Interaction between Bottom-up Saliency and Top-down Control: How Saliency Maps Are Created in the Human Brain

Lucia Melloni\textsuperscript{1,2,3}, Sara van Leeuwen\textsuperscript{1,2,3}, Arjen Alink\textsuperscript{1,2,5} and Notger G. Müller\textsuperscript{2,5,4}

\textsuperscript{1}Department of Neurophysiology, Max-Planck Institute for Brain Research, 60528 Frankfurt am Main, Germany, \textsuperscript{2}Brain Imaging Center, Goethe University, 60528 Frankfurt am Main, Germany, \textsuperscript{3}Cognitive Neurology Unit, Goethe University, 60528 Frankfurt am Main, Germany and \textsuperscript{4}Department of Neurology, Otto-von-Guericke University, 39120 Magdeburg, Germany

The Author 2012. Published by Oxford University Press. All rights reserved.

Whether an object captures our attention depends on its bottom-up salience, that is, how different it is compared with its neighbors, and top-down control, that is, our current inner goals. At which neuronal stage they interact to guide behavior is still unknown. In a functional magnetic resonance imaging study, we found evidence for a hierarchy of saliency maps in human early visual cortex (V1 to hV4) and identified where bottom-up saliency interacts with top-down control: V1 represented pure bottom-up signals, V2 was only responsive to top-down modulations, and in hV4 bottom-up saliency and top-down control converged. Two distinct cerebral networks exerted top-down control: distractor suppression engaged the left intraparietal sulcus, while target enhancement involved the frontal eye field and lateral occipital cortex. Hence, attentional selection is implemented in integrated maps in visual cortex, which provide precise topographic information about target-distractor locations thus allowing for successful visual search.

Keywords: attention, functional magnetic resonance imaging, retinotopic mapping, saliency, top-down

Introduction

Visual search is ubiquitous: We engage in it several hundred times a day, for example, when looking for a pencil on the desk or a road on a map. Traditionally, visual search is divided into 2 types: one where target selection is purely bottom-up (sensory) driven and another that requires top-down control. In the former, an object pops out and automatically grasps attention due to its saliency (e.g., a red tomato among green apples). In the latter, endogenous control is required to meet the observer's goals. These 2 forms of attention must be coordinated to guide behavior. In natural and laboratory settings, it has been shown that hardly any search is purely bottom-up or purely top-down driven (Wolfe 2007; Einhäuser et al. 2008). This has led to the proposal that both types interact and create an integrated saliency map in which the relative sensory strength (the degree of difference to the other objects) and the relevance of each object in a scene are topographically represented (Treue 2003; Wolfe 2007; Müller et al. 2009).

Because the visual system possesses high spatial resolution, it is thought that this map is distributed across early visual cortical areas (Treue 2003) and is under the control of a frontoparietal attention network. Accordingly, studies in animals have shown that responses in lower visual areas (V1, V2, and V4) depend on the stimulus saliency but also on its behavioral relevance (Kastner et al. 1997; Nothdurft et al. 1999; Lee et al. 2002; Ogawa and Komatsu 2004; Burrows and Moore 2009). However, direct evidence for saliency maps in the human brain and how they are organized along the visual hierarchy is missing. Furthermore, it is currently unknown how saliency is computed across the cascade of striate to extrastriate areas and where those signals interact with top-down guidance. Answering these questions requires simultaneous access to many areas, which remains a challenge in electrophysiological recordings in animals, and experimental designs that manipulate both bottom-up and top-down control.

Here, we combined functional magnetic resonance imaging (fMRI) in human subjects with a visual search task in which bottom-up saliency and top-down control were independently manipulated and applied the logic of animal electrophysiology to our data analysis. In particular, we made use of localizer and retinotopic mapping procedures enabling us to explicitly assign retinotopic activation to target or distractor stimuli across the whole visual hierarchy (V1 to hV4). Only a retinotopically specific differentiation between target and distractors qualifies an area as a physiological correlate of a salience map. This procedure enabled us to unequivocally identify brain areas as carriers of saliency maps. We further asked how these maps are influenced by bottom-up or top-down manipulations, whether there is a gradient in modulation across the visual hierarchy, and which areas exert top-down control over this differentiation. We expected bottom-up saliency to be reflected in V1 (Levitt and Lund 1997; Nothdurft et al. 1999; Beck and Kastner 2005; Zhaoping and May 2007), while top-down control was expected to modulate responses in extrastriate areas (Luck et al. 1997; Kastner et al. 1998; Beck and Kastner 2005). A likely site of integration of bottom-up saliency and top-down guidance was area hV4 (Ogawa and Komatsu 2006; Burrows and Moore 2009), which has previously, on the basis of electroencephalographic dynamics, been suggested to harbor an overall saliency map (Töllner, Zehetleitner, Gramann, et al. 2011; Töllner, Zehetleitner, Krummenacher, et al. 2011).

Beyond early visual areas, we also investigated the sources of top-down control at the level of the whole brain. In visual search, top-down attention can operate by enhancing target salience or via suppressing distracting stimuli (Müller and Ebeling 2008). It is currently unknown whether the same networks control both processes.

In the current study, we made use of the high spatial resolution offered by fMRI and provide definitive evidence for
hierarchically organized saliency maps in human visual cortex, with increasing saliency levels from V1 to hV4 and differential modulations depending on whether saliency is determined by top-down, bottom-up, or a combination of both: In V1, saliency maps were based purely on bottom-up signals, whereas hV4 qualified as the locus of convergence between bottom-up saliency processing and top-down control. Target enhancement was mainly controlled by the frontal eye field (FEF), and distractor suppression was traced back to activity in the left intraparietal sulcus (IPS).

Materials and Methods

Experimental Procedures

Participants
Seventeen healthy volunteers participated in the fMRI study (mean age, 26 years; age range, 18-34 years; 11 females). All subjects were right handed, had normal or corrected-to-normal vision, and no history of neurological or psychiatric disorders. One further subject took part in the fMRI study; however, her neuroimaging data were discarded due to technical artifacts. All subjects gave written informed consent before participation in accordance with the Declaration of Helsinki and approved by the local ethics committee.

Visual Search Experiment: Stimuli and Task
Four gratings (spatial frequency = 3 cycles/degree [cpd], 1.5° × 1.5° visual angle [vda]) were arranged around the fixation point of an imaginary circle (vda eccentricity from fixation to the center of each grating). Stimuli were either green (CIELab 87.23, -85.84, 82.85) or purple (CIELab 59.92, 97.72, -60.51). Visual stimuli were presented using an MR-compatible gooseneck system with 2 organic light-emitting diode displays (MR Vision 2000; Resonance Technology, Northridge, CA) with a resolution of 800 × 600 pixels at a refresh rate of 60 Hz, located at a virtual distance of 1.2 m from the subjects. The visible screen size subtended 30° × 22.5° vda in the horizontal and vertical plane, and the gray background luminance was 10.09 cd/m². Presentation software (version 10.3) was used for stimulus presentation and response collection.

Subjects were instructed to indicate the location of a target stimulus—a horizontal grating—as quickly and accurately as possible, by pressing 1 of 4 linearly arranged buttons on a custom-made response box with their right hand. The outer left button was mapped to the lower left quadrant, the left button to the upper left quadrant, the right button to the upper right quadrant, and the outer right button to the lower right quadrant. The overall saliency of the target was manipulated by altering the degree of bottom-up saliency as well as the availability of top-down guidance. To alter bottom-up saliency, we manipulated the distractors and created 3 conditions: target singleton in orientation (TSO), target singleton in orientation and color (TSOC), and target singleton in orientation and a distractor singleton in color (TSO-DSC) (Fig 1a). Orientation was the task-relevant dimension, while color was always task irrelevant. We manipulated top-down control by presenting conditions in a blocked or a mixed design. In a control experiment (n = 12; age range, 19–51; 10 females), we confirmed that orientation was processed in a bottom-up (pop-out) fashion in conditions TSO and TSOC, which was evidenced by the absence of a set size effect (TSO F1,11 = 2.180, P = 0.168; TSOC F1,11 = 879, P = 0.369) when presenting conditions either in a set size of 4 or 8 elements.

The experiment was completed in 2 sessions. In each session, subjects completed 3 blocked runs, one per stimulus array, and 3 mixed runs. Blocked runs contained 128 trials (25 per target location and 25 “null” baseline trials, i.e., displays in which only the fixation cross was visible on the screen, 2-back balanced randomization); mixed runs consisted of 131 trials (32 trials per stimulus arrays and 32 “null” baseline trials, 2-back balanced randomization). Three dummy trials were presented at the beginning of each. Each experimental trial lasted 4 s. A trial began with a fixation cross presented for 500 ms, followed by the stimulus display for 600 ms, replaced by a question mark which remained on the screen for 2900 ms (Fig 1b). Responses were recorded from stimulus display onset.

Localizer and Retinotopic Mapping
Each experimental session included a localizer experiment designed to determine the cortical representation in early visual areas of stimuli similar in location and size to the grating stimuli presented in the array (Supplementary Fig. 1). Subjects had to keep fixation while a contrast reversing black and white checkerboard was alternately presented in 1 of the 4 stimulus locations (spatial frequency: 1.2 cpd; inversion frequency: 8 Hz; luminance white: 43.95 cd/m²; luminance black: 1.295 cd/m²). Stimuli were presented in blocks of 16 s with 16 s fixation intervals that served as baseline (total duration 9 min 24 s).

Standard retinotopic mappings (for a detailed description, see Engel et al. 1994) were collected for each subject to assess the boundaries of retinotopic cortical areas V1, V2, V3, and hV4. In short, we presented a wedge-shaped black and white checkerboard pattern (18.4° vda), contrast reversing at a frequency of 4 Hz, which slowly rotated clockwise for a cycle of 360° within 66.5 s. Subjects had to fixate on a white cross while the wedge rotated. Twelve repetitions of full rotations were presented (total duration 13 min 48 s).

Magnetic Resonance Imaging Procedure
Functional and anatomical MRI data were acquired with a 3-T Siemens Allegra scanner (Erlangen, Germany) with a standard head coil. Functional data for the localizer and visual search experiment were collected using a T2*-weighted echo-planar imaging (EPI) sequence with the following parameters: repetition time (TR): 2000 ms; echo time (TE): 30 ms; flip angle (FA): 77°; field of view (FOV): 220 mm; voxel size 3.4 × 3.4 × 3.4 mm; gap thickness: 0.3 mm. We acquired 266 or 274 volumes containing 35 transversally oriented slices for each blocked and mixed functional run, respectively. Two hundred and eighty volumes were collected for the localizer experiment. For the retinotopic mapping, we acquired 396 volumes containing 34 slices per functional run (EPI, TR: 2080 ms; TE: 30 ms; FA: 80°; FOV: 210 mm; voxel size 3.3 × 3.3 × 3.3 mm; gap thickness: 0.22 mm). Spatial distortions in the EPI images were corrected using a point spread function (Zaitsev et al. 2004). High-resolution anatomic images were obtained using a T1-weighted magnetization prepared rapid acquisition gradient echo sequence (160 slices; TR: 2500 ms; TE: 3.93 ms; FA: 12°;
Statistical analyses were conducted in SPSS (v17, SPSS Inc.), and in Matlab using custom code. The first 2 volumes of each event-related run were discarded to preclude T1 saturation effects. Preprocessing of the functional data included motion correction (Fornisano et al. 2005), linear trend removal and temporal high-pass filtering at 0.0054 Hz, and slice scantime correction with sinc interpolation. The 2D functional images were coregistered with the 3D anatomical data, and both 2D and 3D data were spatially normalized into the Talairach coordinate space (Talairach and Tournoux 1988). Functional 3D data were spatially smoothed with a Gaussian kernel (8 mm full-width at half-maximum) for the whole-brain analysis only. To create inflated surface reconstructions, the gray-white matter boundary was segmented, reconstructed, smoothed, and inflated (Kriegeskorte and Goebel 2001) separately for each hemisphere.

Our rapid event-related fMRI study used closely spaced trials, leading to a substantial overlap in the resulting hemodynamic responses. Nevertheless, when the precautions of a balanced randomization are taken (2-back randomization, see above), the underlying hemodynamic responses can be assessed by deconvolution (Dale and Buckner 1997). A deconvolution analysis estimates the hemodynamic response function for each trial on the basis of a general linear model (GLM). Ten stick predictors, one per volume, were defined to cover the temporal extent of a typical hemodynamic response (20 s).

To investigate interactions between top-down control and stimulus-driven attentional selection, we performed 2 analyses: a region of interest (ROI) analysis in early visual areas (V1 to hV4) and a whole-brain analysis. The first analysis investigated how saliency is reflected in early visual areas, and which factors modulate it: bottom-up saliency, top-down guidance, or an interaction between bottom-up and top-down control, while the second analysis investigated the cerebral networks involved in top-down control, particularly target enhancement and distractor suppression.

## ROI-based analysis
First, the borders of areas V1-hV4 were determined for each subject based on standard retinotopic mapping procedures (Engel et al. 1994; Kastner et al. 1998). Then, ROIs for the cortical representation of the stimulus location were determined from the localizer experiment in each subject and mapped out for each early visual area (V1-hV4) as previously defined by retinotopic mapping. The average number of voxels acquired per visual area was 376 standard deviation [SD] = 130) for V1, 393 (SD = 124) for V2, 409 (SD = 120) for V3, and 348 (SD = 116) for hV4. The average t-value at which regions were mapped in each visual area was t = 3 (SD = 0.2). Parameter estimates (beta values) of time courses of activation were extracted for each experimental condition in each subject in each ROI separately for target- and distractor-related activity. The observed peaks (points 3–4) of the group-averaged hemodynamic response function within V1-hV4 were extracted and subjected to ANOVAs. Activity from the different quadrants was collapsed because we did not have a quadrant-related hypothesis for retinotopic areas. A first analysis included an rmANOVA with factors visual areas (V1, V2, V3, and hV4), stimulus type (target and distractor), block type (blocked and mixed), and stimulus array (TSO, TSO-DSC, and TSOaC). To further study the neural manifestation of saliency, we examined the relative processing of target as compared with distractor stimuli. For that, we computed an index of differential activation (saliency index, SI) in which beta values at the peak of the blood oxygen level-dependent (BOLD) response of distractors were subtracted from those of the target per condition and visual area and then submitted to an rmANOVA with visual areas, block type, and stimulus array as factors. Positive numbers indicate that targets elicited higher BOLD activity than distractors. In theory, the SI should subsume all factors that define the saliency of a stimulus at a given moment in time (Töllner, Müller, et al. 2011), that is, top-down perceptual set, bottom-up salience, and intertrial factors. Our study explicitly manipulated bottom-up and top-down factors, and thus, we expected their contribution to be reflected in the SI. In contrast, we did not expect a high contribution of intertrial history since due to restriction imposed by the deconvolution analysis of our event-related fMRI design we introduced a 2-back balance of the history of conditions, that is, all possible combinations of condition triplets are equally likely in our design. This effectively controls, at least in short runs of trials, for the intertrial history.

Whole-brain analysis. Statistical analyses were conducted volume based. Multisubject random effects GLMs were computed over the third to sixth predictor, thus covering the peak of the BOLD response while accounting for variability in the peak time in different subjects and brain regions (for a similar approach, see Serences et al. 2005), corrected for serial correlation and z-normalized. To correct the multisubject statistical maps for multiple comparisons, we performed 3D cluster-size thresholding (Thirion et al. 2007) by use of a Monte Carlo simulation. The original map for direct contrasts was thresholded at t66 = 4 and for interactions at t66 = 3. The estimated cluster size threshold for P < 0.05 (1500 iterations) was 79 mm2/381 voxels and 165 mm3/189 voxels, respectively.

Of note is the fact that to investigate the effects of top-down guidance on stimulus saliency, the stimulus arrays were embedded in different block types, which were collected in separate fMRI runs. Such a manipulation, as shown in previous psychophysical studies (Wolfe et al. 2003; Müller et al. 2009), modifies search strategies leading to measurable behavioral changes, that is, reduction of the interference of a highly salient distractor or enhancement of the facilitatory effects of a double singleton target. Note, however, that only the instructions but not the stimuli differed between conditions. It is thus conceivable that this form of top-down guidance introduces differences in the baseline, as it effectively constitutes a state variable and should be in place even before a stimulus is presented. For instance, weighing of a stimulus dimension is expected to occur before the stimulus is presented, but its effect will only be seen once the stimulus display is shown. In order not to confound the effects of top-down with trivial effects such as the passage of time, we carefully pseudorandomized the order in which the blocked and mixed runs were acquired over subjects using Latin squares. Thus, any significant difference arising from a comparison between blocked and mixed runs must be systematic and cannot be explained by scanner noise, etc. In addition, each run was normalized to its baseline, which puts all runs numerically onto the same level.

## Results
We collected BOLD signals during a visual search task, in which subjects had to localize a target stimulus (a horizontal grating)
as quickly and accurately as possible. We varied the strength of "bottom-up saliency" of the target by modifying the distractor stimuli, causing the target to pop out relatively more or less in 3 separate stimulus arrays (Fig. 1). The stimulus array contained either 1) a target that was singleton (i.e., no other stimulus in the display shared the same feature) in the task-relevant dimension, namely orientation (TSO), 2) a target that was singleton in 2 dimensions, one relevant (orientation) and one irrelevant (color) (TSOC), or 3) a target that was singleton in the relevant dimension (orientation), presented together with a distractor that was singleton in an irrelevant dimension (color) (TSO-DSC), creating a scenario of high competition between target and distractors. Thus, both stimulus arrays TSOC and TSO-DSC contained color singletons that were task irrelevant, preventing any incentive to attend to them deliberately (Yantis and Egeth 1999). However, in stimulus array TSO, the color singleton facilitated target detection (Krummenacher et al. 2001), whereas in TSO-DSC, the color singleton interfered with the correct identification of the target (Theeuwes 1994). Note that the actual target stimulus never changed.

"Top-down control" was manipulated by altering the context in which each stimulus array occurred. They could either be presented in a blocked or mixed order, thus providing subjects with a priori information about whether the irrelevant dimension should be enhanced or suppressed (Wolfe et al. 2003; Müller et al. 2009). We predicted that presenting condition TSOC in isolation (blocked) would enhance color processing, reflecting the facilitating effect of the color singleton (saliency enhancement). In contrast, in condition TSO-DSC, when the highly salient color distractor competed with target selection based on orientation, we expected that color processing would be suppressed, hence diminishing its interference (distractor suppression). In the mixed, trial types were presented in randomized order. Thus, since subjects did not know whether color processing would facilitate or hinder target detection in the upcoming trial, a differential modulation reflecting saliency enhancement and distractor suppression was not expected. Hence, top-down effects should reduce the interference of an irrelevant salient distractor and/or augment facilitation when the irrelevant feature coincides with the target.

**Task Performance**

Error rates were low (mean 1.57%, standard error 0.337%) but differed between stimulus arrays ($F_{1,929.32.793} = 10.491$, $P < 0.001$): subjects’ committed less errors on stimulus array TSOC than on TSO-DSC ($P = 0.001$) and TSO ($P = 0.036$), confirming that a double singleton target is more easily localized. No difference was observed between TSO and TSO-DSC ($P = 0.275$). No other effects were significant ($P > 0.4$).

Figure 2a illustrates the mean lnRTs. For illustration purposes and to ease the comparability with previous studies, Supplementary Figure 2 shows the RT back-transformed to milliseconds. Clear differences in lnRT were observed between stimulus arrays ($F_{1,795.30.508} = 51.026, P < 0.0001$). Critically, top-down control modulated the processing of bottom-up saliency (block type x stimulus array, $F_{1,301.30.665} = 4.599, P = 0.021$). To explore the effect of an irrelevant color singleton on target detection and how it is modulated by block type, we computed an index of facilitation and interference. In brief, lnRT of condition TSO (without a color singleton) was taken as baseline and subtracted from the other conditions containing a task irrelevant color singleton that could either facilitate (TSOC) or interfere (TSO-DSC) with target detection. Negative values indicate facilitation while positive values indicate interference. Condition TSO showed facilitatory effects both in the blocked and mixed conditions (both means > 0, both $P < 0.00001$, Fig. 2b). We found a bigger advantage for the blocked than the mixed condition ($t_{17} = 2.634, P = 0.017$), confirming the effect of top-down control even in efficient visual search. The color singleton distractor in condition TSO-DSC only interfered when top-down control was reduced, that is, in the mixed condition ($t_{17} = 2.066, P = 0.054$). However, it did not produce interference in the blocked condition when top-down control was available ($P > 0.3$). Together, these results indicate that top-down control can alter the weighting of a stimulus dimension (Wolfe et al. 2003; Müller et al. 2009), such that when subjects can exploit a priori information, for example, that an irrelevant but salient feature never belongs to the target, this feature can be suppressed, resulting in reduced interference.

**Top-Down and Bottom-Up Effects in Visual Areas**

To investigate the correlates of saliency in visual areas (V1-hV4) and how these maps are influenced by bottom-up and top-down manipulations, we used localizer and retinotopic mapping procedures (see Materials and Methods, Supplementary Fig. 1). This allowed us to explicitly assign retinotopic activation to target or distractor stimuli across the visual hierarchy and thus to identify brain areas as carriers of a saliency map.
Figure 3 shows the average estimated BOLD response for target and distractor ROIs per visual area (Supplementary Figs 3 and 4 show BOLD responses per experimental condition and distractor type). Retinotopic subregions containing a target showed higher activation than those containing a distractor \((F_{1,16} = 19.154, P < 0.001\)) from V1 to hV4 (all \(I_{16} > 3.3, \text{all } P < 0.01\)). Thus, correlates of saliency (the differentiation between target and distractors) could be observed throughout the visual hierarchy. Importantly, this differentiation was modulated by block type \((F_{1,49.23.380} = 6.062, P < 0.05\)) and stimulus array \((F_{1,16} = 4.281, P < 0.05\)). Furthermore, activation levels increased as a function of visual area \((F_{2,170.39.658} = 13.847, P < 0.001\)), and this increase was modulated by stimulus type \((F_{2,199.35.182} = 6.130, P < 0.01\)). Target activation increased more steeply than distractor activation. No other main effects or interactions were significant (all \(P > 0.1\)).

Having established that areas V1–hV4 differentiate between targets and distractors, we investigated how saliency was affected by relative bottom-up and top-down influences. For that, we computed an index of differential activation (SI) in which the BOLD response for distractor activation was subtracted from target activation (see Materials and Methods). This index reflects a measure of how distinct targets are from distractors. In a full ANOVA model which included all visual areas, we found that SI increased linearly as a function of visual area \((F_{2,195.35.184} = 6.130, P = 0.04\)) and differed in the presence of top-down: SI was higher for blocked than for mixed conditions \((F_{1,16} = 6.026, P = 0.026\)). Also, bottom-up saliency modulated SI levels \((F_{1,49.35.390} = 4.281, P = 0.035\)): TSO showed the highest SI levels, followed by TSO and TSO-DSC. Both conditions TSO and TSO-DSC had higher SI than condition TSO-DSC (both \(P < 0.006\)). Thus, conditions in which the target popped out most and where there was no interference from a salient distractor induced the highest SI levels. No other effects or interactions were significant (all \(P > 0.1\)).

We then investigated how bottom-up and top-down factors influence SI throughout the visual hierarchy. In V1, we found that SI was exclusively modulated by stimulus array \((F_{1,35.21.51} = 4.025, P = 0.047\)). As can be seen in Figure 4, stimulus array TSO (with the highest bottom-up saliency) exhibited higher SI than TSO \((I_{16} = 2.251, P = 0.039\)). No significant differences were observed for stimulus arrays with lower bottom-up saliency (TSO, TSO-DSC \(P > 0.1\)). These results suggest that V1 saliency is sensitive to manipulations of bottom-up information but not top-down. The opposite effect was found in V2, with higher SI for blocked than for mixed conditions \((F_{1,16} = 5.057, P = 0.039\)).

Networks for Top-Down Suppression and Enhancement

The previous analysis demonstrated a hierarchy of saliency maps along the visual hierarchy and the influence of top-down control in early visual cortex. However, it does not reveal the network involved in top-down control, and whether the constituents of this network depend on whether top-down acts by suppressing a highly salient distractor or by enhancing a less salient target. We thus conducted a whole-brain analysis to reveal the networks carrying top-down signals. We first investigated brain areas reflecting top-down control related to the “suppression of highly salient distractors,” which is the scenario for condition TSO-DSC. Top-down control was evidenced as less interference in the blocked than in the mixed condition. Then, to identify brain areas primarily involved in the top-down suppression of highly salient distracters, we contrasted the blocked versus the mixed condition for TSO-DSC with the blocked versus the mixed condition for TSO. In the blocked condition, both stimulus arrays shared activity related to target enhancement,
however, only stimulus array TSO-DSC had a highly salient distractor and should thus exhibit distractor suppression. As can be seen in Figure 5, the contrast revealed a cluster in the left IPS. Hence, top-down control in the presence of a highly salient distractor engaged the left IPS.

Since it is not clear whether relative decrements in the local saliency of highly efficient pop-out search lead to increments in top-down control and which brain areas underlie this effect (Wolfe et al. 2003), we turned to the highly efficient conditions. Double singletons like TSOC are associated with higher local saliency and thus might require less top-down guidance (Krummenacher et al. 2001). Alternatively, bottom-up saliency in pop-out displays could be sufficient to direct attention without additional top-down control, and thus, no brain regions should show differential activation. However, when we compared the 2 efficient pop-out stimulus arrays TSO and TSOC in the blocked condition, we found higher activation in the right FEF for stimulus array TSO than TSOC (Fig. 6a); left FEF was also found at a more lenient threshold ($P < 0.01$, uncorrected). This suggests that top-down control is engaged even in highly efficient search.

To reveal brain areas involved in "target enhancement" through top-down control, we contrasted the blocked versus the mixed condition for stimulus array TSO with the blocked versus the mixed condition for stimulus array TSOC. Top-down control related to target enhancement would imply altering the weights of the task-relevant feature (orientation), which would be most evident for the condition with relatively lower bottom-up saliency (TSO). As can be seen in Figure 6b, we found activation in the right FEF (also left FEF at $P < 0.05$, uncorrected) and 2 clusters in the left lateral occipital complex (LOC). Thus, the networks involved in top-down control differed depending on whether top-down control is engaged due to competition with a highly salient distractor or due to relative decrements in bottom-up signals.

**Brain–Behavior Correlations**

To further elucidate the relationship between frontal and visual areas in visual search, we correlated reaction times (lnRT) with fMRI signals focusing on the pop-out stimulus arrays (TSOC, TSO) when top-down control was highest, that is, in the blocked condition. The direction of the association allowed us to infer whether brain areas reflect local saliency or top-down control. Figure 7 shows that activity in early visual cortex, V1 and hV4, "negatively" correlated with lnRT in both conditions, TSOC and TSO. In contrast, activity in frontal regions, right and left FEF, "positively" correlated with lnRT in both conditions. No other correlations were significant (2 LOC clusters). In addition, across subjects activity in hV4 inversely correlated with right FEF for both TSO and TSOC conditions (TSO $R = -0.453$, $P = 0.034$; TSOC $R = -0.468$, $P = 0.029$ one sided). No other correlations were significant. Thus, while higher activity in V1 and hV4 correlated with faster reaction times, higher activity in FEF correlated with slower reaction times. This indicates that when bottom-up saliency is high, seen as higher activity in V1 and hV4, less top-down control is required, seen as lower activity in FEF.

**Figure 4.** Index of saliency: ROIs (visual cortex). Parameters estimate for the conditions of interest in V1, V2, V3, and hV4. Parameters estimate at the observed peak (points 3–4) of the group-averaged hemodynamic response function for target and distractor activity were extracted and then subtracted to create an index of differential activation (SI) per condition and visual area. Note that factors modulating saliency differ across the visual hierarchy: An effect of stimulus array was evident in V1, while V2 activity differed for task set. In hV4, we observed an interaction between stimulus array and task set. Note that no effect of stimulus array is observed in the blocked nor in the mixed condition in area V2 and V3 (all $P > 0.14$, all $\eta^2 < 0.11$). We note nonetheless that areas V2 and hV4 display a similar BOLD signal pattern for the blocked trials that is indeed congruent with top-down modulations but only area hV4 showed a significant interaction.

**Figure 5.** Top-down control in the presence of salient irrelevant distractors. The suppression of an irrelevant distractor activated the left IPS as measured by the contrast [TSo-DSC$_{\text{block}}$ vs. TSo-DSC$_{\text{mixed}}$] > [TSo$_{\text{block}}$ vs. TSo$_{\text{mixed}}$].
Finally, we investigated the relationship between the behavioral index of interference and activation in each of the relevant areas, which were found active when the stimulus array TSO-DSC was contrasted with other stimulus arrays (left IPS as well as activity in V1 and hV4). However, no significant correlations were observed for those contrasts (all $P > 0.1$).

Discussion

Interactions between Top-Down Control and Bottom-Up Saliency in Visual Cortex

We made use of the exquisite spatial resolution offered by fMRI to investigate how saliency signals are computed across the visual hierarchy and what determines the responses of a given area, that is, bottom-up saliency, top-down control, or an interaction of both. By manipulating task strategies, we could show that top-down attention modulates bottom-up saliency, thus suppressing the detrimental effects of highly salient distractors. Our findings highlight the flexibility and advantages offered by top-down control (Leber and Egeth 2006; Einhäuser et al. 2008; Müller et al. 2009).

We provide evidence of hierarchically organized saliency maps in human visual cortex. We show that in V1, bottom-up saliency has a clear effect on the amplitude of the BOLD response: The stimulus array with the highest pop-out saliency (TSOC) elicited the highest activation. This is in line with the proposal that V1 carries a purely bottom-up saliency map...
(Zhaoping and May 2007) as we did not observe a top-down modulation. In contrast, we found an effect of block type in V2 and hV4, revealing that top-down control influences visual processing at a very early stage. Previous studies have also demonstrated that top-down control modulates extrastriate areas but not V1 (Luck et al. 1997; Kastner et al. 1998). In monkeys, spontaneous firing rates in areas V2 and V4 increase when attention is directed into the receptive field (RF) even when no stimulus is present (Luck et al. 1997), reflecting a top-down signal (Moore 2006). Top-down control shows a gradient across early visual cortex: The earliest and strongest attentional modulations occur in V4, followed by intermediate effects on amplitude and latency in V2 and very late and weak effects in V1 (Buffalo et al. 2010).

We found the earliest locus of convergence between bottom-up and top-down attentions in hV4. Experiments in macaques have also revealed that both bottom-up activation and top-down activation are represented in V4 during visual search (Ogawa and Komatsu 2006). Importantly, population responses are higher for targets than for distractors (Fecteau and Munoz 2006). Recently, it was shown that V4 neurons respond more vigorously to pop-out arrays than to conjunction arrays, suggesting that V4 neurons are modulated by bottom-up saliency (Burrows and Moore 2009). However, when top-down attention was directed outside the RF, this selectivity was eliminated. Thus, the mere presence of a pop-out stimulus is not sufficient to drive selection by V4 neurons but requires top-down attentional modulation. Furthermore, in line with our results and previous behavioral (Krummenacher et al. 2001; Weidner and Müller 2009) and computational (Soltoni and Koch 2010) visual search studies, these authors found higher saliency modulation for double singletons (unique in orientation and color) than for single singletons (unique in one feature only). This is complemented by recent work showing that the N2pc, an event-related potential component thought to reflect attentional selection and to originate in area hV4 or inferotemporal cortex (Hopf et al. 2006), arises earlier in response to double feature singletons as compared with single singletons (Töllner, Zehetleitner, Krummenacher, et al. 2011).

The specific nature of the interaction found in hV4 suggests that subjects were able to use top-down information to enhance target saliency. While in the mixed condition, all stimulus arrays were processed similarly, in the blocked condition, when certain features of bottom-up information could be strategically enhanced to detect the target stimulus faster, we observed that stimulus arrays TSO and TSO produced more activation than stimulus array TSO-DSC. This indicates that top-down control interacts with bottom-up stimulus processing already at the level of hV4. The behavioral results support our interpretation: Larger facilitatory effects of a double pop-out target and smaller interference effects from a salient distractor were observed in the blocked as compared with the mixed condition. These results show that top-down guidance can alter stimulus saliency at early stages of processing (hV4), also providing support to the Dimensional Weighting Theory which postulates that saliency maps are penetrable by top-down strategies (Töllner, Müller, et al. 2011).

Crucially, these results cannot be explained by differences in the physical properties of the stimuli, as the effects represent an interaction. That is, despite differences in sensory stimulation, the very same physical stimuli did not elicit differential effects in the mixed but only in the blocked conditions. The effects are also not attributable to differences in eye movements, as fixation stability did not differ between experimental conditions (see Supplementary Analysis: Eye Tracking Data Analysis and Supplementary Fig. 5).

Interestingly, we observed that bottom-up signals coded at the lowest level of V1 interact with top-down signals at the higher level of hV4 apparently bypassing area V2. Indeed, direct anatomical connections from V1 to V4 have been shown in the macaque monkey in several studies (Yukie and Iwai 1985; Ungerleider et al. 2008), which could explain why bottom-up effects emerged in V1 and V4 but bypassed V2. However, considering the similar response profile in V2 and hV4, it is also possible that some bottom-up effects are mediated via V2. Further studies will shed light on this issue.

Theoretical accounts of saliency computation, such as Guided Search (GS) (Wolfe 2007) and Dimensional Weighting Account (DWA) (Müller et al. 1995) postulate the existence of an integrated (feature independent) saliency map for which neurophysiological evidence has been found (reviewed in Fecteau and Munoz 2006). One key prediction of the GS and DWA architecture is that the overall saliency map should reside always at the same level of the hierarchy, despite how much top-down or bottom-up resources are engaged. An alternative view (Treue 2003), however, proposes that saliency maps are distributed along the visual processing pathway and that perceptual selection is embodied in the activity of the area that best matches the task at hand. A key prediction of the distributed architecture is that the area exhibiting selection would depend on the task at hand, and the strategy employed to accomplish it. Our results are more consist with the later view as we found evidence of selection at different hierarchical levels depending on the engagement of top-down resources, that is, clear signs of selection were observed in hV4, but mostly when top-down was high (blocked condition). This result, in addition to the correlation analysis (see below) showing association between RT and activity levels at several stages of the hierarchy (V1, hV4, and FEF), led us to conclude that saliency is computed in a distributed fashion, much in line with Treue’s proposal. Further support for such an account comes from the study of Burrows and Moore (2009), who found that the mere presence of a highly conspicuous stimulus (pop-out) is not sufficient to drive selection by V4 neurons but requires top-down attentional modulation. In sum, our results provide evidence for a multistage selection process with a gradient of saliency maps in which either bottom-up, top-down, or a combination of both is represented in early visual area.

**Top-Down Control Facilitates the Search of a Salient Target**

We found that visual search involving highly salient stimuli is not exclusively stimulus driven but also engages top-down control (Müller 2003; Wolfe 2007): A double singleton target was detected faster when it could be predicted (in blocked trials) than when it could not (in mixed trials). Also, evidence of saliency was observed at the level of hV4, but only when top-down attention was available. Furthermore, we observed that stimulus array TSO, despite containing a highly salient target, induced higher activation in FEF (a potential source of top-down modulation; Moore 2006) than TSO-D, the condition with the highest bottom-up saliency. Top-down control exerted by
frontal regions might then act by altering the weights of task-relevant features (e.g., orientation) represented in intermediate visual areas (LOC) (Larsson et al. 2006; Montaser-Kouchaki et al. 2007), enhancing target saliency.

The inverse sign of the neurobehavior correlations observed in early visual and frontal areas (faster RTs correlated with higher activation in V1 and hV4 and lower activation in FEF), suggests that attentional selection is a multistage process, where the strength of the bottom-up signal determines the stage at which selection occurs, and the stage at which top-down control is recruited. Other studies lend support to this interpretation, as they have also found increased BOLD activity in frontal regions (FEF) of the monkey and human brain for pop-out displays (Anderson et al. 2007; Wardak et al. 2010) and increased BOLD activity in FEF as target saliency is decreased (Paus 1996; Luna et al. 1998). Furthermore, bottom-up stimulation is not sufficient to drive selection in V4 neurons but requires top-down control (Burrows and Moore 2009), and microstimulation of FEF modulates early visual areas only when bottom-up signals are present (Ekstrom et al. 2008). Importantly, attention has been shown to increase the coupling between FEF and V4, which is initiated by FEF (Gregoriou et al. 2009). Together, the current findings show that attentional selection, even of highly salient stimuli, involves top-down control lending support to theories, such as "Guided Search," GS (Wolfe 2007), or "Dimension Weighting," DW (Müller et al. 1995), which consider attentional selection to result from the interaction of bottom-up salience and top-down guidance. In addition, they highlight that attention is embodied in a network of areas involving frontal, parietal, and early sensory cortices.

**Top-Down Control in the Presence of a Highly Salient Distractor**

Top-down control in the presence of a highly salient distractor recruited different areas than highly efficient search, in particular the left IPS. The specific involvement of IPS in suppressing salient distractors fits well with previous studies showing that interference from highly salient distractors increases alongside activity in occipital regions when transcranial magnetic stimulation (TMS) is applied over left IPS (Mevorach et al. 2010). This suggests that left IPS suppresses salient but irrelevant signals in early visual areas. Along the same lines, it is known that applying TMS over the left IPS immediately before target presentation increases the interference of a highly salient distractor (Mevorach et al. 2009). Surprisingly, in our study, we did not observe a correlation between the index of interference and activity levels in left IPS. This negative result should be interpreted with caution as the interference levels observed in our study where relatively small. Considering the causal evidence provided by previous studies combining fMRI and TMS, we believe that the absence of a correlation is likely due to the small dynamic range of the effect in our data. Further studies will shed light on this issue. Altogether, left IPS appears to be involved in suppressing the influence of highly salient distractors. Here, prior knowledge of the salient but irrelevant feature results in the active suppression of that feature.

**Conclusion**

We provide clear evidence of hierarchically organized saliency maps in human visual cortex, with pure bottom-up signals represented at the level of V1, and a gradient of top-down modulation observed in extrastriate areas. While V2 is mostly responsive to top-down modulations, hV4 qualifies as the locus of convergence between saliency processing and top-down guidance, mainly exerted by FEF. Attentional selection might then be understood as a multidimensional process in which strong bottom-up signals are progressively deemphasized and combined with the goals of the observer, allowing the requirements of an ever-changing environment to be met (Treue 2005). The inverse relationship between activation in early visual areas and frontal regions and the speed of target localization in pop-out displays highlights that both top-down and bottom-up attention contribute to target selection (Wolfe 2007). Furthermore, 2 distinct functional networks were engaged in the top-down control of target enhancement and distractor suppression: FEF and LOC were active when bottom-up saliency signals of the target should be enhanced; conversely, a parietal network was engaged when salient bottom-up distractors should be suppressed. Thus, the engagement of bottom-up and top-down attention can be flexibly regulated in order to accomplish the task at hand; and the system can adjust both the involvement of top-down resources when bottom-up information is not strong enough and the mechanism through which top-down control guides attentional selection.

**Supplementary Material**

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/

**Funding**

Max Planck Society, Dr Paul and Cilli Weill Foundation, August Scheidel Foundation, and German Ministry of Education and Research (MU 1364/3).

**Notes**

We are indebted to Caspar M. Schwiedrzik for his insightful comments and help with data analysis, Ayelet N. Landau for helpful comments on a previous version of this manuscript, Wolf Singer for his generous support, and 2 anonymous reviewers for their comments during the revision of the manuscript. Conflict of Interest: None declared.

**References**


Copyright 2012 Oxford University Press for the Association of British Neurological Surgeons.


Percept Psychophys. 29:1021-1035.


