

Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli

Division of Biology, California Institute of Technology,
Pasadena, CA, USA, &
Institute of Computational Science,
Swiss Federal Institute of Technology (ETH),
Zurich, Switzerland

Wolfgang Einhäuser



Ueli Rutishauser

Computation and Neural Systems,
California Institute of Technology,
Pasadena, CA, USA



Christof Koch

Division of Biology, California Institute of Technology,
Pasadena, CA, USA



In natural vision both stimulus features and task-demands affect an observer's attention. However, the relationship between sensory-driven ("bottom-up") and task-dependent ("top-down") factors remains controversial: Can task-demands counteract strong sensory signals fully, quickly, and irrespective of bottom-up features? To measure attention under naturalistic conditions, we recorded eye-movements in human observers, while they viewed photographs of outdoor scenes. In the first experiment, smooth modulations of contrast biased the stimuli's sensory-driven saliency towards one side. In free-viewing, observers' eye-positions were immediately biased toward the high-contrast, i.e., high-saliency, side. However, this sensory-driven bias disappeared entirely when observers searched for a bull's-eye target embedded with equal probability to either side of the stimulus. When the target always occurred in the low-contrast side, observers' eye-positions were immediately biased towards this low-saliency side, i.e., the sensory-driven bias reversed. Hence, task-demands do not only override sensory-driven saliency but also actively countermand it. In a second experiment, a 5-Hz flicker replaced the contrast gradient. Whereas the bias was less persistent in free viewing, the overriding and reversal took longer to deploy. Hence, insufficient sensory-driven saliency cannot account for the bias reversal. In a third experiment, subjects searched for a spot of locally increased contrast ("oddity") instead of the bull's-eye ("template"). In contrast to the other conditions, a slight sensory-driven free-viewing bias prevails in this condition. In a fourth experiment, we demonstrate that at known locations template targets are detected faster than oddity targets, suggesting that the former induce a stronger top-down drive when used as search targets. Taken together, task-demands can override sensory-driven saliency in complex visual stimuli almost immediately, and the extent of overriding depends on the search target and the overridden feature, but not on the latter's free-viewing saliency.

Keywords: attention, eye-movements, human, top-down, bottom-up, salience

Citation: Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, 8(2):2, 1–19, <http://journalofvision.org/8/2/2/>, doi:10.1167/8.2.2.

Introduction

In natural vision, human observers sequentially allocate focal attention to subsets of the scene (James, 1890). Such attention shifts are typically associated with shifts in eye-position (Rizzolatti, Riggio, Dascola, & Umiltà, 1987). Notwithstanding the importance of observer's task in guiding eye-movements when viewing natural scenes (Buswell, 1935; Yarbus, 1967) or performing everyday activities (Land & Hayhoe 2001), purely sensory-driven models predict gaze allocation in complex stimuli surprisingly well (Dickinson, Christensen, Tsotsos, & Olofsson, 1997; Itti & Koch, 2000; Koch & Ullman, 1985; Parkhurst, Law, & Niebur, 2002; Peters, Iyer, Itti, & Koch,

2005; Tatler, Baddeley, & Gilchrist, 2005). This raises the question as to what extent sensory-driven ("bottom-up") and task-dependent ("top-down") signals interact in guiding human attention in complex visual stimuli.

Visual search tasks allow controlled, feature-specific manipulations of top-down attention. The relation of reaction times to set size provides a good measure of search difficulty and has been the basis of elaborate models of covert visual search (Bacon & Egeth, 1997; Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989). Recent research on overt visual search, i.e., search by gaze allocation, extends such models: Rao, Zelinsky, Hayhoe, and Ballard (2002) extend Wolfe et al.'s guided search model by using multi-scale filters to predict search in complex scenes; similarly, the model of Navalpakkam

and Itti (2006) selectively up-regulates features in a saliency map to maximize the targets' signal-to-noise ratio. Other models exploit scene statistics (Oliva, Torralba, Castelhano, & Henderson, 2003) or Bayesian optimality (Najemnik & Geisler, 2005) to learn spatial priors for search. Despite a large body of modeling work, the interaction of top-down and bottom-up signals for overt visual search in complex stimuli remains largely unknown.

Although a number of brain areas have been implicated in the representation of saliency (for reviews, see Colby & Goldberg, 1999; Treue, 2003), electrophysiological data suggest that areas of high convergence between top-down and bottom-up projections play a particular role in its computation, such as the pulvinar (Robinson & Petersen, 1992) or V4 (Bichot, Rossi, & Desimone, 2005; Mazer & Gallant, 2003; Ogawa & Komatsu, 2006). Neuronal activation in the frontal eye-fields (FEF) reflects bottom-up saliency and task-demands (Thompson & Bichot, 2005). Using micro-stimulation, Armstrong, Fitzgerald, and Moore (2006) recently demonstrated that projections from FEF modulate the size of V4 receptive fields akin to attentional modulation, providing a mechanism for top-down modulation of bottom-up signals. Consequently, understanding the interaction between bottom-up and top-down signals is key to understanding the neural substrates of saliency computation.

Henderson, Brockmole, Castelhano, and Mack (2007) show that typical measures of bottom-up saliency do not account for fixations of observers engaged in a visual search task. Consistent with this psychophysical finding, Ipata, Gee, Gottlieb, Bisley, and Goldberg (2006) show that actively ignoring a stimulus of high bottom-up saliency reduces responses in LIP, an area associated with high-level representation of attention. Although this is suggestive of a task fully overriding bottom-up saliency, several issues remain open: Can search for an unrelated item override a robust bottom-up bias? Is such overriding immediate? Does it depend on the type of search, i.e., on whether the appearance of the target is exactly known? Is bottom-up information used for target search, even if its saliency is ignored? Here we apply different modifications (contrast gradients, flicker) to natural images to bias bottom-up saliency. We measure how fixation locations depend on task (free-viewing, template search, oddity search) and target location relative to the bottom-up bias. Using this setting we quantify the interaction of bottom-up (sensory-driven) and top-down (task-dependent) signals on human overt attention with naturalistic stimuli.

Methods

Stimuli

All stimuli were based on 128 photographs of outdoor scenes, which contain few or no man-made objects

(Zürich Natural Image Database; Einhäuser, Kruse, Hoffmann, & König, 2006a) and are available from the authors' Web page (<http://n.ethz.ch/~einhaeuw/download/ZurichNatDB.tar.gz>). Stimuli were in 8-bit grayscale at a resolution of 1024×768 pixels. To vary the natural appearance of the images, different levels of noise were added to their phase. Images were transformed into Fourier space, noise of mean zero and standard deviation σ_{noise} was added to the phase of half of the coefficients (the phases of the remaining coefficients were changed accordingly to preserve symmetry), and the image was transformed back. Since earlier experiments (Einhäuser et al., 2006b) had demonstrated a sharp transition between subjective "natural" and "non-natural" appearance at around $\sigma_{\text{noise}} = 0.4\pi$, we used four different levels of noise to achieve about equal amounts of "natural" and "non-natural" looking images: no noise, $\sigma_{\text{noise}} = 0.2\pi$ (Figure 1); $\sigma_{\text{noise}} = 0.6\pi$ as well as entirely random phase drawn from a uniform distribution between 0 and 2π . In Experiment 4, only $\sigma_{\text{noise}} = 0.2\pi$ was used. No image was used more than once (Experiments 1 and 2), twice (Experiment 3), or thrice (Experiment 4) at the same noise level for the same observer.

Bottom-up saliency

We used two different mechanisms to bias sensory-driven saliency towards one side of the image, static gradients of luminance-contrast ("static condition") or 5 Hz flicker ("dynamic condition").

Static condition

In the static condition (Experiments 1 and 3), we increased saliency towards one side of the image by applying a contrast gradient as follows (Figure 1):

$$I(x, y) = I_0(x, y) + \beta(x)(I_0(x, y) - \langle I_0 \rangle), \quad (1)$$

where I_0 denotes the intensity of the raw image (after applying noise, see above), $\langle I_0 \rangle$ its mean intensity over the image and β is given by

$$\beta(x) = \frac{(2x-L-1)}{L} \text{ ("right side salient")} \quad (2a)$$

or

$$\beta(x) = -\frac{(2x-L-1)}{L} \text{ ("left side salient")}, \quad (2b)$$

where x denotes the horizontal image coordinate in pixels ($x = 1 \dots 1024$) and L the width of the image ($L = 1024$).

If values of $I(x, y)$ exceeded the 8-bit dynamic range, they were clipped (minimum 0, maximum 255). This adds a non-linear distortion of intensity values. However, most pixels are sufficiently far from the bounds of the dynamic

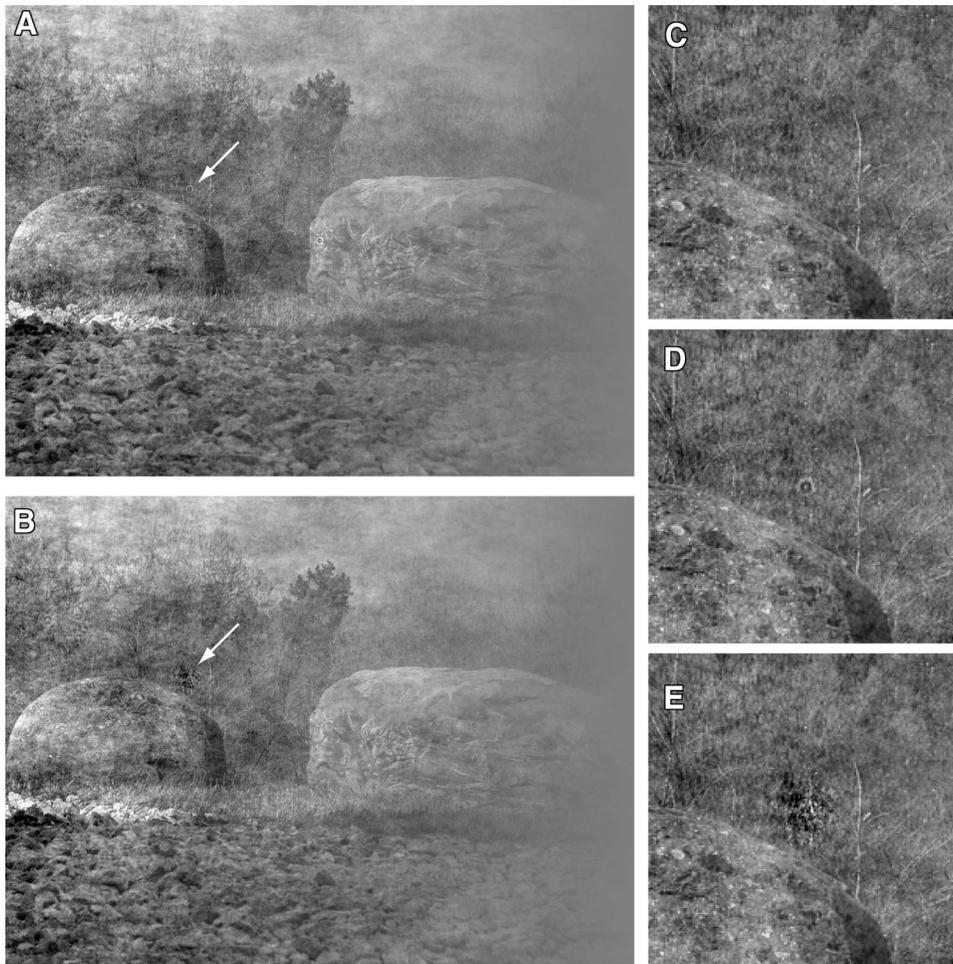


Figure 1. Stimuli in static condition (Experiments 1 and 3). One of the test images with small amount of phase noise ($\sigma_{\text{noise}} = 0.2\pi$) and high-contrast (salient) side towards the left. (A) Template search (Experiment 1): A “bull’s-eye” is present in the center (search template) and towards the top right of the left stone (search target, arrow not present in actual display). (B) Oddity search (Experiment 3): Locally increased contrast defines the target (for illustration target is at the same location as in panel A, arrow not present in actual display). (C) Magnification of 200×200 pixel region around target location before adding the target. (D) The same region as in panel C with bull’s-eye target (magnification of panel A). (E) Same region as in panel C, with oddity target (magnification of panel B). The figure is intended for illustration only. High-resolution versions of the examples are available at <http://n.ethz.ch/~einhaeuw/download/jovTopDownSuppl/>.

range. Across all 1024 stimuli with gradient (4 phases \times 128 images \times 2 gradient directions), only $0.3\% \pm 0.3\%$ (mean \pm *SD* over stimuli) of pixels were affected by clipping.

Note that the present method of introducing contrast gradients differs slightly from an earlier proposal: Here we modulate from zero contrast ($\beta = -1$) to doubling the contrast ($\beta = +1$) within the same image, whereas in Einhäuser et al. (2006b), we varied contrast from no modification to either zero or doubling ($\beta = 0$ to $\beta = \pm 1$). There are 2 reasons for this change: First, the present modification leaves the center of the image unmodified in all conditions; second, the overall dynamic range of the modification is double that of the earlier study. This doubles the horizontal bias in free-viewing because a negative modification in one direction has an effect equivalent to the positive modification in the opposing direction (linearity; Einhäuser et al., 2006b).

Dynamic condition

In the dynamic condition (Experiment 2), 5-Hz flicker replaced the contrast gradient (Figure 2). The flicker modulated one side of the image around half the dynamic range of the static condition, i.e., for stimuli with saliency bias towards the right side β was given by

$$\beta(x, t) = \begin{cases} 0 & x < \frac{L}{2} \\ 0 & x > \frac{L}{2} \text{ and } t \in \{[0 \text{ ms}, 50 \text{ ms}], [100 \text{ ms}, 150 \text{ ms}], \dots\} \\ -0.5 \frac{(2x-L-1)}{L} & x > \frac{L}{2} \text{ and } t \in \{[50 \text{ ms}, 100 \text{ ms}], [250 \text{ ms}, 300 \text{ ms}], \dots\} \\ +0.5 \frac{(2x-L-1)}{L} & x > \frac{L}{2} \text{ and } t \in \{[150 \text{ ms}, 200 \text{ ms}], [350 \text{ ms}, 400 \text{ ms}], \dots\} \end{cases} \quad (3)$$

where t denotes time after stimulus onset. A saliency increase to the left side was achieved analogously by interchanging $x < L/2$ and $x > L/2$ in Equation 3.



Figure 2. Movie of stimulus in dynamic condition (Experiment 2). Stimulus of Figure 1 in dynamic condition, salient side (flicker) to the right. Target location as in Figure 1.

Search targets

We used two different types of targets for visual search, “bull’s-eye” patterns (“template search”) and local increases of luminance-contrast (“oddyity search”).

Bull’s-eye targets (template search)

For template search (Figures 1A, 1D, and 2), “bull’s-eyes” were embedded into the images after adding phase noise and applying the contrast modification. They were constructed around a location (x_0, y_0) , where $I(x, y)$ denotes the image intensity before adding the bull’s-eye:

$$I'(x, y) = I(x, y) + 0.5G(x, y)I(x, y), \quad (4)$$

with

$$G(x, y) = \exp \left[-\frac{(x - x_0)^2}{50} - \frac{(y - y_0)^2}{50} \right] \times \cos \left[\frac{3\pi}{8} \sqrt{(x - x_0)^2 + (y - y_0)^2} \right] \quad (5)$$

(all units in pixels) defining the “bull’s-eye” as circular Gabor pattern. One of the bull’s-eyes was always embedded centrally as template for the observer, another one (target) at a location unknown to the observer (see below).

High-contrast targets (oddyity search)

For “oddyity search” (Figures 1B and 1E), we defined the target by locally increasing luminance contrast in analogy to the global gradients of Equation 1. The image I' with target at (x_0, y_0) is given by

$$I'(x, y) = I(x, y) + 2\gamma(x, y)(I(x, y) - \langle I(x, y) \rangle), \quad (6)$$

where $\langle \cdot \rangle$ denotes spatial average as in Equation 1, $I(x, y)$ the image intensity before adding the target as in Equation 4, and

$$\gamma(x, y) = \exp \left[-\frac{(x - x_0)^2}{256} - \frac{(y - y_0)^2}{256} \right] \quad (7)$$

defines the local contrast modification (increase) around the target location (x_0, y_0) . As for the contrast gradients, values beyond the dynamic range are clipped, occasionally yielding an additional non-linear distortion on the intensity values.

Note that *negative* contrast modification is an alternative option for defining the oddity target. We used the positive modification for the following reasons: (i) both negative and positive *strong* local modifications attract fixations, but only negative modifications exhibit a substantial size dependence (Einhäuser et al., 2006b). Pilot experiments also suggested that local low-contrast spots are quickly found, whereas high-contrast targets are not (for visibility, see also Experiment 4). Hence, positive modifications are “harder” targets. (ii) Local negative

modifications have little to no effect in low contrast regions, whereas locally increased contrast is effective and visible even in regions where contrast is already very high (until all pixels reach the bounds of the dynamic range).

Search conditions

The location for the target (oddity or template search) was restricted to two 45° segments of an annulus: The target was located at least 128 pixels (3.6° of visual angle) and maximally 384 pixels (half image height, 10.9° visual angle) away from the image center. In addition, the target could only occur within an angle of $\pm 22.5^\circ$ from the midline. In “unbiased search,” the target was randomly placed with regard to the bottom-up saliency (that is, it was equally likely to be found in the more than in the less salient side). In “inverse bias search,” the target was always on the low-salient side (low contrast/no flicker) for stimuli with saliency bias, and on a random side for stimuli without saliency bias.

Subjects

Nineteen volunteers from the Caltech Community (age 20–37) participated in the experiments, five in each of [Experiments 1–3](#) and four in [Experiment 4](#). All subjects had uncorrected normal vision and were naive to the purpose of the study. All subjects gave written informed consent to participation. All experiments conformed to National and Institutional Guidelines for experiments in human subjects and with the Declaration of Helsinki.

Setup

Experiments were conducted in a dark room specifically designed for psychophysical experiments. Observers were seated 80 cm from a 20-inch CRT monitor, on which the stimuli subtended 29×22 degrees of visual angle. The gamma of the monitor was corrected to achieve a linear mapping between pixel values and displayed luminance. Maximum luminance was set to 26 cd/m^2 , while ambient light levels were below 0.01 cd/m^2 . All presentation and data analysis were performed using the MatLab (Mathworks, Natick, MA) programming environment and its Psychophysics and Eyelink Toolbox extensions (Brainard, 1997; Cornelissen, Peters, & Palmer, 2002; Pelli, 1997, <http://psychtoolbox.org>).

Throughout the experiment, the observer’s right eye-position was recorded at 1000 Hz using an Eyelink-1000 (SR Research, Osgoode, ON, Canada) non-invasive infrared eye-tracking system. We used the manufacturer’s software for calibration, validation, drift-correction, and determining periods of fixation. The calibration of the eye-tracking system was validated every 48 trials and

recalibrated if necessary. An additional drift-correction was performed whenever an observer failed to fixate within about 1.4° (50 pixels) of an initial central fixation cross within 5 s.

In all experiments and conditions, each trial started with a central fixation cross which observers had to fixate for 500 ms to trigger stimulus onset. During free-viewing tasks, stimuli were presented for 3 s; during search tasks, stimuli were presented until the observer fixated the target for 500 ms or 10 s had elapsed without the observer finding the target. In [Experiments 1–3](#), observers had to judge after each presentation in all tasks and conditions, whether or not the presented stimulus had been “natural,” which we defined as “resembling the image of a real-world scene.”

Experiment 1

In [Experiment 1](#), each block consisted of 96 trials. In a third of the trials (32), the salient side was left, in a third the salient side was right, in the remainder no contrast gradient was added to the stimulus. Each of the 4 phase-noise levels was used equally often (8 times per block) for each of these 3 directions (salient side left, salient side right, no gradient). In the first and fourth block (“free-viewing”), subjects had no task but still had to judge whether the image had been natural or not. In the 2nd block, subjects searched for a bull’s-eye target, which was embedded with equal probability on either side of the stimulus (unbiased search). In the 3rd block, the bull’s-eye target was always embedded in the low contrast, i.e., low-salient, side of the image (inverse bias search).

Experiment 2

[Experiment 2](#) was identical to [Experiment 1](#) with the exception of the static bias (contrast gradient) being replaced by the dynamic one (flicker).

Experiment 3

[Experiment 3](#) includes oddity search in addition to template search, which is a replication of [Experiment 1](#). Each block in [Experiment 3](#) consisted of 48 trials. As in [Experiments 1](#) and [2](#), saliency was biased to the left, to the right or unbiased each in a third of the trials, and each noise level was used equally often for each bias. In four blocks (1st, 6th, 7th, and 12th), observers performed a free-viewing task. In the remaining blocks, observers performed unbiased search (2nd, 4th, 8th, and 10th) or inverse bias search (3rd, 5th, 9th, and 11th). In half of these blocks, the target was the bull’s-eye as in [Experiments 1](#) and [2](#), i.e., observers performed template search (2nd, 3rd and 10th, and 11th block for CH, SS, and TD, or 4th, 5th, 8th, and 9th block in CR and JB). In the remaining blocks, observers searched for a spot of increased contrast, i.e.,

they performed oddity search. Note that the total number of trials for free-viewing and template search was the same as in Experiments 1 and 2, and 2×96 trials of oddity search were performed in addition (Table 1). Due to the larger amount of trials, Experiment 3 was split in two sessions on consecutive days: blocks 1 through 6 were performed on the first day, blocks 7 through 12 on the second day.

Experiment 4

Experiment 4 uses reaction time measurements to assess the visibility of targets at known locations. Stimuli were as in the search tasks of Experiment 3, but the target could only occur at one of two predefined locations on the horizontal midline ± 128 pixels ($\pm 3.6^\circ$) from the center. In half of the trials (target-present trials), targets occurred at one of these locations; in the other half (target absent trials), no target was present. Observers were asked to “respond as quickly as possible” as to whether or not a target was present by pressing one of two buttons of a gamepad. Subjects were instructed to fixate the center, which was marked by the search template (template search) or a small fixation cross (oddity search). Trials were aborted if subjects broke fixation, i.e., deviated more

than 1.4° from the center for more than 100 ms. Each observer performed 4 blocks of 96 trials each (Table 1): the equivalent of unbiased and inverse bias search for each of the search targets (template/oddity).

Instructions

Task description

Observers were instructed to minimize their head-movements during the experiment. For the “natural”/“non-natural” judgment task, “natural” was defined as “representing the image of a real world scene.” Observers had to fixate the center to start each trial, but were explicitly told that they “are free to move [their] eyes” when the “picture is shown.” Before the first block (“free viewing” in all cases), no instruction regarding search was given. Before the first template search block, observers were instructed that they “will be presented a central ‘bull’s-eye’ like target,” that “a second identical target is hidden somewhere in the image,” and that their task is “to find this second target as quickly as possible and fixate it with [their] eyes until the picture disappears.” Before the first oddity block, they were instructed that they have to “find as quickly as possible a region of increased contrast,

Experiment	Block no.	Trials/block	Bias	Task	Target Location	Target type
1	1, 4	96	Static	Free-viewing	–	–
	2	96	Static	Search	Random	Bull’s-eye
	3	96	Static	Search	Low contrast side	Bull’s-eye
2	1, 4	96	Dynamic	Free-viewing	–	–
	2	96	Dynamic	Search	Random	Bull’s-eye
	3	96	Dynamic	Search	No flicker side	Bull’s-eye
3–group 1 (CH, SS, TD)	1, 6, 7, 12	48	Static	Free-viewing	–	–
	2, 10	48	Static	Search	Random	Bull’s-eye
	3, 11	48	Static	Search	Low contrast side	Bull’s-eye
	4, 8	48	Static	Search	Random	High contrast
	5, 9	48	Static	Search	Low contrast side	High contrast
3–group 2 (CR, JB)	1, 6, 7, 12	48	Static	Free-viewing	–	–
	2, 10	48	Static	Search	Random	High contrast
	3, 11	48	Static	Search	Low contrast side	High contrast
	4, 8	48	Static	Search	Random	Bull’s-eye
	5, 9	48	Static	Search	Low contrast side	Bull’s-eye
4–group 1 (AK, MM)	1	96	Static	Detect	Random	Bull’s-eye
	2	96	Static	Detect	Low contrast side	Bull’s-eye
	3	96	Static	Detect	Random	High contrast
	4	96	Static	Detect	Low contrast side	High contrast
4–group 2 (AC, PJ)	1	96	Static	Detect	Low contrast side	High contrast
	2	96	Static	Detect	Random	High contrast
	3	96	Static	Detect	Low contrast side	Bull’s-eye
	4	96	Static	Detect	Random	Bull’s-eye

Table 1. Paradigm overview.

i.e., a spot that might appear darker, brighter, or more in focus than its surrounding.” Before the inverted blocks, subjects were told that the task is “exactly the same as in the previous block.” Before the final free-viewing block, observers were told that they “do not have to search,” and that the “task is exactly the same as in the first block.” To encourage quick search, subjects were told that the number of trials is constant and that payment is independent of performance. If subjects asked questions regarding the purpose of the experiment, they were told that “we are interested as to where you look at,” and—in case this did not satisfy them that “the full purpose can only be revealed after the experiment is completed.”

In the search tasks of [Experiments 1–3](#), the stimulus disappeared as soon as the target was found or if 10 s had elapsed. Auditory (high pitch vs. low pitch sound) and text feedback was provided immediately afterwards to indicate successful or non-successful conclusion of a search trial. In [Experiment 4](#), auditory and text feedback was used to indicate the correctness of the response or trial abortion due to broken fixation.

Debriefing

After the experiment, subjects were asked whether they “noticed differences between the search blocks.” If they did not report any, they were explicitly asked whether some “condition had been easier.” If they did not make any reference to target location in their reply, subjects were further asked, whether they “had noticed that the target appeared preferentially in specific locations.” Only after this debriefing was complete, the subjects were—on request—told that there were differences between the search blocks and that the target in some conditions only had occurred in the low-contrast side.

Eye-movement analysis

Fixations

Unless otherwise stated, all eye-movement analysis was restricted to periods of fixations. These were defined using a combined velocity and acceleration threshold according to the eye-tracker’s default settings. The distance of subsequent fixations falls below 0.5° for only 3.9% of the data, which renders it unlikely that jittering around a single location (e.g., by microsaccades) is counted as more than one fixation in a substantial fraction of cases.

Statistical analysis of eye-movement data

For each task (free viewing/unbiased search/inverse bias search) and subject, we separately analyzed the horizontal location of the first 10 fixations (excluding the “0th” central fixation). Unless otherwise stated, we separately averaged these fixation locations for stimuli with salient side on the left and salient side on the right.

By subtracting these means from the respective means in stimuli without saliency modification, we obtain the time course of the fixation bias induced by the saliency increase. This measure, “relative horizontal fixation location,” is insensitive to any general individual biases to fixation and can thus be averaged over subjects within each task and condition. Due to the dependence of successively fixated locations (if the first fixation is far left of the midline, the second fixation also has a high probability to be left of the midline, etc.), we only analyze the first fixation’s absolute position statistically. Instead, we base the quantitative analysis on the relative position of successive fixations (“saccade directions”). Unlike successive fixations, we can treat successive saccade directions as nearly independent (although the effects of inertia and inhibition-of-return prevent independence from holding in full).¹ Even in a situation, where a saliency modification (gradient, flicker, etc.) perfectly biases fixation, we would not expect a bias on average *saccade directions* over prolonged viewing time: Eventually, eye-position reaches the image boundary. Hence, we do not expect to find any substantial effect on the *mean* saccade direction over all fixations in a trial. Consequently, we analyze each fixation (1st fixation in a trial, 2nd fixation, etc.) separately by means of pair-wise comparisons. For clarity, we report uncorrected significance levels throughout. As we at least perform 10 comparisons (one per fixation) in each condition, we base our main conclusions only on *p*-values smaller than 0.005, which corresponds to a Bonferroni corrected level of 0.05. Note that any remaining dependency between successive saccade directions would require a less strict correction, and Bonferroni therefore provides the most conservative correction. Neither will any of the main conclusions be based on so-called “non-significant” results, i.e., the fact that a null-hypothesis cannot be rejected at a given significance level.

Search time in Experiments 1–3

As the present study focuses on biases in overt attention, for behavioral data (time or fixations to find target) of individuals in search tasks of [Experiments 1–3](#), the reader is referred to the [supplementary material](#).

Latencies and dwell times

We define the latency of the first saccade in each trial as the time from stimulus onset to the onset of the first saccade. For each fixation following the initial one (“0th”), we report dwell times. Again we base this analysis on the definition of saccades and fixations using the eye-tracking software’s default thresholds. An overview over latencies and dwell times is presented in the [supplementary material](#).

Results

Experiment 1: Static biases are actively countermanded

Five naive observers viewed modified photographs of outdoor scenes, while their eye-positions were recorded. As expected from previous experiments (Einhäuser et al., 2006b), observers judged the low noise images almost always as “natural” (no noise: $(97.2 \pm 6.0)\%$; $\sigma_{\text{noise}} = 0.2\pi$: $(95.7 \pm 4.3)\%$ – mean \pm SD across subjects) and high noise images as “non-natural” ($\sigma_{\text{noise}} = 0.6\pi$: $(7.2 \pm 15.2)\%$; random phase: $(6.7 \pm 13.7)\%$). The fact that this judgment does not depend on the task ($p = 0.57$ for main effect of task, two-factor ANOVA on task and noise-level) suggests that observers paid attention to the stimulus also in free-viewing, and that the search task does not interfere with this global task.

In the 1st and 4th experimental block, observers viewed each image for 3 s and only performed the natural/non-natural judgment. The first fixation is significantly biased to the high-contrast side ($p = 0.0008$, uncorrected t -test, Figure 3A). Successive fixations remain preferentially on this high-saliency side, reaching a plateau at approximately the second fixation. Consistent with the absolute position, the first saccade shows a significant direction bias to the high-saliency side ($p = 0.0002$, Figure 3B). The second tends to continue into this direction, after which the position bias remains constant with the exception of a slight rebound at the 5th fixation. This shows that—under free-viewing conditions—the bottom-up bias induced by the contrast gradient starts immediately, with the first fixation after stimulus onset.

In the 2nd experimental block, observers had to detect a target that was hidden with equal probability on either side of the image in addition to the natural/non-natural judgment. In this “unbiased search” condition, the bias induced by the gradient disappears entirely and immediately (uncorrected t -tests: $p > 0.11$ for any fixation, Figure 3C). Similarly, none of the saccade directions has any significant bias in any direction (Figure 3D, $p > 0.07$). Hence, our experiments show no evidence for a bias on fixation during search. This suggests that a search task can quickly override the robust sensory-driven effect of the contrast gradient.

In the 3rd experimental block, the search target always occurred on the low-contrast, i.e., low-saliency, side. Note that this was not a spatial bias as low contrast could appear equally likely on either side. In this “inverse biased” search, the first fixation is already biased to the low contrast side ($p = 0.0005$, t -test, uncorrected, Figure 3E). While this inverted bias starts already with the first saccade, the trend to saccade towards the low contrast side remains for the first 8 saccades, reaching (uncorrected) significance for the first 6 (Figure 3F). This

suggests that even in trials in which initial saccades follow the bottom-up bias to go towards the high-contrast region, there is a constant top-down drive towards the low contrast side. On average, however, even for the first saccade, this top-down drive (at least partially) supersedes the bottom-up drive. This suggests that observers are not merely ignoring the high-salient cue but actively countermanding it. In addition, the comparison of the time courses of saccade direction reveals that net bottom-up effects (drive to high-contrast side in free viewing, Figure 3B) persists only for the first few saccades, while top-down drive (to low contrast side in inverted bias search, Figure 3F) can persist throughout a trial.

Quick and implicit learning of biases

In debriefing, none of the observers reported to be aware of the fact that the target always occurred in the low-contrast portion of the image. Although we did not aim at testing implicit versus explicit learning formally, this result suggests that observers may *implicitly* learn to countermand bottom-up saliency. This raises the question as to how rapid this implicit learning takes place over the course of the experiment and what the driving signals are. To obtain a single bias measure for an individual fixation, we compute its horizontal distance from the vertical midline. We assign the distance a positive sign if the fixation falls on the high-contrast side and a negative sign if it falls in the low-contrast side. Figure 4A shows this measure of fixation bias as function of trial number (excluding the 32 trials/block, in which no gradient was present) for the second fixation of each subject. As known from the aforementioned analysis and Figure 3, fixations are, on average, biased towards the high-contrast side during the free-viewing blocks (trials 1–64 and 193–256), have no bias for the unbiased search block (trials 65–128), and are biased towards the low-contrast side during inverse bias search (trials 129–192). In the light of inter-subject and trial-to-trial variability, we analyze the mean bias (black circles in Figure 4) and average it across a sliding window of width 8 trials (solid black line in Figure 4A). For the 2nd fixation depicted in Figure 4A, learning is so rapid that we cannot identify a clear monotonic decrease at the beginning of each block. Given the averaging window, this implies that learning happens within 4 trials or less, i.e., very rapidly. Does this only apply to early fixations within a trial? Figure 4B depicts the average traces (akin to the black line in Figure 4A) for all fixations. In all fixations, the biases are learnt and unlearnt very rapidly. The slowest effect is found for the unlearning of the free-viewing bias for later fixations, which takes about 10 trials. It is tempting to speculate that in initial trials of the unbiased search condition, observers are “reset” to a bottom-up mode, if they fail to find the target within the first few fixations. As soon as observers have learnt that the target appears with equal probability

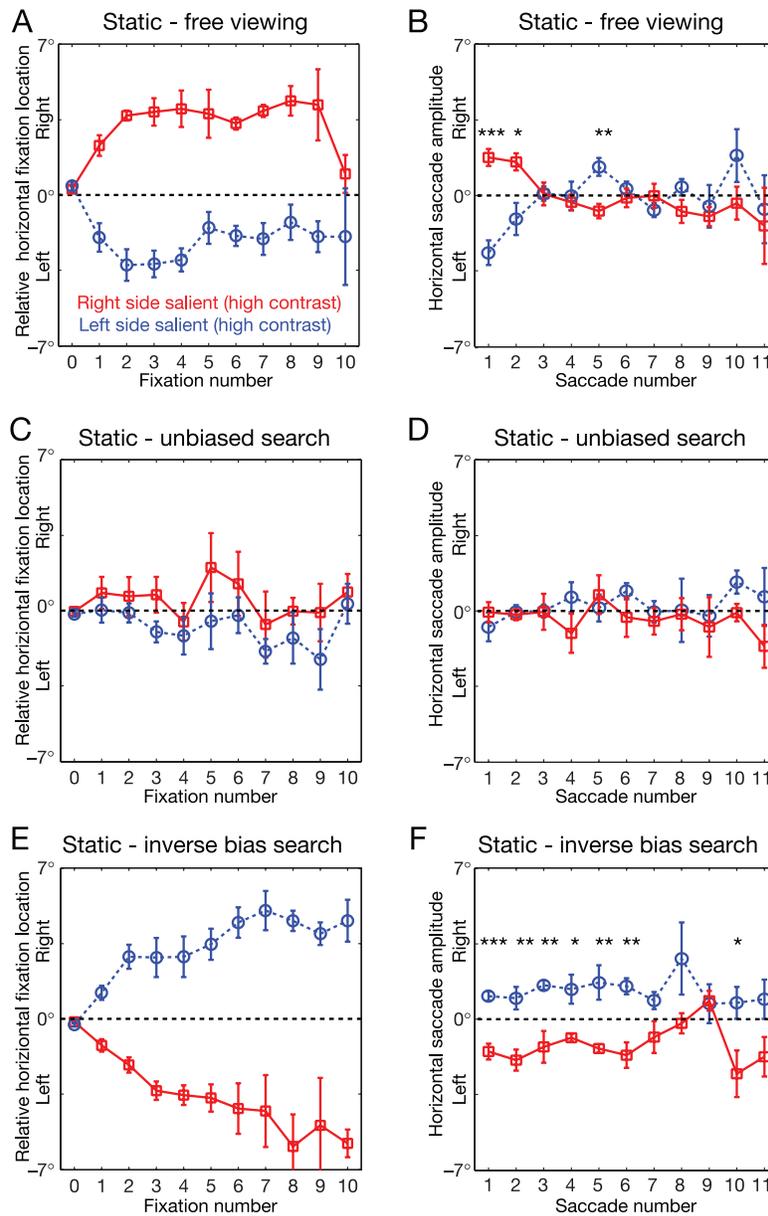


Figure 3. Fixation bias as function of number of fixations in Experiment 1. (A, C, E) Red solid lines: mean horizontal fixation location in images with “salient” (high-contrast) side on the right; blue dotted lines: high-contrast side to the left. All data relative to mean fixation location in images without contrast gradient of the same subject and task. Error bars denote SEM across subjects. Fixation 0 denotes initial central fixation (triggering stimulus onset). (B, D, F) Saccade direction between successive fixations (saccade 1: fixation 0 to fixation 1, etc.). Markers denote gradient direction as in panels A, C, E. Significance markers denote results of post hoc *t*-tests results. Sensory-driven saliency that is effective at biasing eye-movements in free viewing (A, B) can be rendered ineffective (C, D) or even be countermanded (E, F) when searching for a target hidden in the image. In panels B, D and F significance markers denote results of (uncorrected) *t*-tests. Due to the (trivial) mutual dependence of successive fixations’ location, no significance markers are provided in panels A, C, E.

on either side, top-down dominance will persist throughout, even if the target is not found early in the trial. In any case, learning and unlearning to countermand the bottom-up bias is not only implicit, but also rapid, within less than 10 trials.

What is the driving force behind this rapid learning? Although this study is not designed to answer this question quantitatively, we qualitatively inspected the

first trials after a change of task. In the first unbiased search trials, subjects typically retain some tendency to fixate the high-contrast side. After they for the first time find the target in the low contrast region, they start searching there in subsequent trials. After having found the target occasionally in the low-contrast region, the initial bias has entirely vanished. A similar pattern is observed for inverse bias search. As soon as a target is

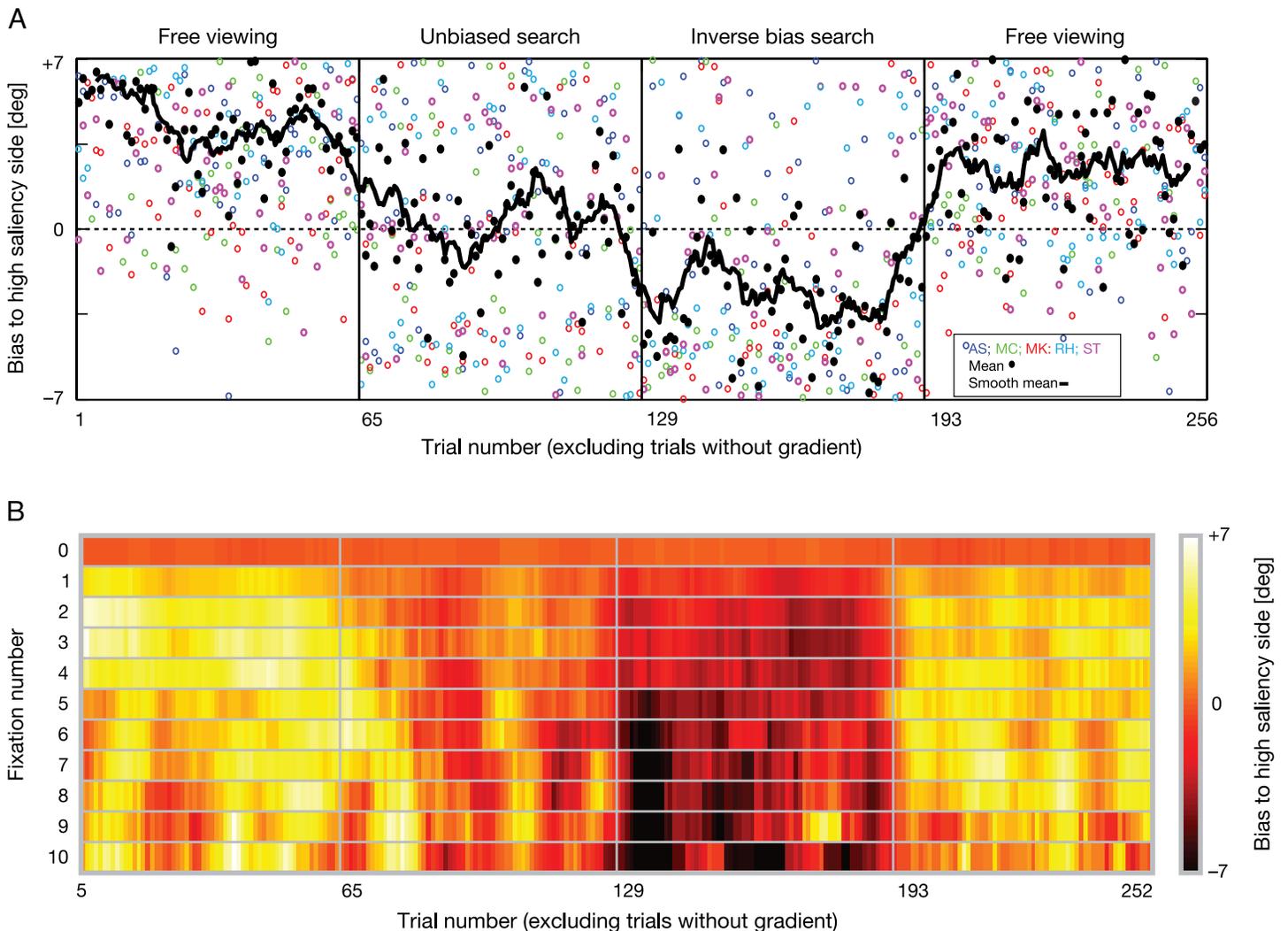


Figure 4. Speed of implicit learning. (A) Fixation bias towards the high-contrast side for the second fixation of each trial and individual subject (open circles). Black filled circles denote mean across subject for each trial, solid black line a sliding window average. Trials without gradient have been excluded from the time axis. Note that the averaging window is symmetric, i.e., non-causal and reaches 4 data points back in time. (B) Color-code depicts the bias towards the high-contrast side averaged over subjects and adjacent trials for each fixation number. (The line for the second fixation, i.e., the third line from the top, corresponds to the black trace of panel A.)

once found quickly in the low-contrast region, observers keep this successful strategy and preferentially start searches there. Longer stretches (about 3 trials) of stimuli without gradient, however, can make the stimulus-driven bias partly reoccur. The qualitative trial-by-trial inspection of the data thus supports the quantification across all trials: learning (and unlearning) of biases happens very rapidly.

Experiment 2: Does quick countermanding result from too low saliency?

Could the result of [Experiment 1](#), that saliency can be quickly overridden and countermanded, merely be explained by the assumption that contrast gradients were

not salient enough? To test this possibility, we performed the same experiment, but replaced the static gradient by a dynamic bias on saliency, a 5-Hz flicker in either the left or right side of the stimulus. The first fixation shows a significant bias towards the flickering side ($p = 7 \times 10^{-5}$, [Figure 5A](#)), and consequently the first saccade is significantly biased in this direction ($p = 0.0002$, [Figure 5B](#)). However, unlike for the static bias, there is no further drive towards the high-saliency side for the second saccade. To the contrary, the bias is significant towards the low-salient side for saccades 3 and 4 ($p = 0.03, 0.02$) respectively. Although it is unclear whether this happens because prolonged looking at flicker is repulsive or because our flicker is less salient per se, we achieve the desired effect: free-viewing saliency, which we operationally define as the attractiveness of the high-saliency side, is weaker than in the static condition.

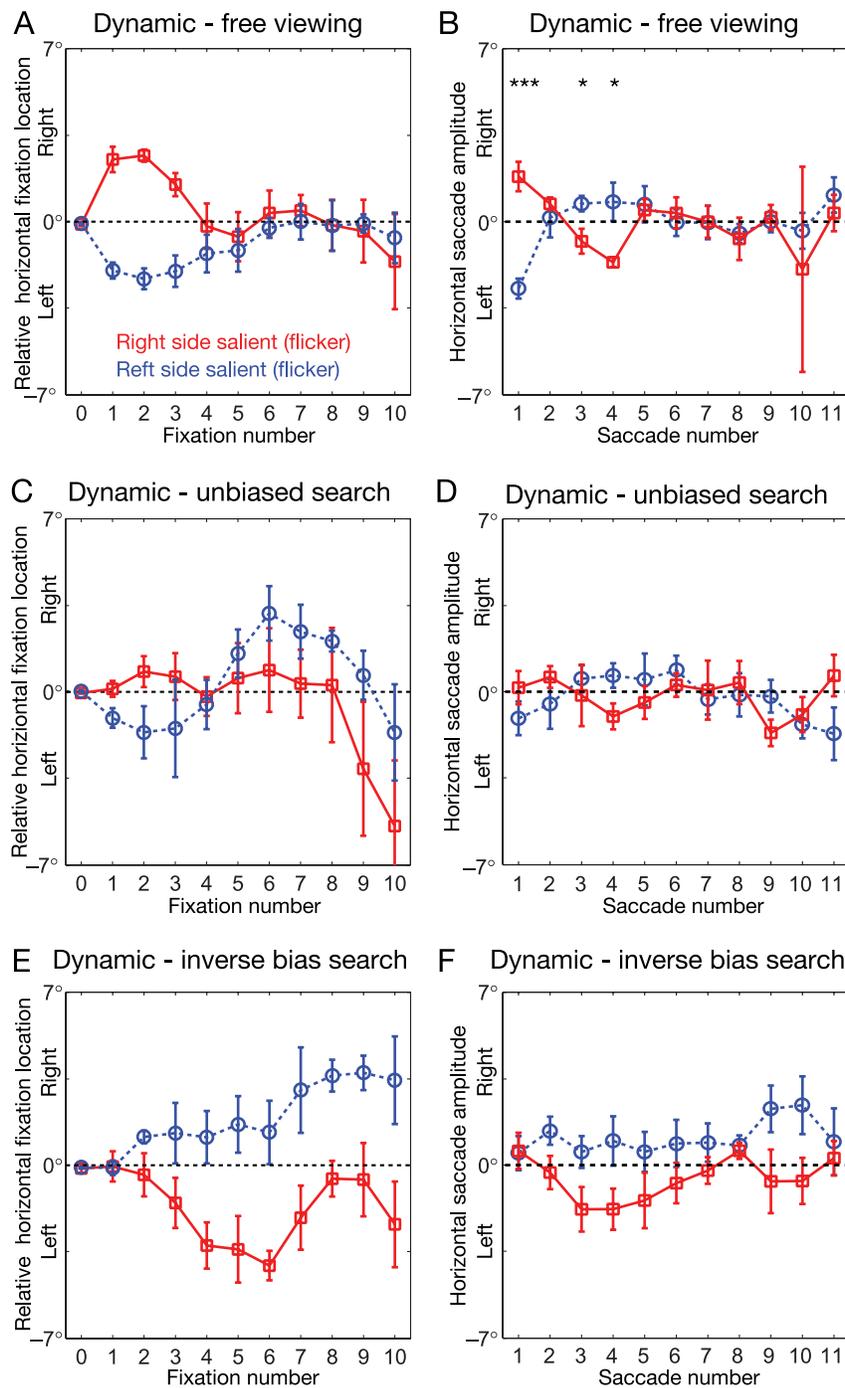


Figure 5. Fixation bias in dynamic condition (Experiment 2). Red solid lines: mean horizontal fixation location in images with “salient” (flickering) side on the right; blue dotted lines: flicker to the left. Notation as in Figure 3. (A, B) free viewing, (C, D) unbiased search, (E, F) inverse bias search.

Hence, if the overriding and countermanding of the static gradient were only due to its too weak saliency, one would predict that the dynamic bias is at least as easily and quickly countermanded. Contrary to this prediction, the bias towards the flickering side has a slight tendency to prevail for the first fixation in unbiased search, although this bias fails to reach significance for the position alone ($p = 0.051$,

Figure 5C) and there is no significant bias to any saccade ($p > 0.05$, Figure 5D). More importantly, there are no saccades in the inverted bias search that are significantly biased towards the low-contrast side ($p > 0.06$ for any saccade), although saccades 2 through 7 show a trend towards inversion (Figure 5F). Also, the inversion starts later than in the static condition, as there is no bias for the first two fixations ($p = 0.94$, Figure 5E).

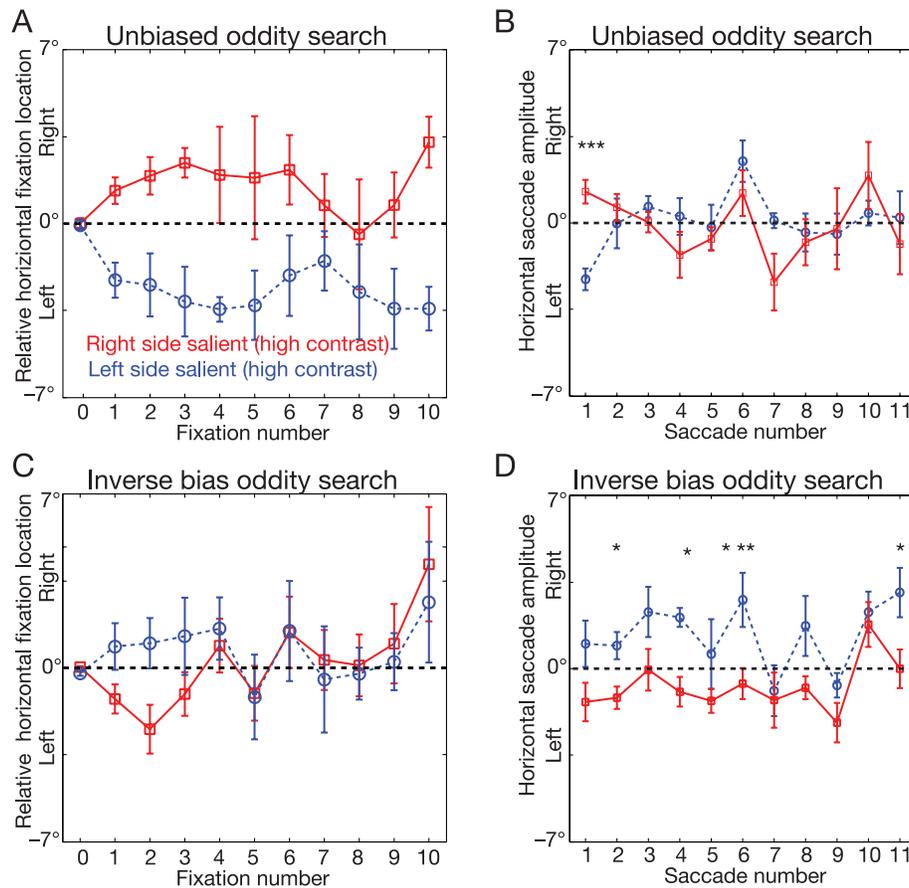


Figure 6. Oddity search (Experiment 3) Red solid lines: mean horizontal fixation location in images with “salient” (high-contrast) side on the right; blue dotted lines: high-contrast side to the left. Notation as in Figure 3. (A, B) Unbiased oddity search, (C, D) inverse bias oddity search.

Consequently, despite its lower saliency in free-viewing, the dynamic bias takes more time to be countermanded by a search task. This implies that insufficient saliency cannot explain why the search task countermands a sensory-driven bias.

Experiment 3—Oddity search

In Experiment 3, observers in some experimental blocks searched for spots of locally increased contrast (“oddity search”). In other blocks we replicated the template (bull’s-eye) search of Experiment 1. Consistent with Experiment 1, the first free-viewing fixation is biased to the high-contrast side ($p = 5 \times 10^{-6}$, t -test) and so is the first saccade ($p = 7 \times 10^{-6}$, t -test). Similarly, none of the unbiased template search fixations show a significant bias ($p > 0.31$ for all fixations) in line with Experiment 1. In inverse bias template search, countermanding occurs immediately (first fixation: $p = 0.01$). Saccades 1 through 11 are biased towards the low contrast side, and for 1 through 4 and 6 through 8, the bias is significant at $p < 0.05$ (data not shown).

In contrast to template search, the bias towards high-contrast prevails for the first fixation in unbiased oddity

search ($p = 0.004$, Figure 6A) and also for the first saccade ($p = 0.0006$, Figure 6B). Analogously, there is only little countermanding of the bottom-up bias in inverse bias oddity search (Figure 6C), and only few saccades show a bias to the low-contrast region (Figure 6D). These data suggest that at least some of the sensory-driven bias prevails in oddity search, whereas template search entirely overrides it. Hence, the extent to which search can override sensory-driven saliency depends on the search type and on the feature defining the target.

Comparison across experiments

So far we have analyzed all experiments and conditions separately. Here we directly compare the same conditions across different experiments. To obtain a single measure of bias, we subtract the mean fixation location for left-side-salient trials from the mean for right-side-salient trials and average across subjects. This differential measure is positive if fixations are biased to the high-saliency side and negative if fixations are biased to the low-saliency side. We analyze eye-position separately for all fixations. It is important to note that consecutive

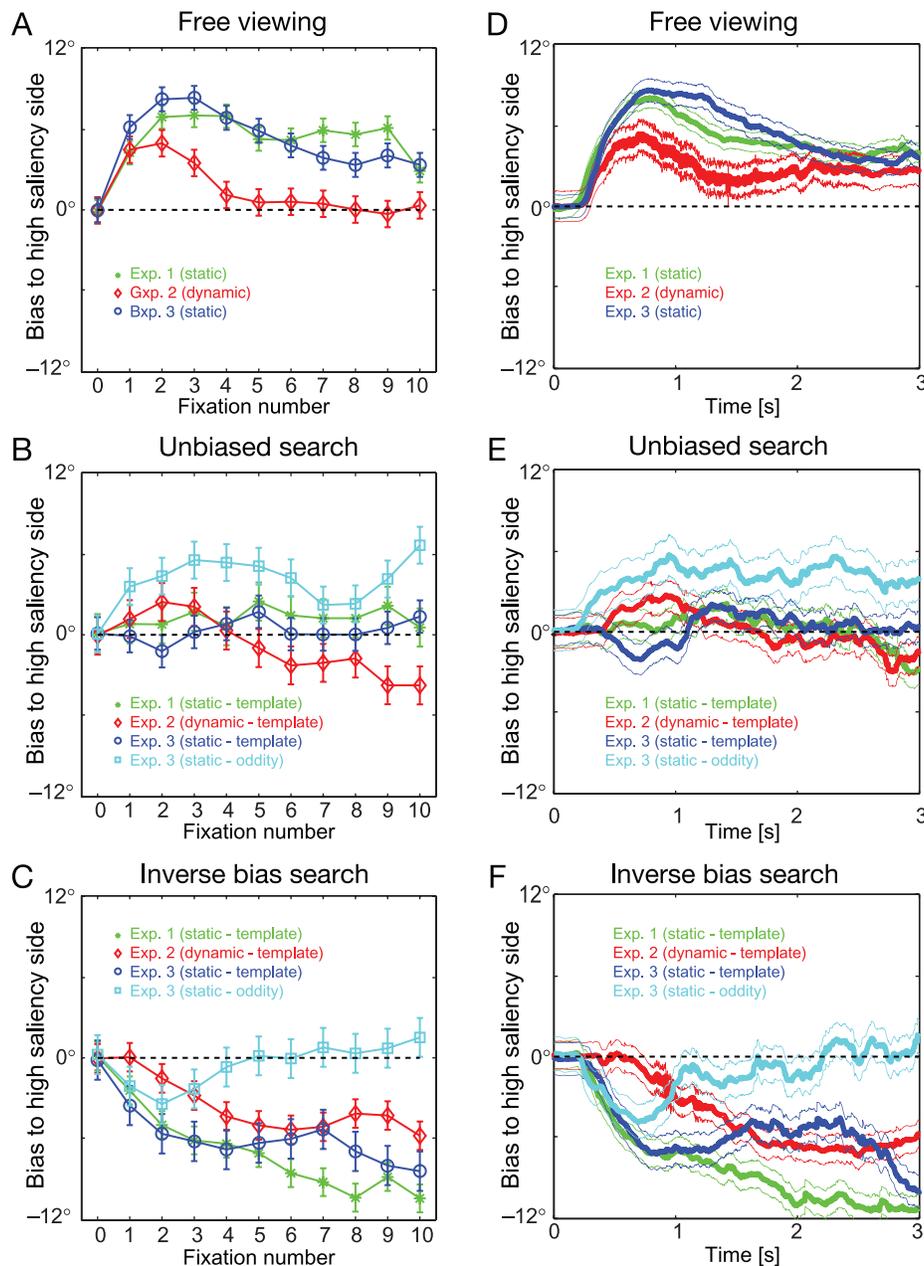


Figure 7. Summary of Experiments 1–3. (A–C) Difference in relative fixation location (right-side-salient minus left-side-salient) plotted against fixation number. Note that this measure of fixation bias is quantitatively different from the bias measure of Figure 4, which measures the distance of individual fixations from the midline. (D–F) Difference in relative eye-position plotted against time elapsed in trial (mean \pm SEM at each time point). (A, D) Free-viewing, (B, E) unbiased search; (C, F) inverted bias search. (A–F) Green: Experiment 1 (static gradient), red: Experiment 2 (dynamic gradient); blue and cyan: Experiment 3 (static gradient). In panels B and C, blue denotes template search (replication of Experiment 1) and cyan denotes oddity search.

fixations are not independent from each other, i.e., the onset and offsets of differences are of particular interest. In free viewing (Figure 7A), the two experiments with static bias (Experiments 1 and 3) are not different for any fixation (minimum over all fixations: $p_{\min} = 0.07$, t -test). The bias induced by the static gradients in Experiments 1 and 3 is larger than the bias induced by the flicker

(Experiment 2) from the first or second fixation onward, respectively. Comparing individual fixations between Experiments 1 and 2, the difference reaches significance at for fixations 2 through 9 ($p = 0.03, 0.02, 0.007, 0.03, 0.005, 0.002, 0.03$, respectively). Comparing individual fixations between Experiments 2 and 3, the difference reaches significance for fixations 4 through 7 ($p = 0.03$,

0.02, 0.008, 0.004). This shows that the bias induced in free-viewing is stronger for the static gradient than for the dynamic flicker not only across all fixations, but also at some individual time points. This does not imply that flicker is generally less salient than contrast but shows that for the specific parameters chosen here, contrast gradients induce a more robust sensory-driven bias on fixation than flicker. In this operational sense, our contrast gradient acts as a more salient cue under free-viewing conditions than the particular instance of flicker.

Next we directly compare all unbiased search experiments (Figure 7B). As expected, we do not find any difference for any fixation between template search in Experiment 1 and Experiment 3 ($p_{\min} = 0.28$). Similarly, the direct comparison yields no evidence for a difference ($p_{\min} = 0.12$) between template search in Experiment 1 (static) and Experiment 2 (dynamic), and only a slight trend to a difference between template search in Experiments 2 and 3 ($p = 0.046$ at 2nd fixation). Hence, unbiased template search is similarly unbiased across all experiments.

For oddity search, however, the bias towards the high-contrast side is consistently larger than for template search in any condition (Figure 7B). For individual fixations, the difference of oddity search to template search is significant for fixation 10 compared to Experiment 1 ($p = 0.04$), and for fixations 2, 5, 6, 7, 9, 10 compared to Experiment 2 ($p = 0.03, 0.005, 0.005, 0.04, 0.046, 0.001$, respectively). The more relevant comparison within the same experiment yields significant differences for fixations 1 through 4 ($p = 0.03, 0.04, 0.03, 0.049$, respectively). This indicates that—even for individual fixations but most importantly across all fixations—the free-viewing bias partly prevails for oddity search. Direct comparison between free-viewing bias and oddity search bias in Experiment 3 indeed provides no evidence for a significant difference between the two for any individual fixation ($p_{\min} = 0.053$). In contrast, the bias in free-viewing is significantly larger than in unbiased template search for fixations 1, 2, 3, 4, 6, 7, 9 ($p = 0.001, 0.003, 0.006, 0.01, 0.01, 0.03, 0.04$). Especially for early fixations, this is clear evidence that unbiased template search overrides the sensory-driven bias, while oddity search has no or less of such an overriding effect.

For countermanding in inverse bias search (Figure 7C), we again find no difference for template search between Experiments 1 and 3 for any fixation ($p_{\min} = 0.08$). Hence, template search in Experiment 3 replicates the findings of Experiment 1 entirely, ensuring that the order of blocks does not have a major effect on outcome. In template search, less countermanding takes place for the dynamic bias (Experiment 2) than for the static bias (Experiments 1 and 3) in all fixations. This difference is significant also for individual fixations. Between Experiments 1 and 2, fixations 1, 2, 3, 8, and 10 reach significance individually (at $p = 0.005, 0.02, 0.007, 0.02, \text{ and } 0.03$ respectively), between Experiments 2 and 3, significance is restricted to

fixations 1 and 2 ($p = 0.01, 0.01$). This confirms the earlier analysis: flicker is harder to countermand than static gradients, especially for early fixations. Comparing oddity and template search reveals that countermanding a static bias is weaker for all fixations in oddity search than for template search. Comparing Experiment 1 template search with Experiment 3 oddity search reveals significant differences for fixations 3 through 10 ($p = 0.04, 0.02, 0.004, 0.002, 0.001, 0.003, 0.003, 0.001$) and for fixations 4 through 9 for the comparison within Experiment 3 ($p = 0.01, 0.02, 0.03, 0.03, 0.007$). Consequently oddity search contributes less to countermanding a sensory-bias than template search, especially for the later phase of search.

Comparison in absolute time

The analysis so far used fixations to define discrete time steps. Any systematic relation between task and dwell time could affect the relative timing in this representation. Hence, we reanalyze the data using the raw eye-positions irrespective of the type of eye-movement. Qualitatively the results are very similar: For free viewing (Figure 7D), the dynamic bias kicks in at about the same time as the static biases. Nevertheless, the bias induced by flicker rebounds pretty quickly, reaching its peak about 100 ms earlier than the static situation (697 ms compared to 789 ms for Experiment 1 and 803 ms for Experiment 3) and then quickly falling back to nearly no bias. In unbiased template search, there is no consistent bias for either static or dynamic conditions. However, for oddity search, some of the free-viewing bias prevails (Figure 7E). Biases in all static inverse search conditions (template or oddity) have the same initial time course (Figure 7F), while countermanding the flicker takes more time. After about three quarters of a second, countermanding by inverse bias oddity search starts to relax, while at the same time countermanding of the dynamic cue takes effect. In all, the analysis over time is qualitatively very similar to the results obtained on individual fixations. The precise definition of fixations and saccades is therefore uncritical to the results presented here.

Experiment 4—Target visibility

The data of Experiment 3, in particular the direct comparisons of Figures 7B and 7E, suggest that oddity search has less capability to override a bottom-up bias. Is this a peculiarity of the oddity search which is defined by the same feature as the bottom-up bias, or is it a result of reduced visibility of the oddity target compared to the template target?

In Experiment 4, subjects reported the presence or absence of a target as quickly as possible while maintaining fixation at the center. The target could occur at one of two fixed locations. Subjects maintained fixation in 94.1%

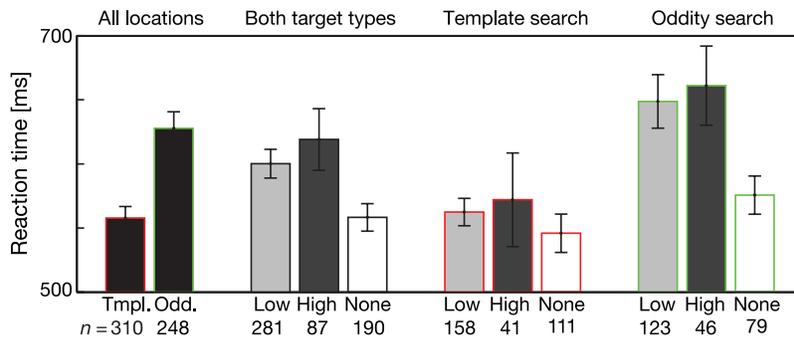


Figure 8. Visibility. Reaction time for all 557 hit (correct and target present) trials (pooled over 4 subjects), mean \pm SEM for different conditions: all template targets, all oddity targets, all targets in low contrast, all in high-contrast, no gradient; template targets in low contrast/high contrast/no gradient, oddity target in low contrast/high contrast/no gradient. Note that y-axis starts at 500 ms. Red outlines denote template, green outlines oddity search, black outline pooling over both. Shading of bars identifies target location (light gray: low contrast; dark gray: high contrast; white: no gradient; black: all locations).

$\pm 4.7\%$ (mean \pm SD over subjects) of all trials, to which further analysis is restricted. In $78.5\% \pm 7.9\%$ of target-present trials, the target was correctly found. The absence of a target was correctly reported in $93.2\% \pm 2.0\%$ of target-absent trials. In both cases, performance in template search ($85.1 \pm 8.5\%$ and $95.4\% \pm 2.1\%$) was better than in oddity search ($71.1\% \pm 13.8\%$ and $91.0\% \pm 5.7\%$, Figure 8). This is the first indication that template targets are more visible than oddity targets. To further analyze this difference, we pooled the hit trials of all observers and performed a two-factor ANOVA on the reaction times with factors search type (oddity vs. template) and target location relative to gradient (target in high-contrast, low contrast, or no gradient). We found significant main effects for target type ($F[1, 552] = 16.02$; $p = 0.0001$) and target location ($F[2, 552] = 4.44$; $p = 0.01$), but no significant interaction ($F[2, 552] = 1.57$; $p = 0.21$). Template targets were found significantly quicker than oddity targets (558 ± 159 ms compared to 628 ms ± 203 ms). A Tukey–Kramer test shows that the significant effect of target location is attributable to the difference between stimuli without gradient (558 ms ± 148 ms) to the stimuli with gradient, whereas there is no significant (at $\alpha = 0.05$) marginal difference between targets in high (619 ms ± 224 ms) or low (600 ms ± 188 ms) contrast regions. These results show that template targets are more visible (more easily found) than oddity targets. Search is easier if there is no gradient, but there is no clear indication of a generally better visibility of targets in the low-contrast region.

Discussion

In the present study, we demonstrate that a visual search task can override and actively countermand sensory-

driven saliency in naturalistic visual stimuli. Such overriding can be almost immediate, i.e., within one fixation, and the overriding speed is not related to the stimulus feature's free-viewing saliency. Overriding is, however, dependent on the feature itself (as revealed by the difference of flicker and contrast gradients) and on the type of search (oddity versus template). Hence, top-down signals modulate bottom-up saliency in a feature-specific way, which constrains models of their interaction.

To bias sensory-driven saliency, we use contrast gradients and flicker. Many studies have found a correlation between contrast and fixation in free-viewing conditions (Einhäuser & König, 2003; Mannan, Ruddock, & Wooding, 1996, 1997; Parkhurst et al., 2002; Peters et al. 2005; Reinagel & Zador, 1999; Tatler et al., 2005), and contrast is also one of the features (besides color and orientation) in the original saliency map model of attention (Itti & Koch, 2000; Koch & Ullman, 1985). Although this does not imply a causal effect of contrast on fixation for local deployment of attention (Einhäuser & König, 2003), the effect of contrast gradients is compatible with a linear relation between large-scale fixation biases and contrast (Einhäuser et al., 2006b). As for contrast, Itti (2006) demonstrated a relation of flicker to fixation in free-viewing natural scenes. The fact that flicker here induces a weaker bias than the contrast gradient in free viewing does not imply that flicker in general is less salient than contrast. Rather our specific choice of flicker parameters (5 Hz, half the contrast range) makes its free-viewing bias weaker by design to show that low saliency cannot account for quick overriding. Similarly, we do not argue that oddity search and template search are inherently of different difficulty. Rather, for our choice of targets, the oddity target may induce a weaker top-down signal, potentially related to its lower visibility at fixed locations.

For free viewing, there is considerable debate as to whether or not the strength of bottom-up features in

driving human gaze in natural scenes decreases over time. Parkhurst et al. (2002) argue that stimulus-driven saliency is higher for early fixations and drops during prolonged viewing to an above-chance asymptotic level. Tatler et al. (2005) challenge this result by demonstrating that it is likely to be a consequence of a common “central bias” of stimuli and fixation, which lets the early central fixations exhibit high saliency. Our data indicates that the bottom-up bias is deployed quickly, although the first fixation is prone to a central bias (the trial is started by a central fixation). We see no evidence that top-down signals take more than the time to the first fixation to deploy (overriding and countermanding is immediate for the static condition). Nevertheless, even in free-viewing, the bottom-up drive towards the high-saliency side ends after a few fixations. Since the plateau is reached less than half way to the image boundary, it seems unlikely that this is only due to boundary effects alone. To the contrary, there is a slight trend back towards the low contrast side and a drive away from the flicker after 3 to 5 fixations. One interpretation of our data therefore suggests that in an initial phase bottom-up and top-down signals drive attention; whereas the effect of bottom-up signals decreases, the effect of top-down signals remains until the task is accomplished. The precise relative time course, however, depends on the bottom-up feature (not on its free-viewing saliency) and on the type of target.

Several studies measure the effect of task on eye-movements. Examples include every-day activities like tea making (Land, Mennie, & Rusted, 1999), judgment of social status in a photograph (Yarbus, 1967), rating preference for a picture (Buswell, 1935), or rating facial attractiveness (Shimojo, Simion, Shimojo, & Scheier, 2003). Notwithstanding the realism of such tasks and the potential role of eye-movements as indicator of cognitive disorders such as autism (Pelphrey et al., 2002), quantification of task related effects typically rely on variants of visual search. The usage of visual search in eye-movement research dates back at least to Buswell (1935), who instructed an observer to “find a person looking out of one of the windows of the tower,” causing remarkable context-driven effect on eye-movement patterns. Most research on visual search, however, is based on reaction time measurements in well-controlled search displays. Following the pioneering work of Treisman and Gelade (1980), the dependence of reaction times versus set size serves as measure of whether a search is serial (reaction time increases linearly with set-size) or parallel (reaction time is independent of set size). The slope of this relation determines the difficulty of the search. Using such a setting, more and more sophisticated models have been put forward to explain the relation of target features to those of attended items. Probably the most influential is “Guided Search” (Wolfe, 1994; Wolfe et al., 1989): Akin to the saliency map, the stimulus is filtered in various feature channels, top-down attention enhances the channels present in the target, resulting in an “activation map”

guiding search. If the target is different in just an individual elementary feature, the activation map has one peak, the target pops out, and search is parallel; if the target is defined by the conjunction of multiple features, the activation map has multiple peaks that have to be searched sequentially. Hence, Guided Search provides a mechanism accounting for Treisman and Gelade’s (1980) data and a framework for integrating bottom-up saliency and top-down feature biases.

The relative weighing of top-down and bottom-up signals for search presents a critical issue for modeling attention. The top-down weights of a feature can be scaled according to its reliability and task relevance (Bacon & Egeth, 1997). However, to optimize search performance, it is not necessarily optimal to boost the target’s features themselves, but rather the signal-to-noise ratio of the target relative to the distractors should be maximized (Navalpakkam & Itti, 2006). This requires knowledge about the distractor distribution and therefore benefits from the integration of bottom-up information. To test whether bottom-up and top-down signals operate independently in a simple search task, van Zoest and Donk (2004) manipulate the saliency of their target (a bar of defined orientation) and a distractor (a bar of another orientation) relative to a background (a grid of equally oriented bars) as well as the relative similarity of target and distractor (their orientation difference). In this setting, they find that bottom-up signals and top-down signals operate independently. The same authors further argue that top-down and bottom-up processes operate on different time scales: early deployment of attention is entirely bottom-up driven, while top-down signals only affect search later (van Zoest, Donk, & Theeuwes, 2004). This finding is consistent with the time course of neuronal activity in area V4, which in early stages preferentially signal stimulus identity, while later stages signal its task-relevance (Ogawa & Komatsu, 2006). Our current results do not rule out that the deployment of top-down signals takes time, as all we can state is their presence from the first fixation, i.e., after about 200 ms to 300 ms. While the time to deployment of top-down signals does not seem to depend on bottom-up saliency *per se*, it nevertheless depends on the bottom-up feature that renders a location salient.

We demonstrate that top-down guidance of visual search does not make the observer oblivious to bottom-up features, but the effect of the features’ saliency is quickly adapted to task-demands. This suggests that the neuronal substrate for weighing bottom-up features occurs late in the visual processing hierarchy. In line with this interpretation, Ipata et al. (2006) find that responses in macaque LIP are reduced for stimuli that have high bottom-up saliency (pop-out) but have to be ignored for the task. It is, however, well conceivable that such task-dependent signals modulate late responses in visual areas, which are sensory-driven early during processing and receive convergent top-down and bottom-up projections (Mazer & Gallant, 2003; Ogawa & Komatsu, 2006).

To model overt visual search in natural stimuli, Rao et al. (2002) replace the features of Guided Search and similar models by general, task-dependent filters. Search starts on a coarse spatial scale and weighing of features is refined at each fixation until the target is found or the finest scale has been reached. This model successfully captures eye-movement and error patterns in overt visual search in complex stimuli. Similarly, when observers search for image patches in natural scenes, target features are elevated at fixated non-target locations (Pomplun, 2006; Rajashekar, Bovik, & Cormack, 2006). For natural scenes, this paradigm, however, does not dissociate between top-down and bottom-up signals: Scene context can bias search (Neider & Zelinsky, 2006; Oliva et al., 2003; Torralba, Oliva, Castelhano, & Henderson, 2006), and thus patches of similar bottom-up structure may be preferentially fixated as consequence of such a contextual, top-down bias. We therefore chose targets unrelated to scene context and nevertheless find a strong task-dependent effect as well as (implicit) learning of target location. This suggests that stimulus features can be used to learn scene context for search, even though their bottom-up saliency is actively ignored.

The finding that visual search can override bottom-up saliency in natural scenes is in line with recent results (Henderson et al., 2007; Rutishauser & Koch, 2007; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006). In addition, we demonstrate that this overriding is not a consequence of insufficient bottom-up saliency. The difference between oddity and template search furthermore suggests that the strength of the top-down signal, quantified by target visibility, may also modulate the degree of overriding. Our data imply that task-specific modulation of individual bottom-up features, as used for covert visual search in simple stimuli, accounts for overt visual search in natural stimuli. Combining such feature-specific top-down signals with (learned) contextual priors on target location (Najemnik & Geisler, 2005; Neider & Zelinsky, 2006; Oliva et al., 2003; Torralba et al., 2006) therefore may provide a promising approach to searching for real-world objects in their natural context.

Acknowledgments

This work was financially supported by the Swiss National Science Foundation (WE, PA00A-111447), the National Institutes of Health USA, the National Science Foundation, and by the DARPA/NGA. The authors thank F. Moradi for discussion and J. Wolfe for comments on earlier versions of the manuscript.

Commercial relationships: none.
Corresponding author: Wolfgang Einhäuser.
Email: wolfgang.einhaeuser@inf.ethz.ch.

Address: Institute of Computational Science, CAB G 82.2, Universitätstrasse 6, ETH Zentrum, 8092 Zurich, Switzerland.

Footnote

¹Strictly speaking, saccade directions are not expected to be *unconditionally* independent, since common factors exist. However, *given the experimental setting*, the direction of saccade n and $n + 1$ are expected to be *conditionally* independent (notwithstanding effects like IOR and inertia). This does not mean that saccade direction is independent of saccade number. It does mean, however, that once the experimental parameters (task/gradient/target/saccade number/viewing time/etc) are known, the direction of saccade n carries no additional information on the direction of saccade $n + 1$. Obviously, this property does not hold for fixated *locations*, since distances between successive locations are drawn from the saccade amplitude distribution, which has finite moments.

References

- Armstrong, K. M., Fitzgerald, J. K., & Moore, T. (2006). Changes in visual receptive fields with microstimulation of frontal cortex. *Neuron*, *50*, 791–798. [[PubMed](#)] [[Article](#)]
- 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](#)]
- 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](#)]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. [[PubMed](#)]
- Buswell, G. T. (1935). *How people look at pictures. A study of the psychology of perception in art*. Chicago: The University of Chicago Press.
- Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Reviews of Neuroscience*, *22*, 319–349. [[PubMed](#)]
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*, 613–617. [[PubMed](#)]

- Dickinson, S., Christensen, H., Tsotsos, J., & Olofsson, G. (1997). Active object recognition integrating attention and viewpoint control. *Computer Vision and Image Understanding*, *63*, 239–260.
- Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience* *17*, 1089–1097. [[PubMed](#)]
- Einhäuser, W., Kruse, W., Hoffmann, K.-P., & König, P. (2006a). Differences of monkey and human overt attention under natural conditions. *Vision Research*, *46*, 1194–1209. [[PubMed](#)]
- Einhäuser, W., Rutishauser, U., Frady, E. P., Nadler, S., König, P., & Koch, C. (2006b). The relation of phase-noise and luminance-contrast to overt attention in complex visual stimuli. *Journal of Vision*, *6*(11):1148–1158, <http://journalofvision.org/6/11/1/>, doi:10.1167/6.11.1. [[PubMed](#)] [[Article](#)]
- Henderson, J. M., Brockmole, J. R., Castelano, M. S., & Mack, M. (2007). Visual saliency does not account for eye-movements during visual search in real-world scenes. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movement research: Insights into mind and brain* (pp. 537–562). Oxford: Elsevier.
- Ipata, A. E., Gee, A. L., Gottlieb, J., Bisley, J. W., & Goldberg, M. E. (2006). LIP responses to a popout stimulus are reduced if it is overtly ignored. *Nature Neuroscience*, *9*, 1071–1076. [[PubMed](#)]
- Itti, L. (2006). Quantitative modeling of perceptual saliency at human eye position. *Visual Cognition*, *14*, 959–984.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506. [[PubMed](#)]
- James, W. (1890). *Principles of Psychology*. New York: Holt.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227. [[PubMed](#)]
- Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*, 3559–3565. [[PubMed](#)]
- Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*, 1311–1328.
- Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*, 165–188. [[PubMed](#)]
- Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision*, *11*, 157–178. [[PubMed](#)]
- Mazer, J. A., & Gallant, J. L. (2003). Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. *Neuron*, *40*, 1241–1250. [[PubMed](#)] [[Article](#)]
- Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*, 387–391. [[PubMed](#)]
- Navalpakkam, V., & Itti, L. (2006). Optimal cue selection strategy. *Advances in Neural Information Processing Systems*, *19*, 1–8.
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*, 614–621. [[PubMed](#)]
- Ogawa, T., & Komatsu, H. (2006). Neuronal dynamics of bottom-up and top-down processes in area V4 of macaque monkeys performing a visual search. *Experimental Brain Research*, *173*, 1–13. [[PubMed](#)]
- Oliva, A., Torralba, A., Castelano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. *IEEE Proceedings of the International Conference on Image Processing*, *1*, 253–256.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of saliency in the allocation of overt visual attention. *Vision Research*, *42*, 107–123. [[PubMed](#)]
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. [[PubMed](#)]
- Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders*, *32*, 249–261. [[PubMed](#)]
- Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*, 2397–2416. [[PubMed](#)]
- Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision Research*, *46*, 1886–1900. [[PubMed](#)]
- Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2006). Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision*, *6*(4):7, 379–386, <http://journalofvision.org/6/4/7/>, doi:10.1167/6.4.7. [[PubMed](#)] [[Article](#)]
- Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*, 1447–1463. [[PubMed](#)]

- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network*, *10*, 341–350. [[PubMed](#)]
- Rizzolatti, G., Riggio, L., Dascola, I., & Umiltà, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*, 31–40. [[PubMed](#)]
- Robinson, D. L., & Petersen, S. E. (1992). The pulvinar and visual saliency. *Trends in Neurosciences*, *15*, 127–132. [[PubMed](#)]
- Rutishauser, U., & Koch, C. (2007). Probabilistic modeling of eye movement data during conjunction search via feature-based attention. *Journal of Vision*, *7*(6):5, 1–20, <http://journalofvision.org/7/6/5/>, doi:10.1167/7.6.5. [[PubMed](#)] [[Article](#)]
- Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*, 1317–1322. [[PubMed](#)]
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, *45*, 643–659. [[PubMed](#)]
- Thompson, K. G., & Bichot, N. P. (2005). A visual saliency map in the primate frontal eye field. *Progress in Brain Research*, *147*, 251–262. [[PubMed](#)]
- Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*, 766–786. [[PubMed](#)]
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136. [[PubMed](#)]
- Treue, S. (2003). Visual attention: The where, what, how and why of saliency. *Current Opinion in Neurobiology*, *13*, 428–432. [[PubMed](#)]
- Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology*, *18*, 321–343.
- van Zoest, W., & Donk, M. (2004). Bottom-up and top-down control in visual search. *Perception*, *33*, 927–937. [[PubMed](#)]
- van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 746–759. [[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., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 419–433. [[PubMed](#)]
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.