Investigating task-dependent top-down effects on overt visual attention

Torsten Betz
Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany ✉

Tim C. Kietzmann
Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany 🏠 ✉

Niklas Wilming
Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany ✉

Peter König
Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany 🏠 ✉

Different tasks can induce different viewing behavior, yet it is still an open question how or whether at all high-level task information interacts with the bottom-up processing of stimulus-related information. Two possible causal routes are considered in this paper. Firstly, the weak top-down hypothesis, according to which top-down effects are mediated by changes of feature weights in the bottom-up system. Secondly, the strong top-down hypothesis, which proposes that top-down information acts independently of the bottom-up process. To clarify the influences of these different routes, viewing behavior was recorded on web pages for three different tasks: free viewing, content awareness, and information search. The data reveal significant task-dependent differences in viewing behavior that are accompanied by minor changes in feature-fixation correlations. Extensive computational modeling shows that these small but significant changes are insufficient to explain the observed differences in viewing behavior. Collectively, the results show that task-dependent differences in the current setting are not mediated by a reweighting of features in the bottom-up hierarchy, ruling out the weak top-down hypothesis. Consequently, the strong top-down hypothesis is the most viable explanation for the observed data.

Introduction

In recent years, research on selective attention has moved to the center of scientific interest (Knudsen, 2007). In investigations of behavioral effects, the study of eye movements is a widely used technique for measuring shifts of attention. Building on a large body of studies examining well-controlled artificial stimuli, more and more investigations are turning to ecologically relevant stimuli like natural scenes. Although research in this field has made considerable progress, many interesting issues remain unsolved. An important aspect of visual attention deals with how goal-directed behavior can alter eye-movement patterns. Very early, it was demonstrated that, depending on the specific task, the same image might be scanned by different saccade sequences (Yarbus, 1967; but see also Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Nelson, Cottrell, Movellan, & Sereno, 2004; Rothkopf, Ballard, & Hayhoe, 2007; Triesch, Ballard, Hayhoe, & Sullivan, 2003). Castelhano, Mack, and Henderson (2009)

show that the task influences which areas in an image might be viewed but leaves other measures of viewing behavior unchanged (like saccade amplitude and fixation duration). At the same time, selected fixation points show systematic relations to local image features (Baddeley & Tatler, 2006; Einhäuser & König, 2003; Reinagel & Zador, 1999).

Taken together, task effects and the correlation of fixated positions with image features raise the question of which causal route the task information takes when influencing viewing behavior. A number of models in the literature either explicitly or implicitly assume different causal routes. A prominent candidate is based on the computation of a saliency map. In a bottom-up directed process, different types of image features are analyzed at several spatial scales and integrated into one joint map (Koch & Ullman, 1985). Top-down directed signals might then act by modulating the relative influence of different features during the integration step (Itti & Koch, 2001; Navalpakkam & Itti, 2005). According to this view, the task modulates viewing behavior via changes in the

bottom-up system. In a radical shift, Ahissar and Hochstein (1997) proposed that all initial bottom-up directed processing is implicit and preattentive. Only after the gist of a scene has been recognized do top-down directed signals lead to attentional vision with scrutiny. In this scenario, the task is expected to influence behavior on a higher level without necessarily changing the bottom-up system. These models are not mutually exclusive, and recent studies have tried to combine properties of these extreme views (Torralba, Oliva, Castelhano, & Henderson, 2006). Still, the different models highlight the need to improve our understanding of the interaction of top-down and bottom-up signals.

In conclusion, one possible route through which top-down signals might change viewing behavior is by adjusting weights of feature maps in a bottom-up driven saliency map. This would happen on a slow time scale—that is, once per task—and will be referred to as "weak top-down" in the remainder of this article. A consequence of this causal route is that the correlation of fixation points with image features (feature-fixation correlation) is expected to be task-specific. Testing different tasks, it should thus be possible to find significant differences in that correlation. Specifically, when differences in the spatial layout of fixation points are observed, these must then be accompanied by changes in feature-fixation correlations. Furthermore, if weak top-down is the only explanation for task-dependent viewing behavior, these changes must be large enough to account for differences in the spatial layout of fixation points. An alternative causal route could directly connect the task with viewing behavior, which does not act through a slow modulation of bottom-up processing; this will henceforth be called "strong top-down". In this case, different tasks would not specifically influence correlation of fixation points with image features. Moreover, it should be possible to modify the distribution of fixation locations without observing relevant changes in relative feature-fixation correlations. As a result, the pattern of feature importance should be insufficient to explain viewing behavior. These two views are not exclusive; it is also conceivable that both routes play a role in the control of viewing behavior.

The study at hand was designed to clarify the influences of these different routes. To approach this issue, eye movements were recorded in a free viewing condition and in two tasks of increasing complexity. Viewing behavior is compared across these tasks, and changes in feature-fixation correlations and their explanatory power for task-dependent viewing behavior are investigated.

An important aspect of this study is the stimulus set chosen. There are manifold reasons to choose web pages as visual stimuli. Firstly, web pages allow an experimental setup closely resembling everyday situations. Web browsing naturally involves a screen, a mouse, and an office environment. 't Hart et al. (2009) recorded gaze movements with a fully mobile setup under truly natural conditions.

They were able to demonstrate marked differences in fixation behavior in the natural setup versus replay of the very same "natural" images under typical laboratory conditions. For web pages, the difference experienced by the subjects between everyday Internet surfing and the laboratory situation is much smaller. Specifically, care was taken to present the stimuli in a setup with screen, mouse, and office environment. Thus, present results presumably generalize well to the respective condition outside of the laboratory. Secondly, the use of web pages as stimuli allows for natural tasks, such as information search or free browsing, which constitute common day-to-day activities. Importantly, these activities are not directly and trivially dependent on specific visual features, avoiding a possible confound. In short, because of easily introduced top-down effects in the form of different tasks, and the associated natural experimental conditions, web pages form an excellent basis for the current study. Furthermore, although task-dependent effects on web pages have been addressed before, the literature is still contradictory on this issue. For instance, Rayner (1998) provided evidence for task-dependent viewing behavior, whereas Pan et al. (2004) could not support that observation in their experiment.

Methods

Participants

Forty-eight volunteers took part in the experiment (12 males and 36 females, ages 24–55, mean age 41.5 years). Inclusion in the study was contingent on reliable eye-tracking calibration. As compensation for participation, subjects received either payment or course credits. All subjects were informed about the experimental procedure, the eye-tracking device, and their right to withdraw from the experiment at any time. They gave written informed consent to participate. Subjects were naive to the purpose of the experiment but were informed about it upon completion of the task. Finally, because the presented web pages were in German, subjects were required to speak fluent German.

Stimuli

The set of stimuli contained 90 screenshots of web pages. These were taken from six different categories: news pages, blogs, landing pages, online shops, company homepages, and information pages. Each category was represented by 15 web pages.

Among the most important aspects of stimulus selection was the likelihood of subjects' prior experience with the
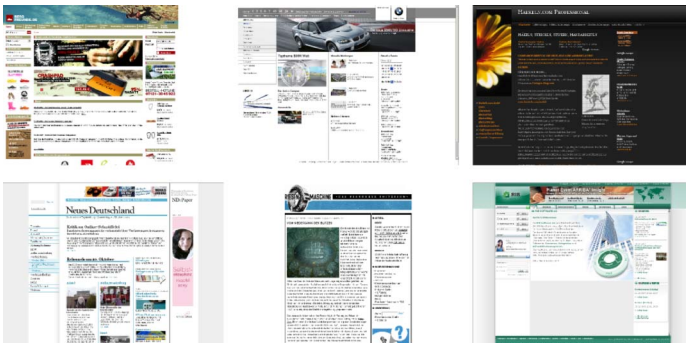
Figure 1. Sample web pages, one from each category.

stimuli. Since the experiment was designed to detect task-dependent top-down effects, all external top-down information that could interfere with the experiment, as in the case of prior knowledge, had to be diminished. To ensure this, only pages expected to be unknown to the subjects were chosen as stimuli.

To maximize natural browsing behavior, stimuli were designed to contain not only the page itself but also a standardized browser template (Firefox). Examples of stimuli from the different categories can be seen in Figure 1.

Each stimulus was shown only once to each subject to prevent possible memory effects. In addition, the sequence of tasks was randomized. The stimuli were selected such that three subjects would cover a presentation of all stimuli in all conditions and thus form one complete data set.

Apparatus

The experimental apparatus was designed to closely resemble a natural environment for the task of web browsing. For this reason, the setup contained a keyboard and a mouse, used by the subjects to provide input to an initial questionnaire and for feedback during the experiment. Distance to the screen was set at 60 cm to resemble normal working conditions. Stimuli were presented on a 19-inch flat screen monitor (SyncMaster 971p, Samsung Electronics, Seoul, South Korea). Screen resolution was set to 1280×1024 pixels; refresh rate was 75 Hz. The web pages had a size of 35.3×25.6 degrees, with the entire display (including the browser template) being 35.56×28.4 degrees.

The eye tracker used was an Eyelink II system (SR Research, Mississauga, Ontario, Canada). This head-mounted system is capable of tracking both eyes; however, only the eye giving a lower validation error after calibration was recorded. Sampling rate was set at 500 Hz. Saccade detection was based on three measures: eye movement of at least $0.1°$, with a velocity of at least $30°/s$ and an acceleration of at least $8000°/s^2$. After saccade onset, minimal saccade velocity was $25°/s$. No headrest

was used because it would have interfered with the natural browsing experience of the subjects.

Tasks

The experiment comprised three tasks. The first was *free viewing,* in which subjects were instructed to simply explore a web page without specific instructions. The second task, *content awareness,* was an extension of the first: after viewing the page, subjects had to select a user group that corresponded to the perceived target demographic of the web page. This task was fairly simple, because a stereotypical user group was included for each web page. For instance, a computer shop web page would include a programmer, schoolchild, priest, baker, and butcher as possible options. Nevertheless, this task required subjects to grasp the main contents of each page. In the third task, *information search,* a web page with links to search results was shown, from which participants could select individual links with the mouse. Here, the task was to consecutively rate the relevance of each of the five listed web pages with respect to the search term. This task demanded detailed analysis of the page, as the content had to be understood and related to the search term.

Because the stimuli were selected from 6 categories of 15 stimuli each, 30 web pages were selected for each task, 5 from each category. In the *free viewing* and *content awareness* tasks, stimuli were completely randomized. In the *information search* task, the presented search results were all taken from the same category and a matching search term was selected accordingly, ensuring that the search term would not be completely unrelated to any of the search results (Figure 2).
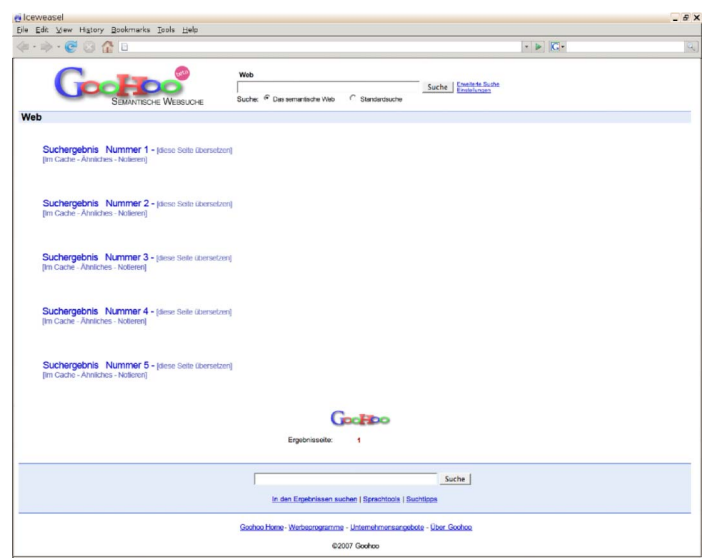

Figure 2. Search results page and mock-up browser window used as a frame for all stimulus presentations.
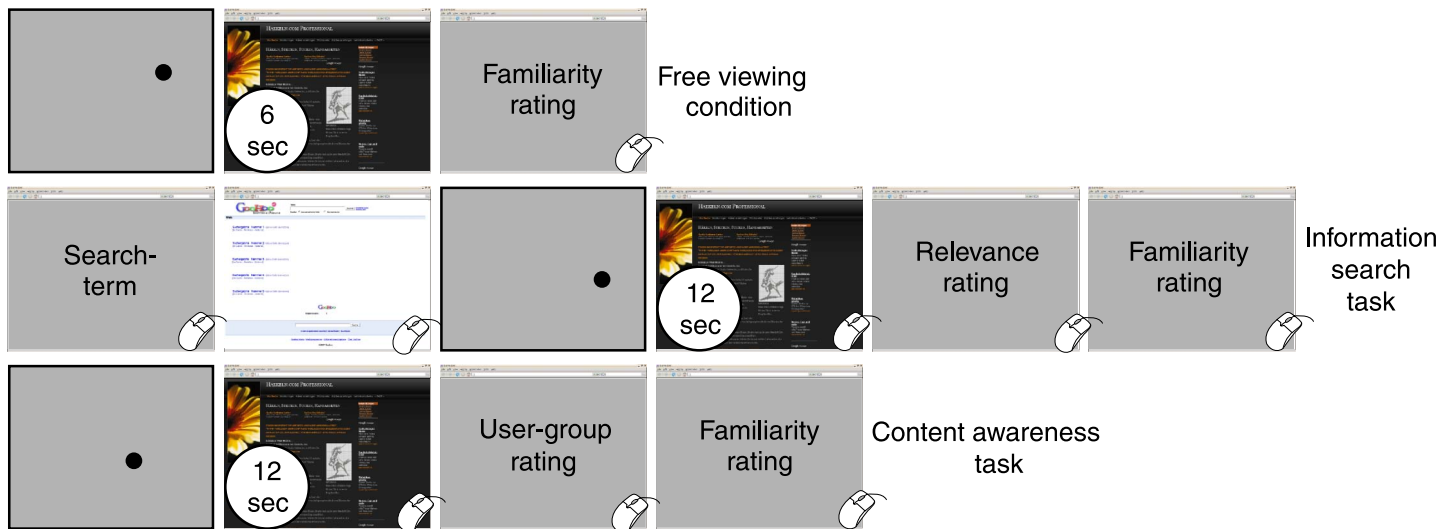
Figure 3. Overview of the procedure in the three different conditions (free viewing, information search task, content awareness task). The mouse icon symbolizes that interaction is possible, either to answer a question or to end stimulus presentation. The large circles indicate the maximal presentation time of a stimulus. Drift correction is depicted as a gray screen with a small black circle.

Procedure

The experiment took place in the Neurobiopsychology Laboratory at the University of Osnabrück. During the initial questionnaire, subjects were seated in a normally illuminated room. For the recordings and calibration, the room was darkened to achieve better tracking results. For the entire duration of the experiment, one experimenter remained in the room for potential questions and to oversee the experimental procedure.

Before the experiment began, participants were informed that the experiment consisted of three tasks, which were explained in detail, and that they were about to be presented with various web pages. They were told that, depending on the task, each stimulus would be followed by one or two questions for which the input would be given with the mouse. The subjects were then asked to complete a questionnaire in a web browser. In addition to simplified data analysis, this allowed the participants to get comfortable with the keyboard and mouse of the setup and to adapt to the web-browsing paradigm of the overall experiment.

After the questionnaire, the room was darkened, the eye tracker was set up, and a first calibration procedure was run. Before the actual experiment, a test run was performed to give the subjects experience with the different tasks. If there were no further questions, the eye tracker was recalibrated and the experiment was started as soon as the mean calibration error fell below 0.3°. After each task block, there was a short break during which the eye tracker could be removed. When the subjects were ready to continue, the eye tracker was set up and calibrated again.

The procedure during the three different tasks was given as follows (see Figure 3). Before presentation of each stimulus, subjects had to fixate a circle appearing at a randomized position on the screen. This was done to compensate for a bias toward the center of the screen (Tatler, 2007). If during this fixation the deviation between measured eye position and fixation circle was too large, the system was recalibrated. Small deviations were corrected by the system via automatic drift correction. After the fixation circle of the drift correction disappeared, the stimulus was presented. For analyses, the first fixation on a stimulus was discarded since it was defined by the location of the fixation mark prior to stimulus onset. Furthermore, only the subsequent 10 fixations on a stimulus were analyzed to ensure a comparable number of fixations on every stimulus. From this point on, the procedure differed from task to task as described below.

In the *free viewing* condition, a stimulus was shown for 6 s, disappeared, and was followed by a question page asking subjects to rate their experience with the shown web page on a scale of 1 to 5. Each step of the scale was explicitly described in a text on the same screen. This question was asked after each stimulus presentation in all conditions and was used to estimate subjects' prior knowledge of the stimuli. Of all familiarity ratings given, 90% were "never seen", and 8% were "seen the domain but not the specific web page". This clearly shows the success of the selection of web pages, which were expected to be unknown to subjects. In the *content awareness* task, the subjects could terminate stimulus presentation by a mouse click as soon as they felt sufficiently confident to answer the question regarding

the possible user group. This user input was introduced to allow greater interaction with the simulated web browser. Two question screens followed the presentation. First, the subjects were required to select one of five possible user groups. Then, they were asked to rate their previous experience with the page, as described above. The most complex task was the *information search* task. After a search term was displayed on a gray screen, a web page resembling the results page of a typical search engine was presented. Prior to the experiment, subjects were told that the quality of the results of the search engine had to be estimated. The subjects were able to select web pages from a list of results, displayed as "Result 1", "Result 2", and so on (Figure 2), by clicking on links. Again, subjects were able to stop the stimulus presentation with a mouse click whenever they felt confident about rating the result with regard to the current search term. Similar to the content awareness task, the subjects were required to provide two ratings on the following pages. However, instead of a more general question about a possible user group, the first question required the subjects to analyze the page content in more detail to rate the relevance of the web page with regard to the search term. The second question asked the subjects to rate their familiarity with the presented web page. After they provided this information, the search page was shown again; however, already visited links were grayed out and no longer able to be selected. After all five pages in the results had been visited, a new search term and a new results page were presented.

Analysis and results

Task influences on the spatial layout of fixations

The aim of the analysis is to characterize the causal routes through which different tasks influence viewing behavior. As a basis for this, a verification and investigation of task-dependent differences is required. A first check of the number of fixations made in the two self-terminated tasks showed no significant differences (information search mean = 35.6, std = 8.2 versus content awareness mean = 35.0, std = 9.1; $p > 0.1$ paired T-test). Due to the reduced viewing times in free viewing, the number of fixations was lower (mean = 24.0, std = 2.2). Only the first 10 fixations were used for further analysis, as 99% of all trials contained at least 10 fixations. More relevant differences in viewing behavior, however, should be observable in the spatial distribution of fixations; hence, the distribution was calculated for each task individually. Because the eye tracker does not record at pixel accuracy and neighboring pixels also influence fixations, 2D histograms of the fixations were smoothed

with a Gaussian kernel (FWHM = 1° = 36 px). Next the integral of the distribution was normalized to one. This resulted in a fixation density map (Figure 4) representing the probability of fixating a location on the screen. The difference between the fixation density maps of the individual tasks was assessed by a symmetric Kullback–Leibler divergence:

$$D_{\mathrm{KL}}(P, Q) = D_{\mathrm{KL}}(P \| Q) + D_{\mathrm{KL}}(Q \| P), \tag{1}$$

$$D_{\mathrm{KL}}(P \| Q) = \sum_i P(i) * \log\left(\frac{P(i)}{Q(i)}\right). \tag{2}$$

This value, unlike the regular KL divergence, can be interpreted as a distance metric between probability distributions. Large values indicate that the distributions are different. To examine whether the differences in fixation distributions in the three tasks were significant, a bootstrapping analysis was applied. This was done for all pair-wise combinations of tasks. All fixation sequences made by subjects in the two tasks to be compared were pooled. Two samples, corresponding in size to the number of fixation sequences made in the two tasks, were drawn randomly with replacement from this pool. For both samples, a fixation density map was constructed in the way described above, and the symmetric Kullback–Leibler divergence between the two maps was computed. This process was repeated 5000 times, resulting in a distribution of KL divergences. If the KL divergence value for comparison of the original fixation density maps was within the highest 1% of values in this distribution, the null hypothesis of no systematic difference between the two distributions was rejected.

Differences in the fixation density maps are clearly visible (Figure 4a). Whereas free viewing and the content awareness task have rather similar distributions, the fixation distribution for the information search task has a different appearance. Statistical analysis confirmed systematic differences between the information search task and the content awareness task ($D_{\mathrm{KL}} = 0.395$, $p < 0.01$), as well as between the information search task and free viewing ($D_{\mathrm{KL}} = 0.367$, $p < 0.01$). No differences between free viewing and the content awareness task could be found ($D_{\mathrm{KL}} = 0.194$, ns). Since all web sites were shown in each task, leaving image statistics and thereby bottom-up influences identical, the cause of these differences must originate from the different tasks.

Feature values at fixation points

According to the weak top-down hypothesis, task-dependent differences in viewing behavior are caused by different feature weightings in the bottom-up hierarchy. These changes should be reflected in the correlation
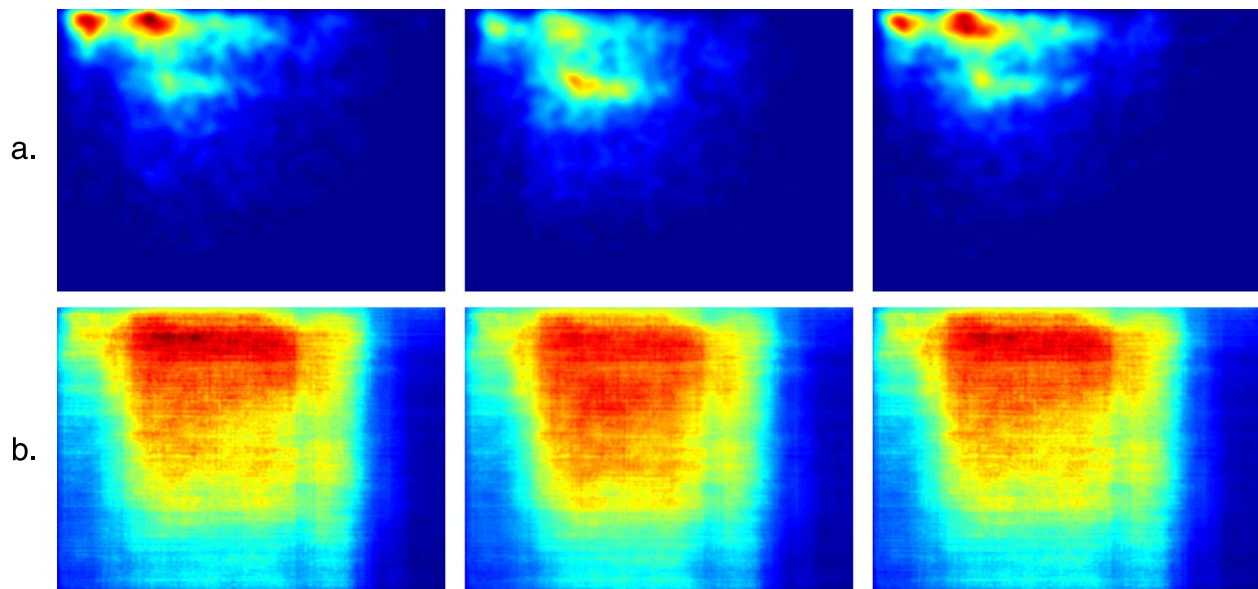
Figure 4. (a) True fixation density maps for the three conditions. Left to right: free viewing, information search task, and content awareness task. While the maps for content awareness task and free viewing look very similar, viewing behavior in the information search task was different. This impression was confirmed by bootstrapping analysis. (b) Salience density maps predicted by saliency models trained on data from free viewing, information search task, and content awareness task. The slight differences present are not sufficient to explain differences in the fixation density maps.

between image features and fixation points. Consequently, the influence of the task on the correlation between bottom-up image features and viewing behavior was investigated.

A common approach is to compare feature values at actual fixations and corresponding control fixations (Acik, Onat, Schumann, Einhäuser, & König, 2009; Einhäuser, Kurse, Hoffmann, & König, 2006; Frey, König, & Einhäuser, 2007; Kienzle, Wichmann, Schölkopf, & Franz, 2007; Tatler, Baddeley, & Gilchrist, 2005; Zhang, Tong, Marks, Shan, & Cotrell, 2008). In the current study, the area under the ROC curve (AUC) was used as a measure of influence of a certain feature on fixation selection. See Tatler et al. (2005) for a discussion of why the AUC is a good measure for quantifying the effect of features on salience. The AUC was calculated through linear interpolation over the ROC curve, which was obtained by separating feature values at control fixations from those at actual fixations. As controls, all fixations made by the same subject in the same task, but on other images than the one in question, were used. Essentially this is needed to avoid the introduction of artificial correlations of fixations with feature values due only to the spatial bias of fixations (Tatler et al., 2005; Zhang et al., 2008). In detail, the effects of eight different features on three to five spatial frequencies were examined: first- and second-order red–green and blue–yellow contrasts, luminance contrast, texture contrast, saturation, and edges (Sobel). Details of feature calculations can be found in Appendix A. The correlation between image features and

viewing behavior was captured by the AUC measure individually for each subject.

Comparing the ability of individual features to distinguish between actual and control fixations revealed that luminance contrast at the highest examined spatial frequency (kernel size 1° of visual angle) scored the highest AUC values, approximately 0.67, in all tasks. This was followed by Sobel, computed on a version of the stimulus downscaled to one quarter of the original width and height (see Appendix A for an explanation), and luminance contrast within a 2° kernel. In the case of contrast features, first-order contrasts gave higher values than second-order contrasts, and high spatial frequencies gave higher values than lower ones. For saturation, lower spatial frequencies work better, but the AUC is much lower than for the contrast features (see Figure 5). Overall, 24 of 27 features investigated showed a correlation with viewing behavior significantly different from chance in each task ($p < 0.05$, uncorrected).

The effect of the task variable on feature-fixation correlations was analyzed with a two-factorial analysis of variance (ANOVA, feature $\times$ task). In this analysis, weak top-down would be evidenced by an interaction of task and feature. Subjects were divided into 16 groups of 3 such that in each group, every image was seen once in each task. The mean AUC value of each group for every feature–task combination was used for analysis. Assumptions of an ANOVA were validated with a Lilliefors test ($p > 0.1$, Bonferroni corrected), visual inspection of qq plots, and Bartlett's test for homoscedasticity ($p > 0.1$).
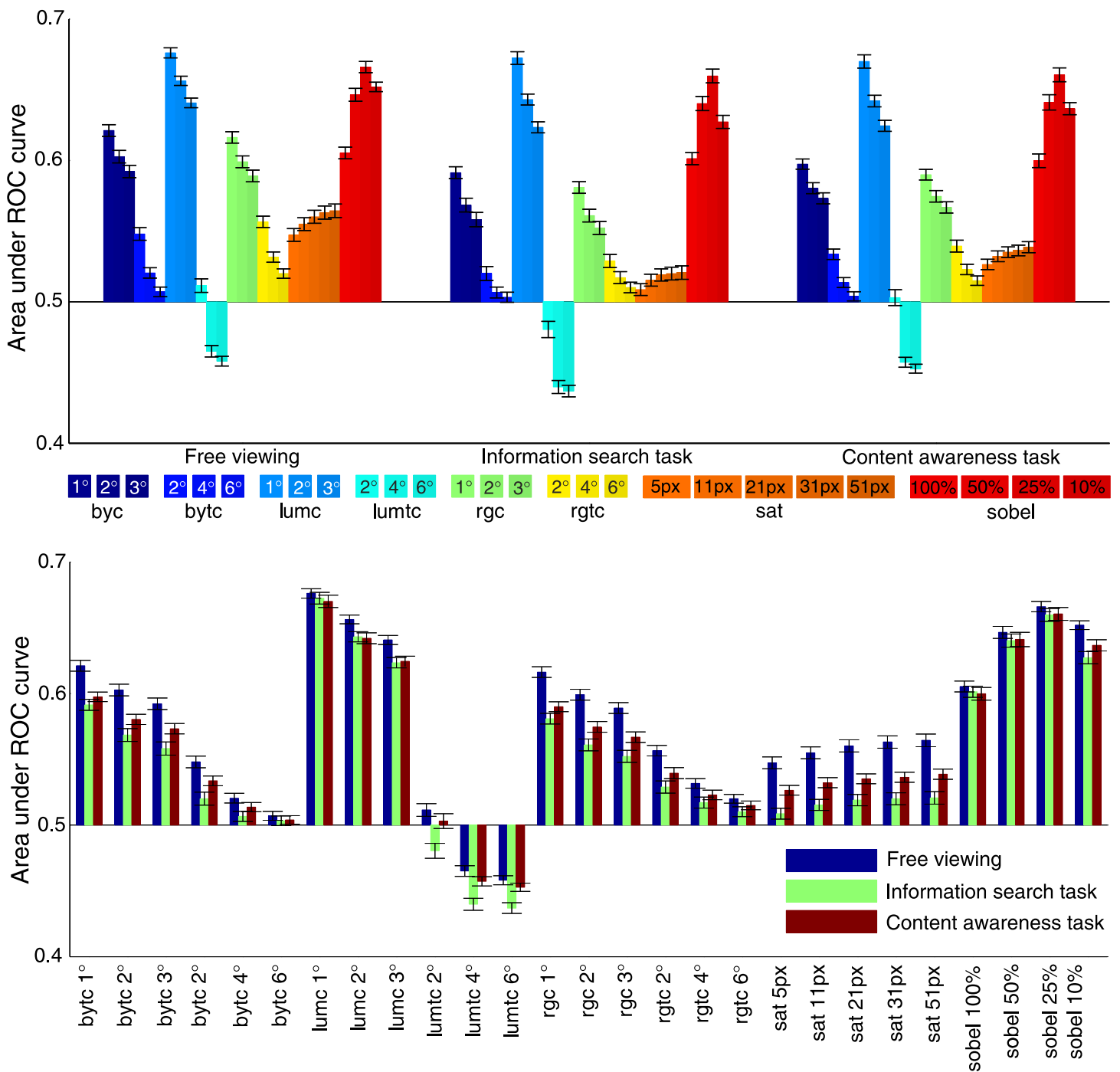
Figure 5. Mean area under the ROC curve for different features and conditions. Error bars represent *SEM*. Values in the upper figure are grouped by task, showing that the relative importance of individual features was almost, but not completely, constant across tasks. Values in the lower figure are grouped by feature, showing that for most features, values were highest in free viewing, lower in the content awareness task, and lowest in the information search task.

Both factors had a significant effect on feature-fixation correlations ($p < 0.001$). Additionally, a significant interaction effect between feature and task was found ($p < 0.001$). This means that AUC values for the different features were significantly different but also that those values differed with task. The interaction effect is in line with the weak top-down hypothesis as it shows that individual features are correlated with fixations differently across tasks. Overall, free viewing gave the largest AUC values, and for most features, the AUC values of the content awareness task were higher than those of the information search task (Figure 5). Post hoc analysis showed that all pair-wise comparisons between tasks were significant (two-way ANOVA, tasks and features as

factors, $p < 0.001$ for all main effects; $p < 0.01$ for all interaction effects).

These data show that each task has two distinct influences on the correlation between fixations and image features. First, AUC values drop with task complexity. Second, the patterns of feature-fixation correlations change with different tasks, consistent with the weak top-down hypothesis.

As a consequence, the question arises whether weak top-down can explain the differences found in viewing behavior. This question becomes especially pertinent when considering that the interaction effect can explain only a small amount of variance in AUC values ($\omega^2_{\text{interaction}} = 0.6\%$, compared with $\omega^2_{\text{feature}} = 91.3\%$ and $\omega^2_{\text{task}} = 2.7\%$).

Time course analysis

The above analysis averaged over time. Because of this, it could be argued that a potential effect of weak top-down might be missed if it was only present in the initial phase of the trials. In order to address this issue, a time-resolved analysis was carried out. For this, AUC values were calculated for each fixation index (all nth fixations). Thus, a third factor, fixation index, could be added to the ANOVA described above. Since weak top-down is revealed in the interaction effect of feature and task, a temporal effect on weak top-down would be visible in a three-fold interaction of feature, task, and fixation index. This, however, was not the case ($p > 0.5$). All factors individually and pair-wise interactions were significant ($p < 0.001$). Although this implies that the distribution of AUC values changes over time, these changes are similar in all tasks. Because this shows that time is not a relevant factor for weak top-down, it was omitted from the following analyses.

Can weak top-down explain the differences in viewing behavior?

In the following, it was investigated whether the small but significant changes in feature correlations over tasks explain at least part of the differences in viewing behavior described in the previous section. This question can be answered by using the bottom-up features as predictors for fixations. If the found differences in feature-fixation correlations do account for the differences in viewing behavior, then the predictions of models that are trained on data from the different tasks should reproduce these differences.

To address this issue, a simple linear model of visual attention was designed, consisting of a weighted sum of all 27 features evaluated in the previous section. The assumption of a linear model is supported by previous studies investigating the interaction of color contrast and luminance contrast (Engmann et al., 2009) and visual

stimuli and auditory stimuli (Onat, Libertus, & König, 2007). Vincent, Troscianko, and Gilchrist (2007) have used a similar model for separating targets from non-targets and found good separability of the two classes, measured with AUC values. However, their model was not trained on eye-tracking data nor directly compared with such data. In the model presented here, weights were trained by multiple linear regression on the eye-tracking data, using the z-scored feature maps as predictors and fixation density maps of individual images (computed as described above) as the response variable. For computational efficiency, all feature maps were resized to a resolution of 231×318 pixels by bicubic interpolation. Training was conducted separately for data from each task. Comparable to the AUC values, the pattern of weights in the combined regression was not identical in all tasks but still showed a rather consistent pattern. Some features with rather large AUC values were assigned negative weights. This can be explained by strong correlations between individual features. Multiple linear regression accounts for this correlation and does not simply assign positive weights to all features positively correlated with the fixation density maps.

After these three different models had been generated, saliency maps for all images in all three tasks were computed. Since the models were trained on the same data now being used to predict, any difference in fixation behavior that can be modeled with the 27 features should be visible in the predictions. Analogous to computing global fixation density maps across all images for each task, the saliency maps of all images were summed. These "salience density maps" can be seen in Figure 4b.

If weak top-down can provide a sufficient explanation for task-dependent viewing behavior, these salience density maps should exhibit differences similar to those found in the empirical fixation density maps (Figure 4a).

Similar to the fixation density maps, differences in model predictions were estimated with the symmetric KL divergence. Results showed that differences between the salience density maps were orders of magnitude smaller than between the fixation density maps (see Table 1). This shows that the model cannot explain the differences in viewing behavior.

Even though the differences between predicted salience density maps were much smaller than between fixation density maps, it might be possible to exploit them to produce task-dependent fixation patterns. This, however, can be accomplished only if the predicted differences

	FV vs. IS	FV vs. CA	IS vs. CA
Empirical data difference	0.368	0.194	0.396
Model prediction difference	0.007	0.001	0.005

Table 1. KL divergences of empirical fixation density maps and predicted salience density maps compared between tasks. *Note*: FV: free viewing, CA: content awareness, IS: information search.

Condition of training data:
- ■ Free viewing
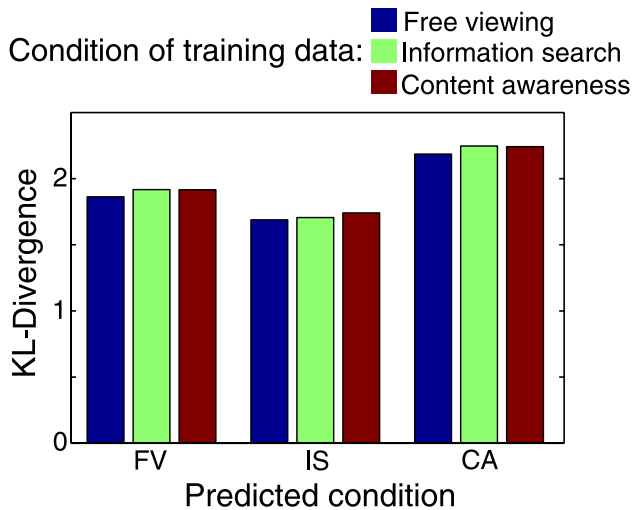- ■ Information search
- ■ Content awareness

Figure 6. KL divergences between salience density maps and fixation density maps. Training with free viewing data predicts each condition best (the blue bar is always the smallest). Additionally, the information search condition is best predicted independent of training data (IS prediction bars are the smallest).

align with the differences in the empirical data. To evaluate this, KL divergences between salience density maps and fixation density maps of all tasks were computed (Figure 6).

If differences in the models could explain differences in viewing behavior, viewing behavior in a task would best be predicted by training with data from the same task and vice versa (i.e., training with data from a certain task gives the overall best prediction for this task). Neither is the case, indicating that the found differences were not aligned with the differences in the empirical data.

The inability of the model to predict the task differences in viewing behavior might be due to poor predictive power of the model in general. In order to rule out this possibility, an analysis that allows a comparison of the model's performance with the attention-modeling literature was conducted. For the required cross-validation, the 90 images and 48 subjects were randomly divided into a training set and a test set at a ratio of 2:1 (60 images and 32 subjects for training, 30 images and 16 subjects for testing). This was done 200 times, with performance calculated as AUC for separating actual fixations from controls. The latter were defined as fixated points of all test subjects on all other test images. The mean AUC value for models predicting viewing behavior in the task on which they were trained was 0.67 (free viewing 0.68; information search 0.67; content awareness 0.67). This increased to 0.80 if random control locations were used. Table 2 shows that this performance is on par with more elaborate models in the literature. Note that the results of these models were based on natural stimuli. To the authors' knowledge, the present study is the first to address modeling of gaze movements on web pages, which unfortunately prevents a more direct comparison.

These data suggest that the inability to account for the task differences in a pure bottom-up model is not due to a general weakness of the model or poor selection of features.

A different approach to evaluating model performance was described in Harel, Koch, and Perona (2007). Instead of pure AUC values, the authors use a performance ratio of the model AUC and an intersubject AUC. The intersubject AUC expresses how well one set of subjects can be used to predict the fixations of another set. In particular, Harel et al. predicted the viewing behavior of one subject from two others. As control fixations, they used random locations. Here, the intersubject evaluation process was repeated 200 times with random assignments of subjects to predicting and predicted groups. In order to keep results comparable to Harel et al., only 2 subjects predicted one other subject. The resulting mean intersubject AUC was similar in the different tasks (free viewing 0.83, information search 0.83, content awareness 0.84, $p > 0.1$, Mann–Whitney U-test). These values are markedly higher than those reported in Harel et al. for natural stimuli, where AUC values range from 0.62 to 0.76 (taken from Figure 3 in their paper). Still, the current model reaches a comparable performance ratio of 0.96. Harel et al. report 0.98 for their best model, 0.84 for Itti and Koch's (2000) model, and 0.84 for Bruce and Tsotsos (2006). The difference in AUC values but rather similar performance ratio can be explained by the higher intersubject variability on natural stimuli than on web pages. It should be noted that using only a small amount of subjects for either predicting or being predicted introduces high variability in the estimate of the intersubject AUC. Additionally, larger numbers of subjects should increase the predictability and therefore intersubject AUC estimates. Therefore, high ratio values between model AUC and intersubject AUC based on only two subjects do not indicate near perfect model performance.

Clearly, these data do not imply that the linear regression modeling approach is superior to the other presented models; rather, they show that the current bottom-up model is capable of capturing essential aspects

Model name	Source	AUC
Itti and Koch	Einhäuser et al. (2006)	0.57
Bruce and Tsotsos	Bruce and Tsotsos (2006)	0.61
Graph-based saliency	Harel et al. (2007)	0.63–0.67
SUN: DoG filters	Zhang et al. (2008)	0.66
SUN: ICA filters	Zhang et al. (2008)	0.67
RBF network	Kienzle et al. (2009)	0.64
LR model (free viewing)	Current study	0.68

Table 2. Comparison of model performance (AUC) with other published results. Importantly, all other models use free viewing natural images and not web pages.

Betz, Kietzmann, Wilming, & König

of viewing behavior on web pages and that the evidence for strong top-down influence is not a side effect of below-average model performance.

Could reading constrain the results?

If most of the fixations in the current paradigm were caused by reading, the analyses and the interpretation of results would face three problems. First, it was suggested in the literature that scene perception and reading could employ different viewing strategies (Rayner, 1998). Hence, if most of the recorded data would contain reading fixations, the scope of the current findings would be limited. Second, salience-based models, as used in the analyses, might not be suited for textual input. Third, effects of weak top-down might be masked if reading was the most prominent mode of exploration. To address these issues, global saccade statistics were compared to the ones observed previously for scene perception and reading. Rayner (1998) reports a mean saccade size of 2° for silent reading and 4° for scene perception. Matching the statistics of scene perception, the mean saccade size of the recorded data is 4°. Additionally, during reading the reported amount of leftward saccades was estimated to be 10–15%. In contrast to these data in the present study, 52% of the 41,765 saccades analyzed in this study were leftward. This matches other data on natural stimuli not containing text available in the authors' laboratory, where 51.3% of 31,785 saccades were from right to left. In conclusion, reading plays at most a minor role in the selection of fixation points.

Discussion

The present eye-tracking study investigated the influence of different tasks on viewing behavior. It aimed to provide deeper understanding of the two causal routes: weak and strong top-down. Weak top-down relies on changes in the relative emphasis of the different features in the bottom-up system. In contrast, strong top-down predicts task-dependent differences in the spatial layout of fixations not caused by changes in the bottom-up system.

To this end, eye movements of subjects were recorded while they viewed web pages in three different tasks. In addition to an analysis of the spatial layout of fixations, ROC methods were applied to investigate the correlation between image features and fixations. To investigate whether weak top-down can be a sufficient cause of different viewing behavior in the tasks, a linear model was trained to predict fixation density maps of each task.

Significant task-dependent differences in the spatial layout of fixations were found. Interestingly, not all tasks elicited specific viewing behavior, which might explain

the conflicting results of Pan et al. (2004) and Rayner (1998). The differences in viewing behavior between free viewing and information search as well as between content awareness and information search are in accordance with Rayner (1998). The fact that no differences between free viewing and the content awareness task were found might explain why Pan et al. (2004) found no influence of the task variable in their study. In that experiment, half the participants were asked to remember the content of web sites; the other half were told just to view the web site. Clearly, the task "remember the content" is conceptually closer to our content awareness task than to the information search task.

Deviant from Tatler (2007), subjects in this study did not show a pronounced central viewing bias, but a tendency to fixate on the top left quadrant of the stimuli. This was not investigated further, since it was not a focus of the current study, but it is an interesting topic for future investigation. The potential importance of the spatial bias for guiding viewing behavior will be returned to further below.

The found differences in viewing behavior prepared the ground for the exploration of the weak and strong top-down hypotheses. For this, a more detailed analysis of feature-fixation correlations was performed. The ROC analyses showed that feature-fixation correlations existed independent of the spatial bias of fixations. These were slightly higher than in comparable studies (Acik et al., 2009; Tatler et al., 2005). In total, 24 of 27 features showed significant correlation with viewing behavior. It should be noted that these correlations on their own do not imply that image features have a causal influence on viewing behavior.

Consistent with a contribution of weak top-down, an ANOVA with task and feature as factors revealed an interaction effect; that is, the relative emphasis of bottom-up features changed between tasks. However, the variance explained by this interaction was very small ($\omega^2_{\text{interaction}} = 0.6\%$). This leads to the question whether these differences can account for the observed differences in viewing behavior, or whether strong top-down is needed to explain the effects.

To answer this question, linear models of the fixation data from each task were trained. If weak top-down is sufficient, the predictions of models trained on data from the different tasks should differ analogously with the observed differences in viewing behavior; however, this was not supported by the modeling results. To begin with, differences between the modeled salience density maps were orders of magnitude smaller than the observed differences in viewing behavior. Furthermore, model differences did not reproduce the pattern of differences in the eye-tracking data. Importantly, the failure of the model to capture task-dependent differences through bottom-up feature modulations cannot be attributed to a general weakness of the approach. Overall, the current results (AUC of 0.67 based on image features and 0.80

with spatial information included) are comparable to more complex and generally accepted attention models in the literature. In addition, the good performance shows that bottom-up modeling is generally applicable to human subjects viewing web pages.

Taken together, the current evidence favors strong top-down as the most influential causal route through which a task can guide viewing behavior; however, some arguments need to be considered before discounting the contribution of top-down modulation of a bottom-up mediated salience (weak top-down).

One might argue that weak top-down was the only cause of changes in viewing behavior in each individual subject, but the changes were not consistent across subjects and therefore average out in the conducted analysis. This would be indicated by high variance of AUC values for individual features across subjects (i.e., high within-cell variances in the ANOVA). This, however, is not the case since 94.6% of the variance in the data could be explained by the two factors and the interaction in the ANOVA model. Moreover, it is not clear why different weak top-down effects across subjects would lead to similar changes in viewing behavior that do not average out.

Another consideration is that not all low-level image features relevant to the bottom-up system were analyzed. One might argue that some low-level feature exists that does receive different weights and thereby explains much of the differences between tasks. This can in principle not be excluded; nevertheless, the selection of features in this study was based on existing models that have been successfully used in predicting visual attention. Moreover, it should be noted that many features are correlated with each other. Thus, the large set of features used in the current work is expected to act as a diagnostic probe for other low-level features.

Finally, high-level features might exist that are able to change viewing behavior in a task-dependent manner. In this case, the observed small changes in feature weights might be the result of changes in weights for higher level features. These would have to be slightly correlated with the features investigated, but not so strongly that the low-level features can already explain the changes in viewing behavior. Such higher level features might be specific object detectors ("menu bar", "heading"), or other (non-linear) compounds of low-level features. Some examples of such features are reported in the literature. Cerf, Harel, Einhäuser, and Koch (2008) found that faces attract more attention than predicted by a bottom-up model. Kayser, Nielson, and Logothetis (2006) found that uninformative image regions are avoided. Additionally, Elazary and Itti (2008) showed that interesting objects correlate with visual salience. For two reasons, this does not pose a difficulty to the strong top-down hypothesis. First, the used low-level features might still act as diagnostic probes, as described above. Second, it is questionable whether these features are still a part of the low-level

system because they already possess some semantic content. Therefore, if these features really are the cause of the observed task effects, it is debatable whether this should be classified as weak top-down. Similarly, some tasks, as for instance *information search,* might require the subjects to put more emphasis on the contained text. If text was interpreted as a feature, this could qualify as weak top-down. This, however, should be visible in the examined Sobel features, which correlate highly with text and therefore act as diagnostic probe. In short, if participants read more text in some tasks, the Sobel AUC should be more pronounced as compared to the other features. This is not the case.

Generally, weak top-down requires a causal influence of low-level features on visual attention. This assumption has been challenged previously by Einhäuser and König (2003). Henderson, Brockmole, Castelhano, and Mack (2007) pursued a similar point, arguing that it is not visual salience that predominantly guides attention but cognitive factors.

In summary, the current data indicate that strong top-down is the dominant mode to adapt the allocation of attention to varying task demands. In the present experiment, no strong evidence for a modulation of bottom-up processing by top-down signals (weak top-down) could be found. At best, it may make a small contribution. This picture might be different in cases where the task specifically asks for attention to a certain feature (e.g., look for the red ball). Hence, it should not be concluded from the current study that strong top-down is the only mode of adapting viewing behavior to task demands. Rather, it was shown that strong top-down is an effective means even in the current tasks, which were not trivially spatial (compared to the hypothetical task "look for the target on the right"): For instance, the content awareness task could be solved by fixations on a multitude of locations. In addition, if the tasks were trivially spatial, higher intersubject AUC values should be expected.

An open question that remains to be investigated in future work is how exactly strong top-down can influence viewing behavior. One possibility would be the introduction of a spatial bias. This bias could emphasize certain parts of the stimuli (e.g., the top-left area, which typically contains the logo or title of a page). Since subjects do not fixate identical points on different web pages, even within a task, a pure spatial bias does not suffice. The importance of bottom-up features is evident in the fact that the AUC measures, as presented in the current work, use spatially biased controls and still display above-chance performance. A potential instantiation of this view was introduced by Torralba et al. (2006), who use a spatial bias toward task-relevant regions to modulate bottom-up saliency.

A further potentially informative aspect in the current study is that the information search task differs from the other two in that it provides prior information to subjects that might be exploited online to evaluate incoming information and, on these grounds, to alter fixation

selection. This type of directed information sampling has been suggested as an important factor for attention (Land & Hayhoe, 2001; Triesch et al., 2003). Web pages provide a semantically rich environment and are therefore well suited for further investigations of this phenomenon. However, it will also be crucial for future work to show in how far strong top-down is a modulating factor of attention while viewing natural scenes. Its existence on web pages makes this a promising endeavor.

Appendix A

Feature calculations

All feature values were calculated in the Derrington–Krauskopf–Lennie (DKL) color space (Derrington, Krauskopf, & Lennie 1984), based on cone excitation measurements in the primate. This space is spanned by three perpendicular axes: the L + M axis is commonly referred to as the luminance axis; along it, the firing ratio of the L and M cones stays constant and luminance increases. Along the L − M axis (red–green axis), the firing of L and M cones changes in such a way that their sum stays constant. This way, luminance is not affected, but color changes from a bluish green to red. Along the S − (L + M) axis (blue–yellow axis), only the excitation of the S cones increases, causing a change in color from a purplish blue to a greenish yellow (Gegenfurtner & Kieper, 2003).

Luminance contrast (lumc) of the region around a fixation was defined as the standard deviation of luminance values (L + M axis in DKL space) in an isotropic patch centered on the fixation, normalized by the mean luminance of the image:

$$LC = \frac{(I^2 * G) - (I * G)}{\bar{I}}, \tag{A1}$$

where

$$G = e^{-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)}, \tag{A2}$$

and I is the matrix of luminance values.

The parameter σ allows for selecting contrasts at different spatial frequencies. It was chosen such that the full width at half maximum (FWHM) of the Gaussian kernel corresponded to 1°, 2°, and 3° of visual angle.

Red–green (rgc) and blue–yellow (byc) color contrasts were defined accordingly for the two color axes in DKL space, but normalization by mean intensity of the respective channel was omitted.

On these three contrast maps, second-order contrasts were calculated using the same method, but doubling the kernel sizes to 2°, 4°, and 6° FWHM. This yielded red–green (rgtc) and blue–yellow (bytc) and luminance (lumtc) texture contrasts.

Additionally, edge values on the luminance channel of the image were computed using Sobel operators as follows:

$$Sob = \sqrt{G_x^2 + G_y^2}, \tag{A3}$$

where

$$G_x = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} * ([+1 \quad 0 \quad -1] * I), \tag{A4}$$

and

$$G_y = \begin{bmatrix} +1 \\ 0 \\ -1 \end{bmatrix} * ([1 \quad 2 \quad 1] * I). \tag{A5}$$

To obtain responses for edges at different spatial frequencies, this Sobel routine was applied to different resized versions of the image's luminance channel, using bicubic interpolation. The resize factors used were 1 (no change in the image), 0.5, 0.25, and 0.1. Finally, the edge maps were smoothed with a small Gaussian kernel (3 pixels FWHM), since the eye tracker does not have pixel accuracy.

Saturation was calculated as the Euclidian distance in DKL color space between a pixel's color value and the center of the isoluminant plane:

$$SAT = \sqrt{RG^2 + BY^2}. \tag{A6}$$

This saturation map was then smoothed with Gaussian kernels of different sizes to obtain weighted mean saturation at different spatial resolutions. Kernel sizes were chosen to cover a wide range of values (FWHM 5 px, 11 px, 21 px, 31 px, 51 px).

Acknowledgments

Author contributions: All authors contributed equally to this work.
Commercial relationships: none.
Corresponding authors: Torsten Betz, Tim C. Kietzmann, Niklas Wilming, Peter König.

Emails: tbetz@uos.de, tkietzma@uos.de, nwilming@uos.de, pkoenig@uos.de.
Address: Institut für Kognitionswissenschaft, Albrecht-strasse 28, 49069 Osnabrück, Germany.

References

Acik, A., Onat, S., Schumann, F., Einhäuser, W., & König, P. (2009). Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories. *Vision Research, 49,* 1541–1553. [PubMed]

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387,* 401–406. [PubMed]

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research, 46,* 2824–2833. [PubMed]

Bruce, N., & Tsotsos, J. (2006). Saliency based on information maximization. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in neural information processing systems* (vol. 18, pp. 155–162). Cambridge, MA: MIT Press. [Article]

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9*(3):6, 1–15, http://journalofvision.org/9/3/6/, doi:10.1167/9.3.6. [PubMed] [Article]

Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (2008). Predicting human gaze using low-level saliency combined with face-detection. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems* (vol. 20, pp. 241–248). Cambridge, MA: MIT Press. [Article]

Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of Physiology, 357,* 241–265. [PubMed] [Article]

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., Kurse, W., Hoffmann, K. P., & König, P. (2006). Differences of monkey and human overt attention under natural conditions. *Vision Research, 46,* 1194–1209. [PubMed]

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision, 8*(3):3, 1–15, http://journalofvision.org/8/3/3/, doi:10.1167/8.3.3. [PubMed] [Article]

Engmann, S., 't Hart, B. M., Sieren, T., Onat, S., König, P., & Einhäuser, W. (2009). Saliency on a natural scene background: Effects of color and luminance-contrast add linearly. *Attention, Perception & Psychophysics, 71,* 1337–1352. [PubMed]

Frey, H., König, P., & Einhäuser, W. (2007). The role of first- and second-order stimulus features for human overt attention. *Perception & Psychophysics, 69,* 153–161. [PubMed] [Article]

Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annual Review of Neuroscience, 26,* 181–206. [PubMed]

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (vol. 19, pp. 545–552). Cambridge, MA: MIT Press. [Article]

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*(1):6, 49–63, http://journalofvision.org/3/1/6/, doi:10.1167/3.1.6. [PubMed] [Article]

Henderson, J. M., Brockmole, J. R., Castelhano, 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 movements: A window on mind and brain* (pp. 537–562). Oxford, UK: Elsevier.

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

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

Kayser, C., Nielsen, K. J., & Logothetis, N. K. (2006). Fixations in natural scenes: Interaction of image structure and image content. *Vision Research, 46,* 2535–2545. [PubMed]

Kienzle, W., Franz, M. O., Schölkopf, B., & Wichmann, F. A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision, 9*(5):7, 1–15, http://journalofvision.org/9/5/7/, doi:10.1167/9.5.7. [PubMed] [Article]

Kienzle, W., Wichmann, F. A., Schölkopf, B., & Franz, M. O. (2007). A nonparametric approach to bottom-up visual saliency. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (vol. 19, pp. 689–696). Cambridge, MA: MIT Press. [Article]

Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review of Neuroscience, 30,* 57–78. [PubMed]

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227. [PubMed]

Land, M., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41,* 3559–3566. [PubMed]

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

Nelson, J. D., Cottrell, G. W., Movellan, J. R., & Sereno, M. I. (2004). Yarbus lives: A foveated exploration of how task influences saccadic eye movement [Abstract]. *Journal of Vision, 4*(8):741, 741a, http://journalofvision.org/4/8/741/, doi:10.1167/4.8.741.

Onat, S., Libertus, K., & König, P. (2007). Integrating audiovisual information for the control of overt attention. *Journal of Vision, 7*(10):11, 1–16, http://journalofvision.org/7/10/11/, doi:10.1167/7.10.11. [PubMed] [Article]

Pan, B., Hembrooke, H. A., Gay, G. K., Granka, L. A., Feusner, M. K., & Newman, J. K. (2004). The determinants of web page viewing behavior: An eye-tracking study. In A. Duchowski (Ed.), *Proceedings of eye tracking research & applications symposium (ETRA 04)* (pp. 147–154). New York: ACM. [Article]

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. [PubMed]

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network, 10,* 341–350. [PubMed]

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision, 7*(14):16, 1–20, http://journalofvision.org/7/14/16/, doi:10.1167/7.14.16. [PubMed] [Article]

Tatler, B. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14):4, 1–17, http://journalofvision.org/7/14/4/, doi:10.1167/7.14.4. [PubMed] [Article]

Tatler, B. W., Baddeley, J. W., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research, 45,* 643–659. [PubMed]

't Hart, M. B., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., Konig, P., et al. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition, 17,* 1132–1158.

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]

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision, 3*(1):9, 86–94, http://journalofvision.org/3/1/9/, doi:10.1167/3.1.9. [PubMed] [Article]

Vincent, B. T., Troscianko, T., & Gilchrist, I. D. (2007). Investigating a space-variant weighted salience account of visual selection. *Vision Research, 47,* 1809–1820. [PubMed]

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum Press.

Zhang, L., Tong, M. H., Marks, T. K., Shan, H., & Cottrell, G. W. (2008). SUN: A Bayesian framework for saliency using natural statistics. *Journal of Vision, 8*(7):32, 1–20, http://journalofvision.org/8/7/32/, doi:10.1167/8.7.32. [PubMed] [Article]