined by the smaller subset of distractors (i.e., if there are fewer red horizontal bars than green vertical bars, then subjects focus on the color dimension and search through the red items). Thus, conjunction search can become efficient if subjects search through the smaller subset. Although these experiments demonstrate that subjects can selectively attend to feature within a dimension (e.g., red within color dimension), they do not indicate whether subjects can select an intermediate feature in a dimension.

Dimension weighting

Several studies in the past investigated top–down guidance to feature dimensions. Their results show that prior knowledge of the target dimension can facilitate search (Found & Muller, 1996; Kumada, 2001; Muller et al., 1995; Treisman, 1988). A prominent theory is the dimension-weighting account (Found & Muller, 1996;

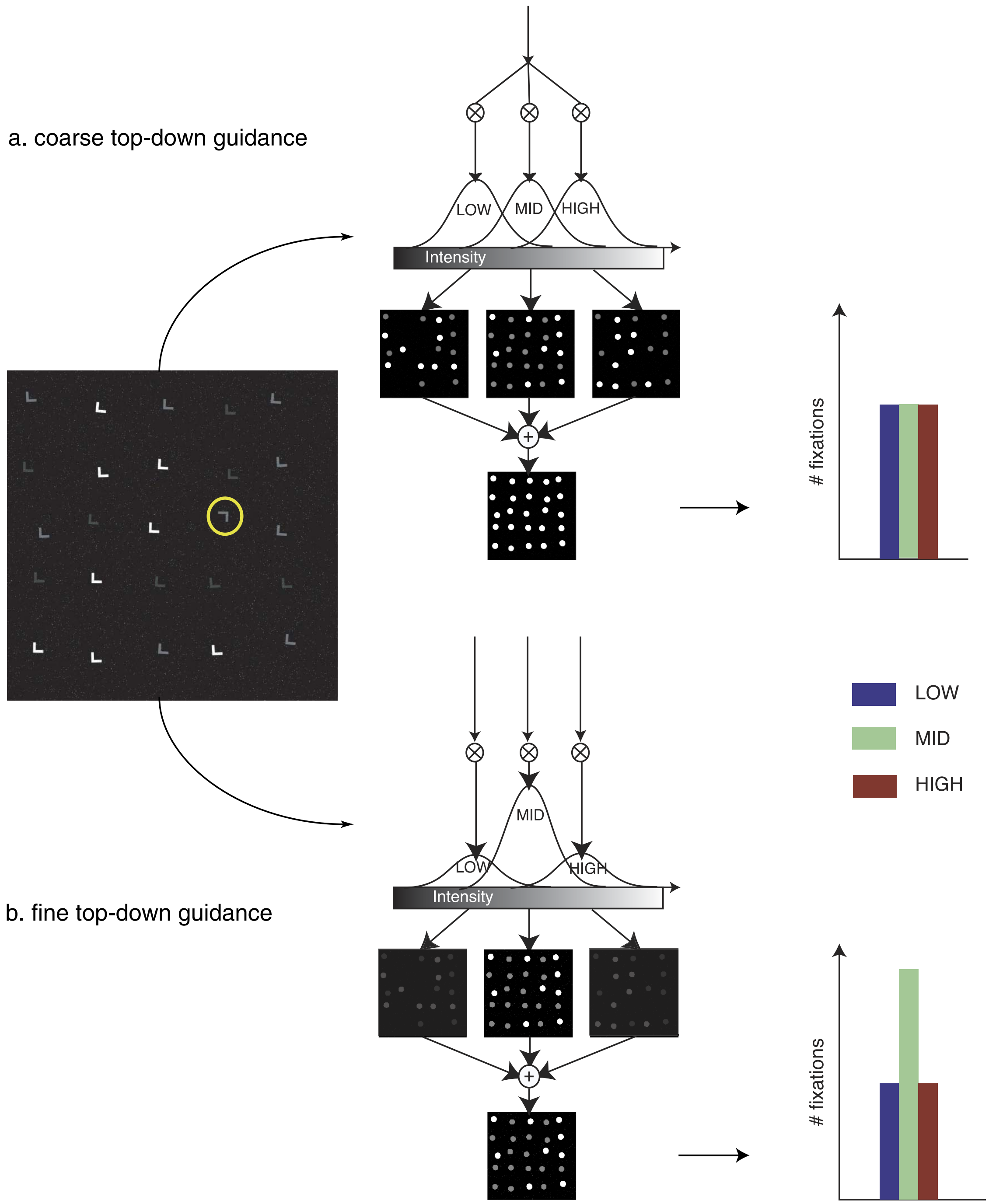

Figure 1. Testing the hypotheses: Consider searching for a MID intensity target (marked by a yellow circle for illustration purposes) among LOW, MID, and HIGH intensity distractors. Let the display be processed by neurons that are tuned to LOW, MID, and HIGH intensity intervals. The feature maps corresponding to the LOW, MID, and HIGH intensity intervals are added to form a saliency map that subsequently guides attention. (a) If top–down guidance were coarse, the gains on LOW, MID, and HIGH intensity intervals would be equal, resulting in equal salience of all items, thereby yielding equal number of fixations on all intervals. (b) In contrast, if top–down guidance were fine grained, the gain on the relevant MID intensity interval would be higher than that on LOW and HIGH intensity intervals, resulting in higher salience of items in the MID interval, thereby yielding higher number of fixations on the MID interval than on the LOW or HIGH interval.

Muller et al., 1995). It suggests that during search for a target, feature dimensions are weighted so that the known target dimension is promoted. The experimental paradigm was as follows: In a within-dimension condition, the target dimension was known and remained constant across trials, but its value varied within that dimension, whereas in the cross-dimension condition, the target dimension varied across trials. Treisman observed shorter reaction times (RTs) in the within-dimension condition than the cross-dimension condition. Muller et al. observed such within-dimension facilitation even between successive trials. This led to a dimension-weighting account, suggesting that the known target dimension receives a higher weight compared with other unknown dimensions, thereby increasing the target's activity in the master map, resulting in faster search for the target. Although these studies suggest weights on feature dimensions (e.g., intensity), they do not suggest weights within a dimension. Here, we wish to investigate the granularity of weights. For instance, is there one weight per dimension or one weight per feature interval within a dimension? In other words, can the coarse feature dimension-weighting account be extended to a finer feature interval-weighting account?

Visual guidance in complex scenes

A recent study (Pomplun, 2006) investigates visual guidance to low-level features in complex natural scenes. The experiment consists of the following paradigm—subjects first preview a target patch (74 × 74 pixels) extracted from the image and subsequently search for the target in the image. Analysis of eye-movement data reveals that subjects saccade to image regions that have similar intensity, contrast, spatial frequency, and orientation as the target. For instance, if the target has MID intensity, there are more saccades to MID intensity regions of the image than to LOW or HIGH intensity regions. This difference in saccadic selectivity was assumed to reflect top–down guidance. The author proceeded to compare the strength of guidance to different feature dimensions, showing decreasing order of guidance for intensity, contrast, spatial frequency, and orientation. The clever experiment design allowed the author to break up each dimension into smaller intervals and infer the spread of guidance through the distribution of saccades as a function of distance from target interval. However, the above experiment suffers from the same confound as experiments on linear separability. Across the LOW, MID, and HIGH search conditions, the author varied not only the target but also the background image. Hence, the measured guidance reflects a combination of both top–down and bottom–up effects, making it difficult to tease apart the contribution of top–down guidance. Indeed, bottom–up effects were not controlled in that experiment. It was not verified whether the regions similar to the target were bottom–up salient or not. Although the proportion of LOW, MID, and HIGH intensity regions was equal when pooled over all images, the proportion was not controlled within a given search condition. It may have been possible that during search for the MID interval target, there were fewer MID interval regions, thereby increasing their bottom–up salience and yielding higher saccadic selectivity. Indeed, the author confirms this by reporting feature-ratio effects; that is, a feature that is present in smaller proportion or ratio in the image attracts higher number of saccades. Such bottom–up effects need to be controlled to allow unambiguous inference on the role of top–down guidance. We achieve this by varying the target stimulus while keeping the background constant. This alteration in the experimental design allows us to investigate top–down guidance without any bottom–up confounds. More details are given in the next section.

Design and analysis of experiments

Contending hypothesis

In this article, we investigate the granularity of top–down signals by comparing two competing hypotheses: (a) top–down guidance is coarse grained versus (b) it is fine grained. As mentioned earlier, the coarse-grained hypothesis is supported by several existing visual search theories. For instance, the dimension-weighting account (Found & Muller, 1996; Muller et al., 1995) of visual search behavior suggests coarse-grained top–down guidance of one gain control term per feature dimension (i.e., the gains on all intervals within that dimension are equal, see Figure 1a). The competing hypothesis is that top–down guidance is fine grained and contains several gain control terms per feature dimension (see Figure 1b; Pomplun, 2006).

Testing the hypotheses

To test these hypotheses, we designed visual search experiments (see Figure 1) where subjects searched for a target belonging to a fine-grained feature interval among distractors belonging to many intervals within a feature dimension (e.g., within the intensity dimension, search for a medium intensity target among distractors of low, medium, and high intensities). Assuming that attention is guided by a saliency map formed by summing feature maps, the coarse- and fine-grained hypotheses generate contradictory predictions on search behavior: According to the coarse-grained hypothesis, the gains on all intervals are equal. Hence, all feature maps contribute equally to the saliency map, resulting in equal salience of items of all intervals, yielding equal number of fixations on each interval. In contrast, the fine-grained hypothesis predicts higher gain on the relevant interval, leading to an increased contribution of the relevant feature map, resulting in higher salience of items of the relevant interval, thereby yielding higher number of fixations. Thus, the fine-grained hypothesis predicts that items of the relevant fine-grained interval should be preferentially fixated. To test these hypotheses, we design the following experiments.

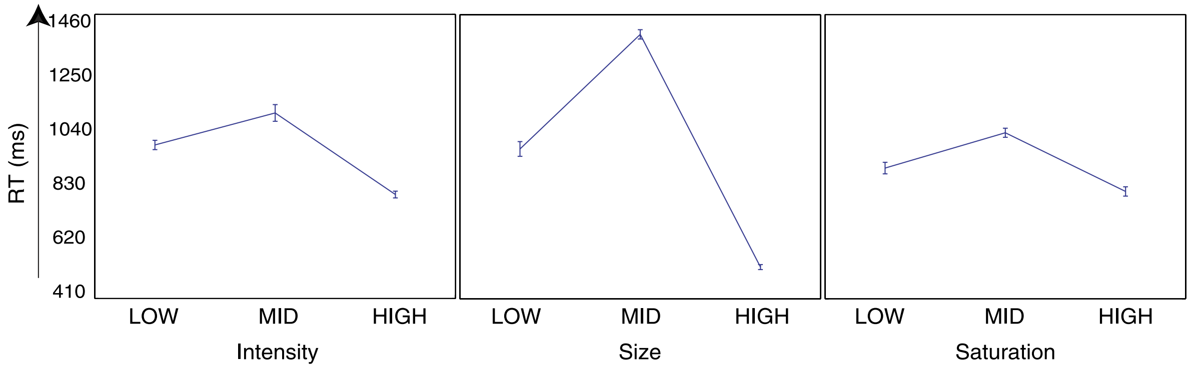

Figure 2. This figure shows the RT for all valid trials in the LOW, MID, and HIGH search conditions within intensity, size, and color saturation dimensions. In all feature dimensions, search was slower in the MID condition than in the LOW or HIGH condition, as demonstrated by the linear separability theory. Note that there is no speed–accuracy tradeoff here as the RT was only computed over valid (correct) trials.

Experiment 1: Intensity

This section describes the design and analysis of eye movements to determine whether top–down guidance can selectively enhance the relevant interval within the intensity dimension.

Design of the stimuli

The details of our experiments are as follows: The intensity dimension is divided into fine-grained feature intervals: LOW, MID, and HIGH. The target and distractor stimuli belong to one of LOW, MID, or HIGH intervals, and the distractors are L shaped, whereas the target is rotated by 180°. This rotation enables recognition of the target but disables preattentive guidance (control experiments reveal a search efficiency of 23 ms/item, indicating that the rotated L target is preattentively indistinguishable from the upright L distractors).

Search conditions

There are three search conditions: LOW, MID, and HIGH, based on whether the target interval belongs to LOW, MID, or HIGH intervals within that dimension, respectively. To avoid confounds due to stimulus-driven bottom–up factors, we maintain the same background stimulus (equal numbers of LOW, MID, and HIGH intensity distractors) across all three search conditions and vary only the target. For example, in the MID condition, subjects search for a MID intensity target among equal numbers of LOW, MID, and HIGH intensity distractors, whereas in the HIGH condition, they search for a HIGH intensity target among equal numbers of LOW, MID, and HIGH intensity distractors. Examples of our displays are shown in Figures 3a, 3b, and 3c.

Additional details of stimuli

Each item is 64 × 64 pixels in size (1.2°). To avoid spatial biasing, the target and distractors can randomly appear anywhere in the invisible 5 × 5 grid that filled the search array. Further, jitter is introduced by rotating each item randomly up to 5°, and random colored noise is added to the display. Stimuli are presented on a 22-in. computer monitor (LaCie Corp.; 1,280 × 1,024; 60.27 Hz double-scan; mean screen luminance, 30 cd/m^2; room, 4 cd/m^2). The search array (1,024 × 1,024 pixels) is embedded on a black background and displayed at the center of the monitor screen (1,280 × 1,024). The display is viewed at a distance of 80 cm, and the viewing angle is 28° × 21°. The stimuli parameters are as follows: In the intensity dimension, the luminance values are as follows: LOW, 4.1 cd/m^2; MID, 21 cd/m^2; HIGH, 112 cd/m^2. These values of LOW, MID, and HIGH intervals are chosen according to the Weber's law. Examples of our displays are shown in Figure 2. To avoid any confounds in inference due to differences in other features, our stimuli are always designed to be identical in all irrelevant feature dimensions and differ only in the intervals within the relevant feature dimension. Thus, in the intensity experiments, all stimuli have the same size, color, and orientation and differ only in the luminance values.

Experimental organization

Subjects perform one search condition a day, for three consecutive days. Each search condition lasts up to an hour and is composed of a maximum of 10 blocks, containing 20 trials each. Subjects run as many blocks as they can (in the range of 8–10) within an hour. Subjects are allowed to take a break in between blocks.

"No Cheat" scheme for response validation

Each trial begins with a central fixation for 250 ms followed by stimulus onset. Subjects search for the target as fast as possible and hit a key upon finding it. Due to boredom or weariness or other factors, subjects may falsely report that they found the target. To avoid such false positives, we introduce a novel "No Cheat" scheme: Upon the key press indicating that the target was found in the display, we flash a grid of two-digit random numbers (of size 0.6° each) for 120 ms and ask the subject to report the random number that flashed at the target's location. Subjects could correctly report the number only if they were fixating at the target location. Online feedback ("Correct" or "Wrong") is provided to the subject based on whether the reported number matches with the flashed number. Only Correct trials (i.e., where the subject correctly reported the number at the location of the target) are considered for analysis of eye-movement patterns. Our choice of the ''No Cheat'' paradigm instead of traditional target absent trials was motivated by the following reasons: Although target absent trials yield more fixations per trial, they are more time consuming. Besides, by validating the subject's response on a per-trial basis, the "No Cheat" paradigm provides a better guarantee that subjects are actively biasing for the target on each and every trial. This also minimizes data wastage by rejecting only the Wrong trials (instead of rejecting the entire block in which it occurs).

Details of eye tracking

A nine-point eye-tracker calibration is performed at the beginning of each block. Each calibration point consists of fixating a central cross, then a blinking dot at a random point on a 3×3 matrix. The experiment is self-paced, and the subjects can stretch before any nine-point calibration. Subjects fixate on a central cross and press a key to start, at which point the trial beings. The eye tracker records from the beginning of the display of the search array to the point when the key is pressed. Each search array image is entirely preloaded into memory. Eye position is tracked using a 240-Hz infrared-video-based eye tracker (ISCAN, Inc.). All analyses are performed offline.

Data cleaning

To verify whether top–down guidance can select the relevant fine-grained interval, we analyzed the eye-movement data of three subjects with normal or corrected vision, who participated for course credit or volunteered. Blocks with bad eye-tracker calibration were not considered for subsequent analysis (0–4% data). Similarly, trials with too many blinks were discarded (0–7% data). As mentioned previously, Wrong trials (incorrect report of the random number flashed at the target's location) were also discarded (0–3% data). It was very rare that subjects indeed found the target but did not report the number correctly (from personal communication with subjects, this error varied between 0% and 2% for different subjects). Subjects had to fixate on a central fixation at the beginning of each trial (to avoid any subject biases toward specific spatial regions). Those trials in which subjects began by fixating more than 3° away from the center were also discarded (0–8% data). All subsequent analyses were performed only on the remaining valid trials.

Reaction time

We computed RT as the time taken to find the target (time from stimulus onset until key press). We compared RTs across the LOW, MID, and HIGH search conditions. As reported in earlier studies (Bauer et al., 1996; D'Zmura, 1991), the RT was significantly higher ($p < .05$) in the MID condition than in the LOW or HIGH condition (see Figure 3a). This replicates the results of previous studies.

Saccadic selectivity

We measured the saccadic selectivity toward LOW, MID, and HIGH intensity intervals in the following manner: We parsed the eye-movement patterns in the valid trials into fixations and saccades and assigned each fixation to the nearest item in the search array (sample eye traces are shown in the first row of Figure 3). Saccadic selectivity for an interval was computed as the total number of fixations that were assigned to items belonging to that interval. For a given search condition (e.g., search for LOW intensity target), we compared saccadic selectivity across different intervals by pooling the trials across all blocks and subjects and performing a paired t test. Statistical analysis revealed a significantly higher saccadic selectivity ($p < .05$) toward the relevant interval than irrelevant intervals. For instance, in the MID condition, search for a MID intensity target leads to more fixations on the MID intensity items than on the LOW or HIGH intensity items. This was consistent for all search conditions (see second row of Figure 3).

Strength of biasing

For a given search condition, we determined the strength of biasing as a function of time by computing the percentage of fixations on the relevant interval for each block. The third row in Figure 4 shows a plot of the strength of biasing as a function of time. Given that there are equal number of items in all three intervals of intensity, there should be a 33.3% chance of fixating each interval. Yet, a t test reveals that the percentage of fixations on the relevant interval is significantly higher ($p \ll .01$) than that predicted by chance (see Table 1). This reveals a clear effect of top–down guidance through selective enhancement of the relevant interval. Also, the strength of biasing seems higher in the LOW and HIGH search conditions (95% confidence interval of [50.58, 57.9] and [53.30, 61.79], respectively) than in the MID condition (95% confidence interval of [37.26, 44.78]). This suggests why the RT is lower in the LOW and HIGH conditions than in the MID condition (Figure 3). Does the strength of biasing vary with time? To answer this

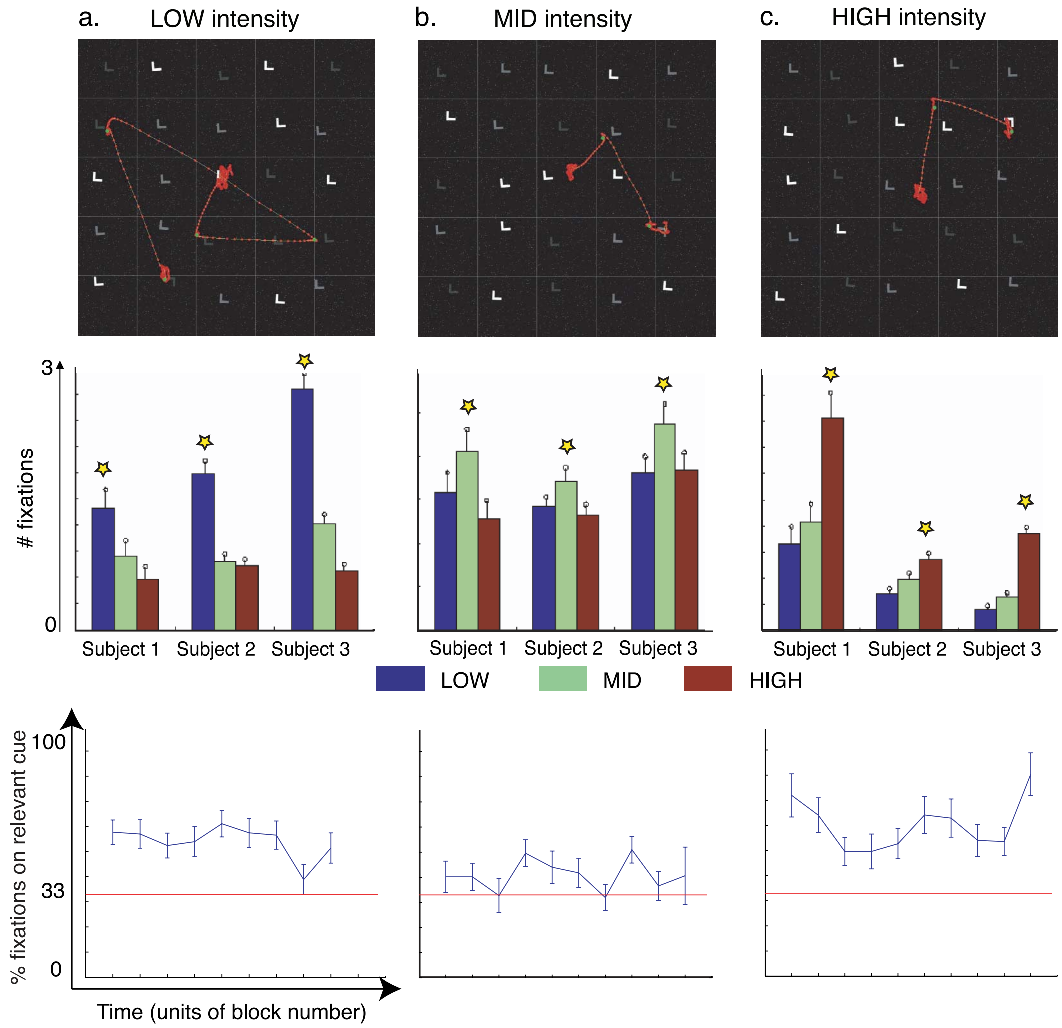

Figure 3. Results in the intensity dimension: (a) The first column shows results during search for a LOW intensity target. The sample eye trace illustrates that subjects tend to fixate on the relevant LOW intensity distractors. Statistical analysis of all trials reveals a significantly higher number of fixations on the relevant LOW intensity items (indicated by a yellow star) than on the MID or HIGH intensity items (paired t test, $p < .05$). Statistical analysis of fixations as a function of time (in units of block number) reveals that the strength of biasing does not change with time (see Table 2). Similar results are observed for the (b) MID and (c) HIGH conditions. As shown in the second column, when subjects search for a MID intensity target, they selectively fixate on the MID intensity distractors, as compared with LOW or HIGH intensity distractors. These results demonstrate that top–down signals can guide attention to the relevant interval within the intensity dimension.

question, we performed a one-way ANOVA (Table 2). Results show that there is no main effect of time (in units of block number) on the strength of biasing.

Experiment 2: Size

To verify the generality of the top–down biasing effect observed in the intensity dimension, we repeated similar experiments and analysis on the size dimension.

Experimental design

In the size experiments, all stimuli have the same luminance, color saturation, and orientation and differ only in the size values. The values for the three conditions (LOW, 0.6°; MID, 1.2°; HIGH, 2.4°) are chosen according to Weber's law. Other experimental details are similar to those in intensity.

Saccadic selectivity

As seen in Figure 4, in all search conditions, there was significantly higher saccadic selectivity (paired t test, $p < .05$) toward the relevant size interval. For instance, in the MID condition, during search for a MID sized target, there were more fixations on the MID sized items compared with LOW or HIGH sized items.

Strength of biasing

The percentage of fixations on the relevant size interval was significantly higher (t test, $p \ll .01$) than the baseline 33.3%

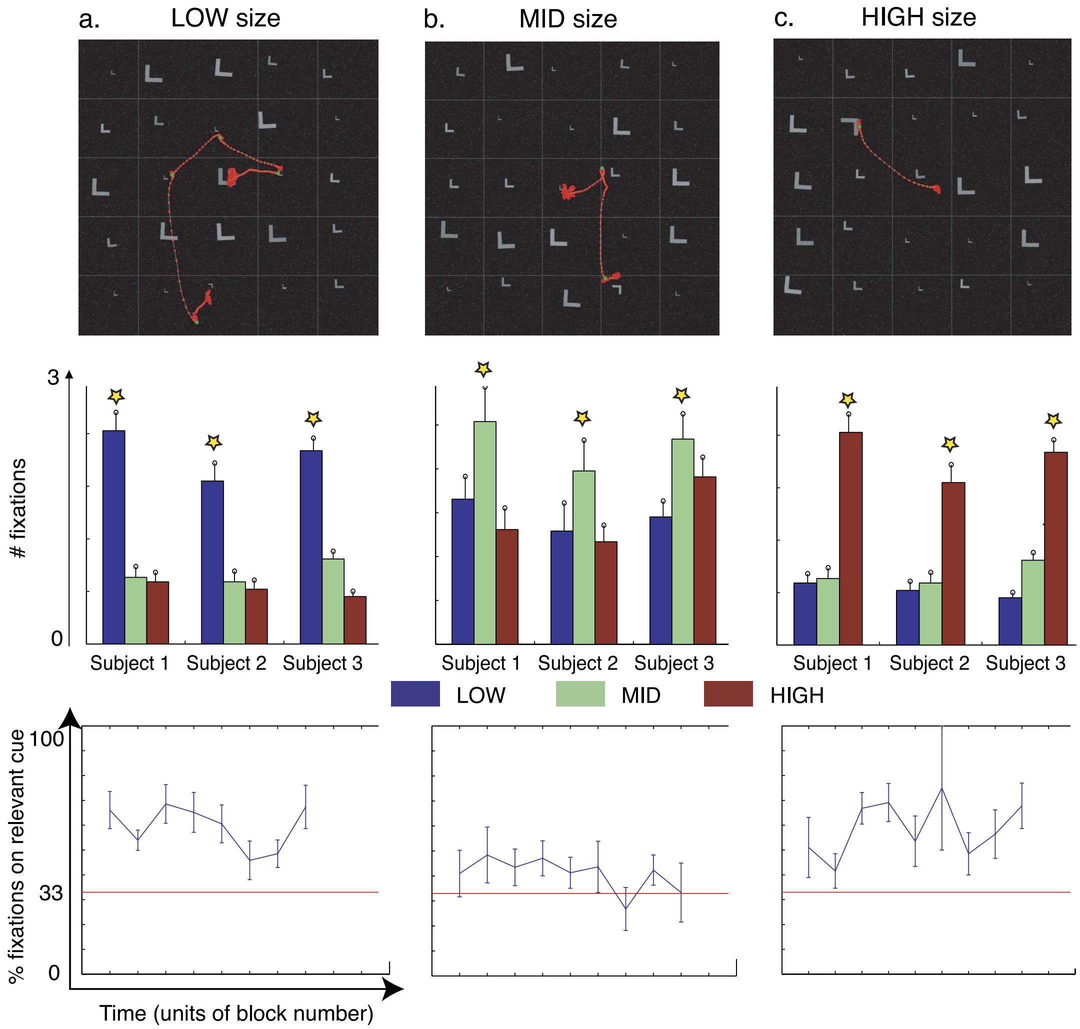

Figure 4. Results in the size dimension: (a) The first column shows results during search for a LOW sized target. The sample eye trace illustrates that subjects tend to fixate on the relevant LOW sized distractors. Statistical analysis of all trials reveals a significantly higher number of fixations on the relevant LOW sized items (indicated by a yellow star) than on the MID or HIGH sized items (paired *t* test, $p < .05$). Analysis of fixations as a function of time (measured in units of blocks from 1 to a maximum of 10) reveals that the strength of biasing does not change with time (see Table 2). Similar results are observed for the (b) MID and (c) HIGH conditions. As shown in the second column, when subjects search for a MID sized target, they selectively fixate on the MID sized distractors, as compared with LOW or HIGH sized distractors. These results demonstrate that top–down signals can guide attention to the relevant interval within the size dimension.

predicted by chance. In the LOW condition, the 95% confidence interval was as high as [54.32, 63.84], whereas in the HIGH and MID conditions, it was [49.94, 62.23] and [36.22, 46.78], respectively. Thus, in all conditions, the strength of biasing was significantly higher than by chance, thereby indicating strong effects of top–down guidance. As with intensity, the results of a one-way ANOVA show that the strength of biasing did not change with time (see Table 2).

	LOW	MID	HIGH
Intensity	$p < 10^{-26}$ [50.58, 57.9]	$p < 10^{-5}$ [37.26, 44.78]	$p < 10^{-25}$ [53.30, 61.79]
Size	$p < 10^{-20}$ [54.32, 63.84]	$p < 10^{-6}$ [36.22, 46.78]	$p < 10^{-12}$ [49.94, 62.23]
Saturation	$p < 10^{-15}$ [44.18, 51.36]	$p < 10^{-36}$ [51.96, 58.39]	$p < 10^{-17}$ [50.10, 59.34]

Table 1. For each dimension tested (intensity, size, color saturation), we find the strength of biasing (computed as percentage of fixations on the relevant feature interval) in the LOW, MID, and HIGH search conditions. A *t* test reveals that in each search condition, the strength of biasing is significantly higher ($p \ll .01$) than the baseline 33.33% predicted by chance. The *p* values and 95% confidence interval (enclosed in brackets) in strength of biasing are reported.

	LOW	MID	HIGH
Intensity	$F(8, 485) = 1.29$, $p = .2448$	$F(9, 491) = 1.16$, $p = .322$	$F(9, 442) = 1.69$, $p = .0912$
Size	$F(7, 442) = 1.45$, $p = .191$	$F(8, 477) = 0.50$, $p = .8521$	$F(8, 447) = 1.35$, $p = .2249$
Saturation	$F(9, 422) = 1.25$, $p = .2651$	$F(9, 461) = 0.60$, $p = .8004$	$F(8, 438) = 1.56$, $p = .1374$

Table 2. For a given dimension (e.g., intensity, size, or color saturation) and a given search condition (e.g., search for a target of LOW, MID, or HIGH feature interval), we determine whether the strength of biasing changes with time by performing a one-way ANOVA test. Results across all conditions show that the strength of biasing does not change significantly ($p \geq .05$) with time (measured in units of block number ranging from 1 up to a maximum of 10).

Experiment 3: Color saturation

Next, we further verify the generality of the top–down biasing effect observed in intensity and size dimensions by repeating similar experiments and analyses on the color saturation dimension.

Experimental design

In these experiments, we desire all stimuli to have the same luminance, orientation, and size and to differ only in the color saturation values. This is trickier because the perceived luminance value of different color saturations is observer dependent. Hence, we run heterochromatic photometry experiments (Pokorny, Smith, & Lutze, 1989) in which the observer adjusts the luminance values of two chromatic lights presented in fast alternation (15–20 Hz) until it appears flicker-free. The stimuli thus generated have the same luminance and size and differ only in the color saturation (LOW: CIE $x = 0.331$, $y = 0.363$; MID: CIE $x = 0.453$, $y = 0.363$; HIGH: $x = 0.621$, $y = 0.363$). Other experimental details are similar to those in intensity.

Saccadic selectivity

As seen in Figure 5, in all search conditions, there was significantly higher saccadic selectivity (paired t test, $p < .05$) toward the relevant saturation interval. For instance, in the MID condition, during search for a MID saturated target, there were more fixations on the MID saturated items than on the LOW or HIGH saturated items.

Strength of biasing

The percentage of fixations on the relevant saturation interval was significantly higher (t test, $p \ll .01$) than the baseline 33.3% predicted by chance. In the LOW condition, the 95% confidence interval was [44.18, 51.36], whereas in the MID and HIGH conditions, it was [51.96, 58.39] and [50.10, 59.34], respectively. Thus, in all conditions, the strength of biasing was significantly higher than by chance, thereby indicating strong effects of top–down guidance. As with intensity and size, the results of a one-way ANOVA show that the strength of biasing did not change with time (see Table 2).

Control experiments

Could the observed results be due to covert attention/recognition only?

One concern is whether the observed saccadic selectivity for the relevant feature interval in Experiments 1, 2, and 3 is due to serial scanning using covert attention and recognition rather than due to parallel processes that provide top–down guidance to the relevant feature interval. While some previous studies suggest that search is serial (Treisman & Gelade, 1980), some others suggest that it is parallel (Desimone & Duncan, 1995), and yet, others suggest that it is a mixture of both (Bichot, Rossi, & Desimone, 2005; Wolfe, Cave, & Franzel, 1989). To address this issue in the context of our search experiments, we conducted additional control experiments in the intensity dimension. We hypothesized that if the observed saccadic selectivity is due to covert serial scanning processes only, then decreasing the presentation time to 120 ms should eliminate the contribution of serial processes and eye movements (Palmer, 1994; Verghese & Stone, 1995); hence, selectivity should disappear. On the other hand, if selectivity is due to a parallel, gain-based mechanism, then even under brief presentation conditions, selectivity for the relevant feature interval should be high. We tested this hypothesis through the following control experiments.

Design of control experiments

Similar to Experiments 1, 2, and 3, we ran three search conditions: LOW, MID, and HIGH, where subjects searched for a target belonging to the LOW, MID, or HIGH intensity interval, respectively. Figure 6a shows a sample trial from the MID condition. Each trial began with a central fixation (for 250 ms), followed by a brief presentation of the search array (for 120 ms). The search array consisted of a 3 × 3 grid of items including one target (rotated L shape) and eight distractors (L shape) belonging to different feature intervals. Pilot experiments in this brief display paradigm revealed that the task was

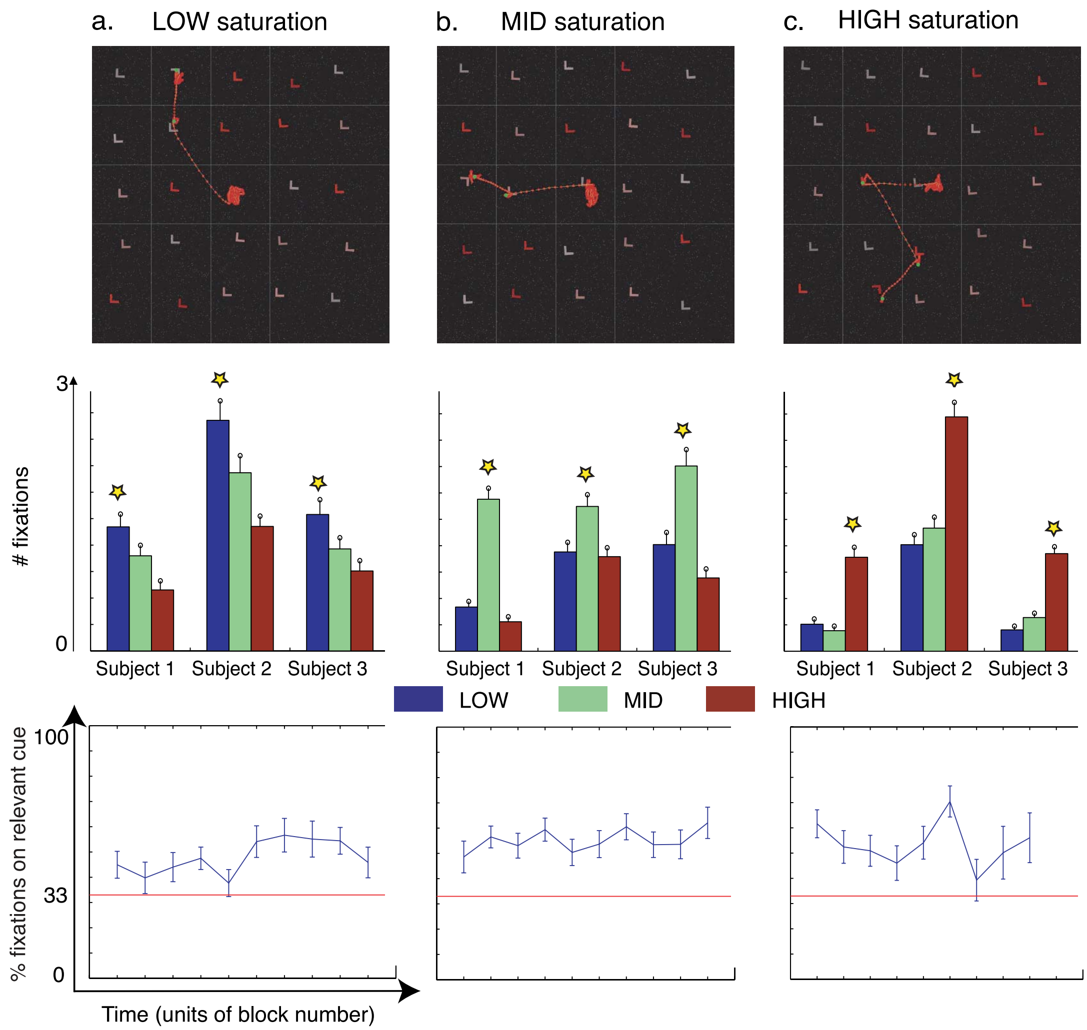

Figure 5. Results in the saturation dimension: (a) The first column shows results during search for a target with LOW saturation. The sample eye trace illustrates that subjects tend to fixate on the relevant distractors of LOW saturation. Statistical analysis of all trials reveals a significantly higher number of fixations on relevant items of LOW saturation (indicated by a yellow star) than those of MID or HIGH saturation (paired t test, $p < .05$). Analysis of fixations as a function of time (measured in units of blocks from 1 to a maximum of 10) reveals that the strength of biasing does not change with time (see Table 2). Similar results are observed for the (b) MID and (c) HIGH conditions. As shown in the second column, when subjects search for a target with MID saturation, they selectively fixate on distractors with MID saturation, as compared to those with LOW or HIGH saturation. These results demonstrate that top–down signals can guide attention to the relevant interval within the intensity dimension.

too difficult with a set size of 5 × 5 items (used in Experiments 1, 2, and 3) and that subjects became frustrated (search accuracy <5%); hence, we decreased the set size to 3 × 3 items. Other parameters such as the size of the target and distractors and interstimulus distance were the same as in Experiments 1, 2, and 3. The search array was followed by a brief presentation of a grid of random two-digit numbers (as part of the "No Cheat" scheme described in Design and analysis of experiments section). Similar to Experiments 1, 2, and 3, subjects were instructed to find the target as fast as possible and report the number at its location. Subjects received feedback on accuracy of target detection. This completed one trial.

Results

Figure 6b shows the results obtained from three naive subjects (who performed 30 blocks of 10 trials each). The task was not easy, as reflected by the low accuracy of target detection (computed as percentage of reports on the target) that varied between 35% and 45% across different search conditions and subjects. Although search accuracy was low, all subjects showed significantly higher number of reports on items belonging to the relevant interval than irrelevant intervals (as determined by a paired t test, $p < .05$). These results confirm that the underlying search mechanism in our experiments is parallel (gain based) rather than serial only.

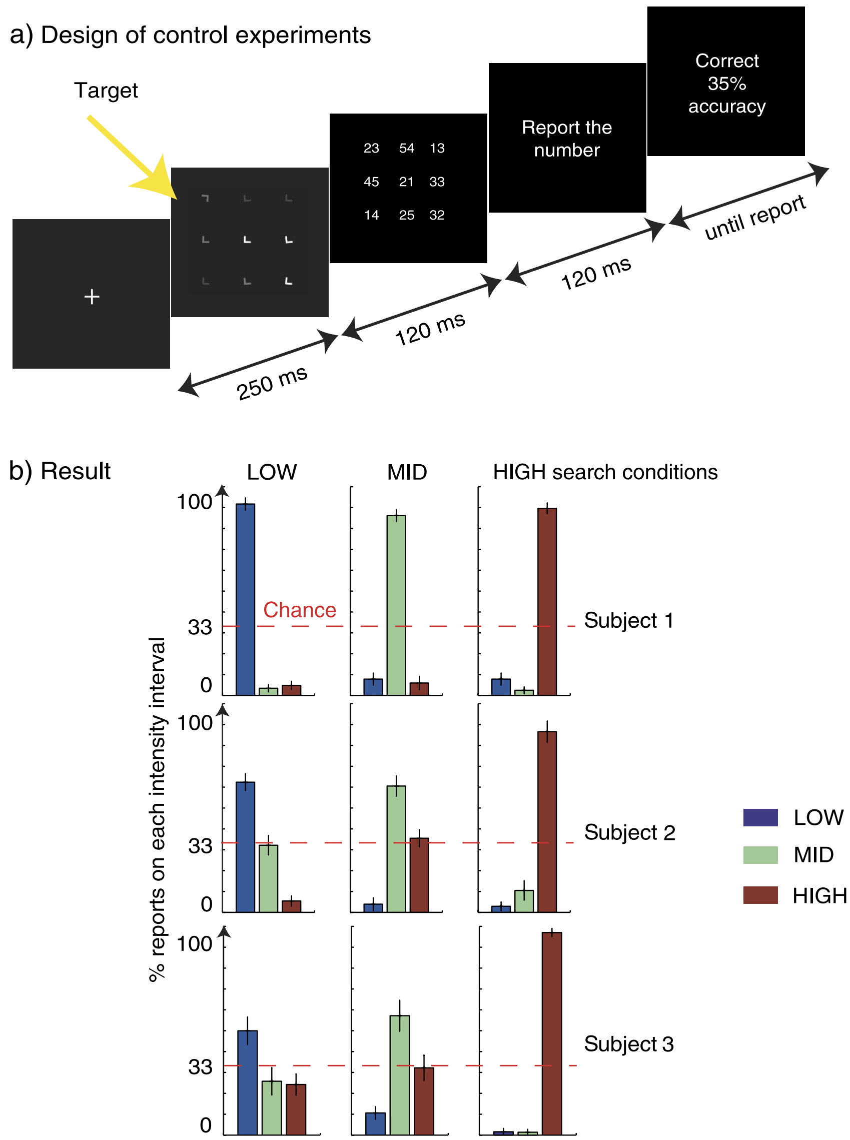

Figure 6. Control experiments and results: (a) Design of the control experiment: Search array was presented for a brief duration (120 ms only) to minimize the role of serial scanning processes. Search array consisted of a 3 $\times$ 3 grid of equal number of items belonging to LOW, MID, and HIGH intensity intervals. In each search condition (LOW, MID, or HIGH), the target was fixed. Subjects were instructed to search for the known target and report the number at its location. (b) The reports were analyzed to determine the percentage of reports on items of each intensity interval. Results of a paired *t* test showed a significantly higher number of reports on items of the relevant interval. For instance, when subjects searched for a MID intensity target, there were more reports on items of the MID intensity interval than on those of the LOW or HIGH intensity interval. These results confirm the role of parallel, gain-based guidance in our search experiments.

Discussion

Reaction time

As reported in earlier studies (Bauer et al., 1996; D'Zmura, 1991; Hodsoll & Humphreys, 2001), in all dimensions tested (intensity, size, and color saturation), the RT was significantly slower ($p < .05$) in the MID condition than in the LOW or HIGH condition (see Figure 3). For instance, search for a medium-sized target was slower than search for a small or big target. This replicates the results of previous studies.

Granularity of top–down attention

In all dimensions tested (intensity, size, and color saturation), our results indicate that subjects could selectively fixate on items belonging to the relevant fine-grained interval defined by the target. These results negate the coarse-grained hypothesis (Figure 1a), which predicts equal number of fixations on all intervals. Instead, they confirm the fine-grained hypothesis (Figure 1b) that, indeed, top–down signals can select the relevant fine-grained interval within a dimension.

Did the target shape provide any guidance?

Although there are 25 items in the display (8 distractors each of LOW, MID, and HIGH intervals, plus 1 target), the average number of fixations is fairly low, between three and six. This raises a concern of whether there was any special guidance due to the target shape. However, pilot experiments confirmed that the target was preattentively indistinguishable from the distractors (RT slope, 23 ms/item, indicating a hard search). This rules out any guidance due to the shape of the target. Also, the low number of fixations (three to six on average) is not an indicator of special guidance to the target for the following reasons: Because most fixations occur only on items of the relevant interval (see Table 1), any model that scans randomly among the 8 items within the relevant feature interval would also predict four fixations on average, which is in agreement with our observations. Thus, the observed results reveal a clear effect of top–down selectivity for the relevant feature interval rather than special guidance to the target shape.

Reconcilitation with previous data

There seems to be an apparent contradiction between the fine-grained hypothesis supported by our results and the "linear separability theory" supported by previous results (Bauer et al., 1996; D'Zmura, 1991; Hodsoll & Humphreys, 2001). The latter reports that search for a MID type target is slower than search for LOW or HIGH type targets (Figure 3), suggesting that top–down signals cannot select the MID interval. On the other hand, the pattern of fixations observed in our results clearly indicate that top–down can select the MID interval. If, indeed, top–down signals can select the fine-grained MID interval, why is search slower? This apparent conflict can be resolved by considering the following model of visual processing: The incoming visual scene is analyzed in each feature dimension by a population of neurons with broad and overlapping tuning curves. The activity of each such neuron is assumed to be modulated by a top–down gain control (similar to Figure 1b). According to this model, a MID type target can be found by selectively promoting the neuron that responds maximally to the MID interval (henceforth referred to as MID neuron). This results in increased salience of all items sharing the MID interval, thereby attracting more fixations as shown in our results. However, because the MID neuron is broadly tuned, it not only responds to the MID type target and distractors but also weakly responds to the LOW and HIGH type distractors. The responses to LOW and HIGH type distractors interfere with search for a MID target, leading to a slow search. A direct consequence of this model is that saccadic selectivity for the MID interval increases as the spacing between LOW, MID, and HIGH increases (i.e., if the LOW and HIGH intervals are widely separated, the MID neuron will respond only to the MID interval, thereby increasing the salience of MID interval items relative to LOW or HIGH). This predicts faster RTs—a prediction that is consistent with existing behavioral reports (Bauer et al., 1996).

Time scale of biasing

Our results show that the strength of biasing does not change as a function of time within a session. This suggests that the top–down bias that is set up initially during the training period (first 20 trials in the session) does not change in the rest of the session (lasting up to an hour). However, this does not rule out short-term priming (Maljkovic & Nakayama, 1994) where the strength of biasing may improve within a few trials, nor does it rule out long-term priming or perceptual priming (Bichot & Schall, 1999) effects where the strength of biasing may improve over a period of days.

Implications for visual search behavior and performance

Although previous studies based on RT measures report that search in the MID condition is slower than in the LOW or HIGH condition, they do not reveal the underlying cause or granularity of top–down signals. Here, we used eye-tracking methods to infer top–down guidance by analyzing whether subjects fixate on the relevant fine-grained interval or not. On the basis of the results of our study, we conclude that top–down signals carry fine-grained information that can specify the relevant feature interval rather than coarse-grained information that can only specify the relevant feature dimension. This holds implications for several existing visual search theories and experiments. Theories such as dimension-weighting accounts (Found & Muller, 1996; Muller et al., 1995), which suggest a single gain control term per dimension, predict equal gain on LOW, MID, and HIGH intervals within the dimension and, hence, cannot account for the greater number of fixations on the MID interval in our MID condition. Clearly, such theories need to be updated from a coarse-grained, one-gain factor per feature dimension to a fine-grained, one-gain factor per feature interval. The conditions for efficient search should be revised: Search should be easy not only when the target and distractors differ in some dimension but also when they differ in some interval within a dimension. This model also accounts for some observed effects in search asymmetry. For instance, faster search for a saturated red than for a desaturated red (Treisman & Gormican, 1988) can be explained by the model as saturated red activates a HIGH interval, whereas desaturated red activates the MID interval and the background activates the LOW interval.

Acknowledgments

Some of these results were presented earlier as a talk (Navalpakkam & Itti, 2004) and poster (Navalpakkam & Itti, 2005a). This research was supported by the National Geospatial–Intelligence Agency, the National Science Foundation, and the Human Frontier Science Program. The authors affirm that the views expressed herein are solely their own and do not represent the views of the U.S. government or any agency thereof.

Commercial relationships: none.
Corresponding author: Vidhya Navalpakkam.
Email: navalpak@usc.edu.
Address: HNB 06, 3641 Watt way, Los Angeles, CA 90089-4620.

References

Bacon, W. J., & Egeth, H. E. (1997). Goal-directed guidance of attention: Evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 948–961. [PubMed]

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

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

Bichot, N. P., & Schall, J. D. (1999). Saccade target selection in macaque during feature and conjunction visual search. *Visual Neuroscience, 16,* 81–89. [PubMed]

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience, 2,* 671–676. [PubMed] [Article]

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

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science, 248,* 1556–1559. [PubMed]

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222. [PubMed]

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458. [PubMed]

D'Zmura, M. (1991). Color in visual search. *Vision Research, 31,* 951–966. [PubMed]

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 32–39. [PubMed]

Found, A., & Muller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a ''dimension-weighting'' account. *Perception & Psychophysics, 58,* 88–101. [PubMed]

Hodsoll, J., & 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. *Perception & Psychophysics, 63,* 918–926. [PubMed] [Article]

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

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews: Neuroscience, 2,* 194–203. [PubMed]

Kaptein, N. A., Theeuwes, J., & van der Heijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1053–1069.

Kumada, T. (2001). Feature-based control of attention: Evidence for two forms of dimension weighting. *Perception & Psychophysics, 63,* 698–708. [PubMed] [Article]

Lee, D. K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature Neuroscience, 2,* 375–381. [PubMed] [Article]

Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 10530–10535. [PubMed] [Article]

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory and Cognition, 22,* 657–672. [PubMed]

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229,* 782–784. [PubMed]

Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *The Journal of Neuroscience, 14,* 2178–2189. [PubMed] [Article]

Muller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics, 57,* 1–17. [PubMed]

Nagy, A. L., & Sanchez, R. R. (1990). Critical color differences determined with a visual search task. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 7,* 1209–1217. [PubMed]

Navalpakkam, V., & Itti, L. (2004, October). Modeling the influence of target and distractor knowledge on visual search. In *Proceedings of the Society for Neuroscience: Annual Meeting (SFN'04).*

Navalpakkam, V., & Itti, L. (2005a). Attention can be guided to the relevant feature category [Abstract]. *Journal of Vision, 5*(8), 1007a, http://journalofvision.org/5/8/1007/, doi:10.1167/5.8.1007.

Navalpakkam, V., & Itti, L. (2005b). Modeling the influence of task on attention. *Vision Research, 45,* 205–231. [PubMed]

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research, 34,* 1703–1721. [PubMed]

Pashler, H. (1987). Target-distractor discriminability in visual search. *Perception & Psychophysics, 41,* 385–392. [PubMed]

Pokorny, J., Smith, V. C., & Lutze, M. (1989). Heterochromatic modulation photometry. *Journal of the Optical Society of America A Optics and Image Science, 6,* 1618–1623. [PubMed]

Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision Research, 46,* 1886–1900. [PubMed]

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience, 5,* 631–632. [PubMed] [Article]

Shen, J., Reingold, E. M., & Pomplun, M. (2000). Distractor ratio influences patterns of eye movements during visual search. *Perception, 29,* 241–250. [PubMed]

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology A Human Experimental Psychology, 40,* 201–237. [PubMed]

Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 652–676. [PubMed]

Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136. [PubMed]

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 15–48. [PubMed]

Treue, S., & Martinez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399,* 575–579. [PubMed]

Treue, S., & Maunsell, J. H. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature, 382,* 539–541. [PubMed]

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y. H., Davis, N., & Nuflo, F. (1995). Modeling visual-attention via selective tuning. *Artificial Intelligence, 78,* 507–545.

Verghese, P., & Stone, L. S. (1995). Combining speed information across space. *Vision Research, 35,* 2811–2823. [PubMed]

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., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 483–502. [PubMed]

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology, 15,* 419–433. [PubMed]

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

Zohary, E., & Hochstein, S. (1989). How serial is serial processing in vision? *Perception, 18,* 191–200. [PubMed]