It depends on *when* you look at it: Salience influences eye movements in natural scene viewing and search early in time

Nicola C. Anderson  
Department of Cognitive Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

Eduard Ort  
Department of Cognitive Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

Wouter Kruijne  
Department of Cognitive Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

Martijn Meeter  
Department of Cognitive Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

Mieke Donk  
Department of Cognitive Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

It is generally accepted that salience affects eye movements in simple artificially created search displays. However, no such consensus exists for eye movements in natural scenes, with several reports arguing that it is mostly high-level cognitive factors that control oculomotor behavior in natural scenes. Here, we manipulate the salience distribution across images by decreasing or increasing the contrast in a gradient across the image. We recorded eye movements in an encoding task (Experiment 1) and a visual search task (Experiment 2) and analyzed the relationship between the latency of fixations and subsequent saccade targeting throughout scene viewing. We find that short-latency first saccades are more likely to land on a region of the image with high salience than long-latency and subsequent saccades in both the encoding and visual search tasks. This implies that salience indeed influences oculomotor behavior in natural scenes, albeit on a different timescale than previously reported. We discuss our findings in relation to current theories of saccade control in natural scenes.

Introduction

One of the fundamental questions facing eye-movement research is the extent to which eye movements in natural scenes are under the control of bottom-up, image-based features and top-down, cognitive factors. The idea that bottom-up image features can influence selective attention in natural scenes arose predominantly from the development of models centered around the salience map (Koch & Ullman, 1985). The salience map is a retinotopic representation containing information concerning the relative conspicuity of individual features in the visual field, like distinct colors or orientations. Theories presuming a salience map (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998) assume that we attend to the location with the highest activation in this map. There is evidence that such models of salience can account for some of the variance in real human eye movements (Foulsham & Underwood, 2008; Itti & Koch, 2000; Parkhurst, Law, & Niebur, 2002; Peters, Iyer, Itti, & Koch, 2005; Underwood & Foulsham, 2006). Converging with this evidence, when scene statistics are directly compared with fixated locations, edge density and luminance contrast tend to be higher at fixated compared to nonfixated locations (Mannan, Ruddock, & Wooding, 1996; Reinagel & Zador, 1999; Tatler, Baddeley, & Gilchrist, 2005). However, even in the earliest discussion of the salience map model, Koch and Ullman (1985) stipulated that the salience map may be modulated by higher cortical centers in a top-down fashion.

Indeed, some of the earliest reports of eye movements in natural scenes observed a large role of top-
down control in the determination of fixated locations. Yarbus (1967) had an observer view a painting (The Unexpected Visitor of I. E. Repin) and showed that viewing behavior varied with instruction. For instance, the observer primarily fixated the faces of the depicted people when asked to give their ages, whereas this was not the case when the observer was asked to remember the positions of the people. More recently, an increasing number of studies have suggested that eye movements are influenced by a variety of top-down factors, such as scene gist (Oliva & Torralba, 2006), objects (Einhäuser, Spain, & Perona, 2008; Nuthmann & Henderson, 2010), domain knowledge and expertise (Reingold, Charness, Pomplun, & Stampe, 2001; Underwood, Foulsham, & Humphrey, 2009), and context and episodic memory (Allmann, 2004; Brockmole, Castelhano, & Henderson, 2006; Richardson & Spivey, 2000). Such pervasive evidence for top-down factors influencing eye-movement behavior in natural scenes has led to the question of whether there is truly a controlling role for salience in oculomotor control and attentional selection at all (Henderson, Brockmole, Castelhano, & Mack, 2007; Land, 2009; Tatler, Hayhoe, Land, & Ballard, 2011).

Perhaps the strongest evidence against a role for salience in oculomotor control arises from the fact that models of bottom-up salience can only account for a modest portion of the variance observed in the eye-movement behavior of observers viewing real-world pictures. That is, evaluations of the extent to which salience models predicted fixation locations have found only a modest relationship between fixated locations and the salience of those locations (e.g., Einhäuser, Spain et al., 2008; Tatler et al., 2005). Accordingly, correlations between image statistics and fixated locations tend to be small (Tatler, Baddeley, & Vincent, 2006). More importantly, demonstrating a correlation between salience and fixated locations does not imply that there is a causal relationship between the two. In particular, the relationship between fixated locations and image statistics may also be accounted for by top-down influences such as the presence of a central bias in scene viewing or a correspondence between objects and salience (Einhäuser, Spain et al., 2008; Nuthmann & Henderson, 2010; Tatler, 2007; for a comprehensive review, see Tatler et al., 2011). Finally, the relationship between bottom-up image salience and eye-movement behavior tends to disappear completely when participants are given a task goal or have special knowledge of the stimuli. For example, bottom-up salience effects are overridden in visual search for a specific target (Einhäuser, Rutishauser, & Koch, 2008; Henderson et al., 2007; Stirk & Underwood, 2007; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006), in encoding and remembering emotive stimuli (Humphrey, Underwood, & Lambert, 2012), and when stimuli are related to personal expertise (Humphrey & Underwood, 2009).

The idea that bottom-up stimulus information plays little role in natural scene viewing contrasts sharply with the strong bottom-up salience effects that arise in experiments employing simpler, more impoverished visual displays, even when participants are given a strong top-down goal (e.g., among many others, Godijn & Theeuwes, 2002; Hickey, van Zoest, & Theeuwes, 2010; Hunt, von Muhlenen, & Kingstone, 2007; Irwin, Colcombe, Kramer, & Hahn, 2000; Lamy & Zoaris, 2009; Siebold, van Zoest, & Donk, 2011; Theeuwes, 1994; van Zoest, Donk, & Theeuwes, 2004a, 2004b; Wolfe, 1994; Zehetleitner, Koch, Goschy, & Muller, 2013). Attention can be captured by irrelevant but salient distractors in color, shape, orientation, and luminance (Siebold et al., 2011; Theeuwes, De Vries, & Godijn, 2003; Turatto & Galfano, 2000; van Zoest et al., 2004a). For example, when asked to search for a small element inside a singleton with a unique dimension (either a unique shape or unique color, depending on the block), participants’ eye movements were captured when another item in the display happened to be unique in the task-irrelevant dimension. Thus, attention—along with the eyes—is readily captured by salient items (Theeuwes, 1992; Theeuwes et al., 2003). However, even the role of salience and the extent to which it influences goal-driven eye movements in such basic displays have been hotly debated (e.g., Folk & Remington, 1998; Theeuwes, 2004; van Zoest et al., 2004a). One emerging idea is that bottom-up stimulus-driven and top-down goal-driven processes both influence oculomotor behavior but operate on different time scales (van Zoest et al., 2004a). For example, in searching among singletons with differing salience, short-latency, first eye movements were consistently directed toward the most salient singleton, with longer latency and subsequent saccades being predominantly goal driven (Siebold et al., 2011).

Does the view that the effects of bottom-up salience and top-down factors are temporally dissociable generalize to natural scenes? This question has thus far yielded mixed answers. For example, early saccades have been found to have a higher salience value than late saccades, both during free viewing (Parkhurst et al., 2002) and in an encoding-retrieval task (Foulsham & Underwood, 2008). Conversely, Tatler et al. (2005) found no significant differences of salience, computed as edge density and contrast, across saccade number in a similar encoding task. Nystrom and Holmqvist (2008) found that elevated contrast values for early fixated locations only occurred in neutral images and not in ones with faces or emotional content. Rather than exploring salience at fixated locations, Einhäuser, Rutishauser, et al. (2008) manipulated images such that one side had higher contrast (and thus higher salience).
They compared the extent to which eye movements were biased toward the higher contrast region when participants were given different task demands. Both in free viewing and presumed bottom-up-driven “oddity search,” a bias in the first saccades for the high contrast side was found. This bias disappeared when participants searched for a presupplied bull’s-eye target.

True bottom-up salience effects are rarely observed in a natural scene visual search task, and they seem easily overridden by top-down goals (Einhäuser, Rutishauser et al., 2008; Henderson et al., 2007; Stirk & Underwood, 2007; Underwood et al., 2006). One potential reason is that it is difficult to disentangle salience effects from goal-driven effects in natural scenes, in particular when the target itself is highly salient. In the work by Einhäuser, Rutishauser et al. (2008), when the template target was located on the low-contrast side of the image and thus was likely more salient than its surroundings, on average less than one fixation was made before the target was found (Einhäuser, Rutishauser et al., 2008, table S3 in supplementary materials). This suggests that the target was acquired immediately after the first saccade into the image. Moreover, in contrast to simple artificial displays, natural scenes are typically heterogeneous, consisting of a multitude of objects and textures. This multitude leads to many somewhat salient locations but it prevents a few individual locations from becoming very salient. Accordingly, even if selection behavior in a natural image is partly influenced by salience, it would be hard to distinguish this influence from top-down-driven saccades to objects of interest that happen to also be salient. Additionally, it is possible that bottom-up salience effects dissipate with time irrespective of top-down control (Donk & van Zoest, 2008) and that the initial saccade latency in an image is simply too long for salience effects to arise. The average latency of the first eye movement in a natural scene is often in the range of around 330 ms (Henderson & Hollingworth, 1999), in which saccades in basic displays tend to already show strong goal-driven components (Siebold et al., 2011; van Zoest et al., 2004a). Salience effects may thus be present in natural scenes but emerge only at the shortest latency initial saccades.

Crucially, temporal effects of salience in natural scenes are generally only explored based on ordinal saccade number (Einhäuser, Rutishauser et al., 2008; Foulsham & Underwood, 2008; Nystrom & Holmqvist, 2008; Tatler et al., 2005). In research with artificial or basic displays, latency is treated as a critical component (e.g., Hickey et al., 2010; Siebold et al., 2011; van Zoest et al., 2004a). In such experiments, early short-latency saccades are more likely to be driven by the bottom-up featural information in the display. It is therefore possible that the mixed results in natural scenes lie in the millisecond-level time course of the deployment of a saccade. Indeed, Mackay, Cerf, and Koch (2012) looked at the time course of saccades throughout scene viewing. They demonstrated that saccadic targeting was different for short-latency saccades than long-latency saccades, with slower eye movements being more strongly directed toward faces in scenes. This effect increased over saccade number and was modulated by task demands. The authors argue that early saccades (initiated within 60–100 ms) are controlled by subcortical circuits (namely, the superior colliculus) and represent a distinct population from later and slower saccades. They suggest that the function of these early eye movements is to automatically and rapidly direct gaze to salient objects in a scene without the need for cortical analysis. However, it remains unclear from that work how these initial rapid eye movements into a scene are exactly targeted, as their comparison was against early saccades that landed on faces in the picture, and whether they were partially anticipatory saccades that were not targeted at all.

In the present work, we explore the time course of the scan paths created in looking at a real-world image. As in research with simple displays, bottom-up salience information may only play a role in influencing eye-movement behavior at the earliest moments of scene presentation. If this is the case, then short-latency, initial eye movements into a display, even in real-world images, are more likely to be directed toward regions of high salience, while long-latency and later saccades should be less influenced by salience. As much of the criticism surrounding the salience model concerns the correlational nature of comparing fixated locations to model-predicted values, we employ a contrast manipulation across the image (Einhäuser et al., 2006; Einhäuser, Rutishauser et al., 2008; Nystrom & Holmqvist, 2008). Furthermore, we investigate the possible influence of salience with and without a visual search task. In Experiment 1 we determine whether the saccade latency is related to the extent to which eye movements are directed toward a high- or low-contrast side of an image. We use an encoding and recognition task that has previously been shown to elicit early salience effects (Foulsham & Underwood, 2008). In Experiment 2 we investigate the same effect in a visual search task previously shown not to elicit early biases toward salient regions (Einhäuser, Rutishauser et al., 2008).

**Experiments 1a and 1b**

In Experiment 1 we manipulated the salience distribution across images by selectively manipulating the contrast in a gradient across the image. In other studies, this has usually implied selectively reducing the
Methods
Participants
Two separate groups of 12 participants each (ages 18–26 years, $M = 22$) were recruited from the Vrije Universiteit Amsterdam and participated in Experiments 1a and 1b, respectively, for course credit or 9 euros. All reported normal or corrected-to-normal vision and were unaware of the purpose of the experiment. The study was approved by the ethics board of the faculty of psychology and education and conducted according to the principles of the Declaration of Helsinki.

Apparatus
The experiments were designed and presented using OpenSesame (Mathot, Schreij, & Theeuwes, 2012), an open-source experiment programming environment integrated with the SR Research Eyelink 1000 tracking system (SR Research Ltd., Mississauga, Ontario, Canada). Stimuli were presented on a 22-in. (diagonal) Samsung Syncmaster 2233RZ with a resolution of 1680 × 1050 pixels and a refresh rate of 60 Hz at a viewing distance of 75 cm. Eye position was recorded via a nine-point calibration and validation procedure. The online saccade detector of the eye tracker was set to detect saccades with an amplitude of at least 0.5°, using an acceleration threshold of 9500°/s² and a velocity threshold of 35°/s. The experiments took place in a dim, sound-attenuated room. The experimenter received real-time feedback on system accuracy on a second monitor located in an adjacent room, and calibration and validation was repeated as needed.

Stimuli
Images were selected from the SUN2012 database (Xiao, Hays, Ehinger, Oliva, & Torralba, 2010) and from Learning to Predict Where Humans Look (Judd, Ehinger, Durand, & Torralba, 2009). The images depicted various exteriors, interiors, and natural scenes and were chosen such that they did not contain any obvious human faces or text. Each image had a native resolution of 1024 × 768 pixels and was converted to grayscale. Images were then further selected such that their mean pixel intensity values across the left and right side of the image conformed to a ratio of at most 3:4. Conforming with these selection criteria, 150 images were used from the SUN2012 database and 92 from Learning to Predict Where Humans Look. Stimuli were presented centrally on the monitor at their native resolution and subtended a visual angle of approximately 22° horizontal × 16° vertical. The rest of the screen surrounding the image was black.

Contrast adjustment
Image contrast adjustment was performed on all selected images using the imadjust function in MATLAB (version 7.12.0, R2011a; the MathWorks, Natick, MA). For Experiment 1a, pixel intensity values on one side of the image were linearly remapped to a range spanning 40% of the original range, in a way that left mean pixel intensity unaffected. Only the variance in pixel intensity was thus reduced. For each image, either the left or right 40% was manipulated in this fashion. For the center 20%, contrast was gradually reduced linearly from full to reduced contrast. Two versions of each image were created, with the contrast gradually reduced from left to right and from right to left (see Figure 1A).

The contrast increase for Experiment 1b was achieved with a similar procedure, by remapping those intensity values within 40% of the mean to the whole intensity range. Intensity values outside of this 40% range were remapped to black and white (see Figure 1C).

The monitor used to display the images use a gamma correction value of 2.2 and was therefore not linearized. See the Supplementary Material for the luminance response curve of the monitor used.

Contrast adjustment and salience
The contrast reduction and contrast increase manipulations were intended to decrease or increase the salience on one side of the image relative to the other. In order to confirm whether these manipulations were successful, salience maps were produced using both the salience model implemented in the Saliency Toolbox (Walther & Koch, 2006) and the AIM model (Bruce & Tsotsos, 2009). The mean salience values for the original and for the contrast-reduced and contrast-increased sides from the output of both models were compared. Note that these models rely on very different assumptions regarding visual features and calculate salience very differently. Nevertheless, they yielded comparable results; both reveal significant changes in
salience due to the contrast manipulation. For the Saliency Toolbox, compared to original salience values \((M = 0.03, \ SD = 0.009)\), reducing contrast indeed reduced salience \((M = 0.01, \ SD = 0.005)\), \(t(241) = 38.30, p < 0.001\), while increasing contrast increased salience \((M = 0.04, \ SD = 0.01)\), \(t(241) = 23.37, p < 0.001\). For the AIM model, compared to the original salience values \((M = 0.33, \ SD = 0.08)\), reducing contrast reduced salience \((M = 0.23, \ SD = 0.06)\), \(t(241) = 48.32, p < 0.001\), while increasing contrast increased salience \((M = 0.41, \ SD = 0.07)\), \(t(241) = 21.43, p < 0.001\). Figure 1 shows an example of the salience map of an unmanipulated image, a contrast-reduced image, and a contrast-increased image.

**Procedure**

Participants were seated with their head constrained in a chin rest and were given verbal instructions regarding the experimental procedure. Calibration and validation of their eye position was performed. In order to encourage free exploration of the images, participants were instructed to explore the images carefully so as to remember them for a later recognition task. Each trial began with a drift-correction screen in which participants were required to press the space bar while fixating a centrally presented fixation cross. Afterwards, the cross remained on-screen for 225 ms, followed by a 75-ms blank screen. This 75-ms blank was introduced in order to facilitate fast first eye movements (Fischer & Ramsperger, 1984; Reuter-Lorenz, Hughes, & Fendrich, 1991). The image was then presented for 2 s. Four hundred images were randomly selected from the image set, 200 with the contrast adjustment on the left and their counterparts with contrast adjustment on the right. These images were presented in a mixed random order. Thus, each participant saw the same image (with a different side adjusted) twice throughout the experiments, with the restriction that this could never happen within three
consecutive trials. Afterwards, participants performed a recognition task on 20 previously presented images and 20 new images selected from the image pool. Participants were asked to press the Z key if they had seen the image before and the slash key (“/”) if they had not seen the image before. Each image was presented once and remained on-screen until a response was made.

To become familiar with the task, participants viewed 10 contrast-reduced (Experiment 1a) or contrast-increased (Experiment 1b) practice images followed by 10 recognition trials (five old and five new images) prior to the experimental trials. Immediately after the practice trials, participants were given feedback on their performance and had an opportunity to ask questions about the experimental procedure. At the conclusion of the experiment, participants were asked to type in a field on the screen whether or not they had “noticed anything strange about the images.” This was done in order to check whether the contrast adjustment was noticed by the participants. The complete experiment lasted approximately 60 min.

Data processing

Previous research employing more basic displays has demonstrated that salience has a reflexive influence on eye movements when saccades are initiated quickly into a display. Thereafter, top-down influences may be mainly responsible for eye-movement control (Donk & van Zoest, 2008; Siebold et al., 2011). The typical pattern is that short-latency initial eye movements into a display are clearly controlled by salience, while long-latency initial eye movements are not. For this reason, we split each saccade into several latency bins to look at the time course of selection behavior. We were interested most in the first saccade into an image, as there is a clear prediction that when initiated quickly, it will be influenced by our contrast adjustment. However, we also included an analysis of subsequent eye movements to look for similar patterns of capture that might emerge later in the trial. These later saccades can also provide some information on how the contrast manipulation may have influenced eye-movement behavior throughout scene viewing.

The data were handled in the same way for Experiments 1a and 1b. First, temporal and spatial outliers were removed. Saccades were removed if their preceding fixation duration was shorter than 90 ms or longer than 700 ms or if they landed outside of the image boundary. Saccades initiated faster than 90 ms were removed in order to exclude eye movements that were anticipatory (if they preceded the first saccade in the trial) or corrective (for later saccades). Outlier removal resulted in 9.36% of eye movements removed from Experiment 1a and 10.83% of eye movements removed from Experiment 1b. Saccades subsequent to Saccade 7 were collapsed into one group. Then saccades were sorted into four bins based on the duration of their preceding fixation, except for the first saccade. The latency of first eye movement was calculated relative to the onset of the display. Recognition trials were not included in the analysis of the eye-movement data.

The primary dependent measure was the proportion of saccades that landed in a higher or lower contrast region. This was either the left or right side of the image, from the center to the edge of the image boundary (i.e., including half of the middle 20% of the image).

A second measure, local luminance contrast at fixation, was calculated (Einhäuser et al., 2006). Luminance contrast at fixation is defined as the standard deviation of the luminance in a patch around fixation (80 × 80 pixels corresponding to approximately 1.72° × 1.72° of visual angle) divided by the mean luminance of the image. To take into account any biases that might arise due to incidental regularities in the local contrast dependent on fixation number, and to account for the central bias known to be most pronounced for early saccades (Einhäuser et al., 2006; Tatler, 2007), we additionally calculated normalized luminance contrast at fixation, which represented luminance contrast at fixation relative to baseline values separately calculated for each fixation number and participant. The baseline for the fixation following the nth saccade was calculated by averaging the luminance contrast values of all fixations following the nth saccade of a given participant over the course of the experiment. These baselines were then subtracted from the values of local contrast at fixation for the corresponding fixations that participants made in a given trial. Thus, values above zero represent local contrast that is higher than the average local contrast for that fixation number, and values below zero represent local contrast that is below average local contrast for that fixation number. This measure reveals any variations in local contrast associated with saccade latency.

In addition, a measure of saccade bias was calculated as the proportion of saccades that switched from the higher to lower contrast side in the image or vice versa. It is the proportion of saccades that started in one type of contrast side (higher or lower) and landed in the other (lower or higher), and represents the likelihood that a given saccade switched sides during viewing. This measure was introduced in order to investigate how the contrast manipulation across our images may have influenced eye-movement behavior later in the trial. It makes explicit whether participants preferred one side of the image region over another. If the salience distribution across the image influences scanning...
behavior, more switches will be made into the higher contrast region of our images, and participants will be more likely to stay in that region.

Results

Experiment 1a

Behavioral results: Participants scored on average 86.04% (SD = 7.19%) correct in the recognition task. Four out of 12 participants indicated in the follow-up questionnaire that they had noticed the contrast manipulation.

Proportion to highest contrast region: Figure 2A shows the proportion of saccades that landed in the higher contrast region of the image in relation to the latency of the preceding fixation. A 7 (saccade number) × 4 (fixation latency bin) within-subjects analysis of variance was conducted on this proportion to investigate to what extent eye movements may have been biased toward the higher contrast side of the image. For clarity, contrasts were planned such that for the saccade number, the first eye movement was always compared to the average of Saccades 2–7, and the fixation latency bin was treated as linear. There was a main effect of saccade number, where first saccades were more likely to land on a higher contrast region, \( F(1, 11) = 64.047, MSE = 0.698, p < 0.001, \eta^2_p = 0.853 \).

In addition, the first eye movement, in general, was more likely than chance to land on the higher contrast region in a one-sample \( t \) test against 0.5, \( t(11) = 13.952, p < 0.001 \). There was a main effect of fixation latency bin, where short-latency saccades were more likely than long-latency saccades to land on a higher contrast region, \( F(1, 11) = 11.904, MSE = 0.001, p < 0.01, \eta^2_p = 0.520 \). Both main effects were qualified by an interaction between saccade number and latency bin, \( F(1, 11) = 29.788, MSE = 0.341, p < 0.001, \eta^2_p = 0.730 \), such that short-latency first saccades were particularly more likely than long-latency first saccades to land on a higher contrast region (see Figure 2A).

Normalized local luminance contrast at fixation: Figure 3A shows the normalized local contrast at fixation following each saccade across latency bins. A 7 (saccade number) × 4 (fixation latency bin) within-subjects analysis of variance was conducted. Because normalized contrast at fixation has an averaged value of 0 for each saccade, our reports in this and subsequent local-contrast analyses are limited to those involving main effects of latency bin and interactions between saccade number and latency bin. There was a marginal effect of latency bin, suggesting that short-latency saccades landed on higher local contrast regions than did long-latency saccades, \( F(1, 11) = 4.197, MSE = 0.005, p = 0.068, \eta^2_p = 0.296 \). There was a marginal interaction between saccade number and latency bin,
F(1, 11) = 4.500, MSE = 0.035, p = 0.060, $\eta^2_p = 0.310$, such that short-latency first saccades tended to land on a higher local contrast region than did long-latency first saccades.

**Saccade bias**: Figure 4A and B shows the likelihood of switching from the higher or lower contrast sides to the lower or higher contrast sides, respectively, across latency bin for the second to seventh saccades. In general, participants were more likely to stay on the side where they were than to make a switch to the other side; the likelihood of a side switch was below 50% for switches both from the normal- to the reduced-contrast region ($M = 0.36$), $t(11) = 7.56$, $p < 0.001$, and from the reduced- to the normal-contrast region ($M = 0.45$), $t(11) = 2.44$, $p = 0.033$. To investigate any systematic dependencies on initial contrast side, saccade number, and fixation latency bin, a 2 (initial contrast side: normal or reduced) $\times$ 6 (saccade number: 2–8) $\times$ 4 (fixation latency bin) within-subjects analysis of variance was conducted on the proportion of saccades that switched from the lower or higher contrast side of the image to the lower or higher contrast side of the image, respectively. Contrasts were planned as linear for saccade number and latency bin. There was a main effect of initial contrast side, $F(1, 11) = 45.244$, $MSE = 0.058$, $p < 0.001$, $\eta^2_p = 0.804$, where the likelihood of switching from the reduced-contrast side to the normal-contrast side was higher than that of the opposite switch. There was also a marginal linear effect of saccade, $F(1, 11) = 3.608$, $MSE = 0.012$, $p = 0.084$, $\eta^2_p = 0.247$, where there was a slight tendency for later saccades to switch contrast sides less often than earlier saccades. There was no main effect of latency bin, nor an interaction between initial contrast side and saccade, initial contrast side and latency bin, or saccade and latency bin, all $F$s < 1. There was an interaction between initial contrast side, saccade number, and latency bin, $F(1, 11) = 8.817$, $MSE = 0.045$, $p = 0.034$, $\eta^2_p = 0.346$, which probably related to the tendency for switches from the reduced- to the normal-contrast side to vary more with latency across saccades than switches from the normal- to the reduced-contrast side.

**Experiment 1b**

**Behavioral results**: Participants scored on average 86.45% ($SD = 8.15\%$) correct in the recognition task. As in Experiment 1a, only four out of 12 participants indicated in the follow-up questionnaire that they had noticed the contrast manipulation.
Proportion to higher contrast region: Figure 2B shows the proportion of saccades that landed in the higher contrast region in relation to the latency of the preceding fixation. As in Experiment 1a, a 7 (saccade number) × 4 (fixation latency bin) within-subjects analysis of variance was conducted on the proportion of saccades that landed in the higher contrast region of the image. There was a main effect of saccade number, where first saccades were directed to the higher contrast region more often than later saccades, $F(1, 11) = 34.353$, $MSE = 0.171$, $p < 0.001$, $\eta_p^2 = 0.757$. In addition, the first eye movement, in general, was more likely than chance to land on the higher contrast region in a one-sample $t$ test against 0.5, $t(11) = 4.939$, $p < 0.001$. However, there was no main effect of latency bin, $F < 1$. There was an interaction between saccade number and latency bin, $F(1, 11) = 7.775$, $MSE = 0.127$, $p < 0.05$, $\eta_p^2 = 0.414$, such that short-latency first saccades were particularly more likely to land on a higher contrast region than long-latency first saccades, compared to the subsequent saccades.

Normalized local luminance contrast at fixation: Figure 3B shows the normalized local contrast at fixation following each saccade across latency bin. A 7 (saccade number) × 4 (fixation latency bin) within-subjects analysis of variance showed no effect of latency bin, $F(1, 11) = 1.865$, $MSE = 0.009$, $p = 0.199$, $\eta_p^2 = 0.145$, nor an interaction between saccade number and latency bin, $F(1, 11) = 1.651$, $MSE = 0.046$, $p = 0.225$, $\eta_p^2 = 0.130$.

Saccade bias: Figure 4C and D shows the likelihood of switching from the higher or lower contrast side to the lower or higher contrast side, respectively, across latency bin for the second to seventh saccades. Again, the likelihood of a side switch was less than 50%, both for switches from the increased- to the normal-contrast region ($M = 0.39$), $t(11) = 6.15$, $p < 0.001$, and for switches from the normal- to the increased-contrast region.
Discussion

In Experiments 1a and 1b, we employed a contrast manipulation across the images in order to manipulate the distribution of salience. The results of both Experiments 1a and 1b revealed that the first saccade was more likely to land on a higher contrast region. In addition, this effect varied with the latency of the preceding fixation duration, where short-latency first saccades into the scene were more likely to land on a higher contrast side relative to long-latency first saccades. In Experiment 1a, where the images were reduced in contrast on one side, the normalized local contrast at fixation following the first saccade relative to the others tended to vary with the preceding fixation latency, where short-latency saccades landed on regions of high local contrast compared to long-latency saccades. Not unexpectedly, saccades subsequent to the first did not vary significantly with the latency of the preceding fixation. Taken together, these results suggest that the early deployment of attention in a natural scene is indeed biased toward regions of high salience.

Interestingly, saccades initiated from one side of the image tended to land on that same side, with the exception of the reduced-contrast side in Experiment 1a: These eye movements were more likely than those initiated in the normal-contrast side to switch sides. In general, observers tended to look more at the higher contrast region of an image and then stay there. These results are consistent with previous work showing a general bias toward higher contrast regions of an image (e.g., Einhäuser et al., 2006) and the elevation of salience at fixation in general (Parkhurst et al., 2002). We extend that work by showing that the extent to which first saccades are biased toward the higher contrast region of the image varies with the preceding fixation latency.

Even though normalized local contrast at fixation showed similar patterns to those of the proportion to higher contrast data, it did not vary with latency in the first saccade of Experiment 1b. It is possible that this result was related to the fact that the increased-contrast images of Experiment 1b generally contained many locations with high salience, resulting in substantial competition of salient regions across the entire image (see Figure 1C). The presence of many salient locations in the increased-contrast images—especially in the center—as opposed to only a few (very) salient locations in the reduced-contrast images, may have partly concealed the occurrence of true salience-driven selection behavior in Experiment 1b. In particular, it may explain why an effect of latency was not found for normalized local contrast at fixation, even though this effect was present in the proportion of saccades to the higher contrast side.

Experiments 1a and 1b confirmed our intuition that the contrast manipulations influence the direction of the first eye movement into a scene and that this effect varies with its latency—a strong indication for the idea that initial selection was truly influenced by salience (e.g., van Zoest et al., 2004a). However, it is unclear whether this effect will survive in the presence of a strong top-down goal.

Experiment 2

In Experiment 2 we used a template visual search task modeled on the work of Einhäuser, Rutishauser et al. (2008) in order to investigate whether the early effects of salience as well as its dependence on preceding fixation latency, as observed in Experiments 1a and 1b, survives in the presence of a strong top-down goal thought to override salience effects (Einhäuser, Rutishauser et al., 2008; Henderson et al., 2007). Unlike Einhäuser, Rutishauser et al. (2008), we varied the visibility of the template target in a manner that depended on the target’s immediate background as well as the visual search performance on the previous trial. This reduced the possibility of biases due to higher visibility of the target in one or the other side of the image, as well as the potential for immediate target acquisition.

Methods

Participants

Twelve participants (ages 18–25 years, M = 21) were recruited from the Vrije Universiteit Amsterdam and participated for course credit or for 9 euros. All
participants reported normal or corrected-to-normal vision and were unaware of the purpose of the experiment. Again, the study and consent procedure were approved by the ethics board of the faculty of psychology and education and conducted according to the principles of the Declaration of Helsinki.

**Apparatus**

The experimental apparatus was exactly the same as that used for Experiments 1a and 1b.

**Stimuli**

One hundred images for the experimental phase and 20 for the practice phase were randomly selected from the images used in Experiments 1a and 1b. Four versions were created of each image as in Experiments 1a and 1b: contrast reduction at the left or right side and contrast increase at the left or right side. To each image, a bull’s-eye was added that acted as a target. The target location was randomly selected on a trial-by-trial basis to appear within an annulus around the center of the screen such that it was greater than 200 pixels (approximately 4.3° of visual angle) and less than 384 pixels away from the image center, as well as 20 pixels away from the horizontal screen edges. Moreover, the target was equally likely to be presented at either side of each image.

The bull’s-eye was defined by increasing and decreasing the intensity of the pixels around its center location \((x_0, y_0)\). First, the bull’s-eye filter \(B\) was defined on the basis of the distance \(d\) of each pixel to the center location. This filter was formed by multiplying a Gaussian density function with a sinusoid:

\[
B_{xy} = \exp\left(-\frac{d^2}{2\sigma^2}\right) \sin\left(\frac{d}{b}2\pi\lambda\right)
\]  

(1)

The parameter values \(\sigma = 5\), \(b = 13.44\), and \(\lambda = 3\) constrained the span of the bull’s-eye to approximately 20 \(\times\) 20 pixels (approximately 0.43° \(\times\) 0.43° of visual angle) and ensured that the bull’s-eye had three rings.

To make the bull’s-eye target approximately equally visible across locations and pictures, the pixel intensity change was made dependent on the prior pixel intensity \(I_{xy}\) and local contrast \(s_l(x_0, y_0)\) in the image. The change to the intensity \(\Delta I_{xy}\) of pixel \((x, y)\) was then defined by

\[
\Delta I_{xy} = \begin{cases} 
B_{xy} s_l(x_0, y_0) c I_{xy} & \text{if } B_{xy} < 0 \\
B_{xy} s_l(x_0, y_0) c(1 - I_{xy}) & \text{if } B_{xy} > 0 
\end{cases}
\]  

(2)

Equation 2 implies that the bull’s-eye brightened or darkened pixels depending on their prior pixel intensity. In a very bright patch, the dark rings would stand out, while in a dark patch, it is the light rings that would stand out. Moreover, the contrast of the ring was higher when there was high contrast at the bull’s-eye location than when there was low contrast. This local contrast \(s_l(x_0, y_0)\) was quantified by the standard deviation of pixel intensity values within the 20 \(\times\) 20 pixel patch surrounding the center location. To avoid situations in which the bull’s-eye became invisible or extremely high in contrast, the standard deviation of the region patch was constrained such that it was never below 0.08 or above 0.35. We included a scaling factor \(c\) that was initialized to 3 but depended on the participants’ performance on the previous trial (see Procedure, below).

Images including the bull’s-eye target were presented centrally at their native 1024 \(\times\) 768 pixel resolution such that they subtended a visual angle of approximately 22° horizontal \(\times\) 17° vertical. The rest of the screen surrounding the image was black.

**Procedure**

Each trial began with a drift-correction screen in which participants were required to press the space bar while fixating a centrally presented fixation cross. The image with the hidden bull’s-eye target was then presented, and participants were required to search for the target. Participants were allowed only eight eye movements to find the target, in order to encourage speeded exploration of the image. Fixation and saccade parameters were monitored in real time, and an eye movement was considered valid and counted toward the total eight allowed if it was larger than 60 pixels (approximately 1.3° of visual angle). A target was considered found if the participant fixated within 100 pixels (approximately 2.15° of visual angle) of the bull’s-eye, after which a small green arrow appeared for 2 s pointing toward the target and indicating successful completion of the trial. If the participant failed to locate the target before making eight eye movements, a small red arrow appeared for 2 s indicating the location of the target.

The visibility of the target was modified on a trial-by-trial basis in relation to real-time performance. If the target was found within three eye movements on the previous trial, the contrast of the target was reduced by reducing the scaling factor \(c\) of the previous trial (see Equation 2) by 0.2. If the target was not found within three eye movements on the previous trial, \(c\) was increased by 0.2 to increase the contrast of the target. Thus \(c\) varied from trial to trial throughout the experiment. This adjustment was done in order to encourage multiple eye movements being made before finding the target; the values were essentially arbitrary and selected based on pilot testing.

Participants were seated comfortably in the chin rest and given written instructions regarding the experimental procedure, including an example of the bull’s-
Results

To remain consistent with the analyses of Experiments 1a and 1b, we chose to analyze the eye-movement data from the reduced-contrast images and increased-contrast images separately. As a control, we also ran a complete 2 (image contrast adjustment: increased or reduced) × 7 (saccade number) × 4 (fixation latency bin) within-subjects analysis of variance both on the proportion of eye movements that went to the higher contrast region and on the normalized local luminance contrast at fixation. In the proportion of eye movements that went to the higher contrast region, there was no main effect of image contrast adjustment, $F(1,11) = 3.026, MSE = 0.107, p = 0.110$, $\eta_p^2 = 0.216$; the only interaction with this factor on this proportion data was with saccade number, $F(1, 11) = 11.906$, $MSE = 0.860, p = 0.005$, $\eta_p^2 = 0.520$, such that first saccades were more likely to land in the higher contrast region for reduced-contrast images compared to increased-contrast images. For normalized local contrast at fixation, there was no main effect of image contrast adjustment, $F < 1$.

Behavioral results

Participants correctly located the target on 84.1% of the trials. The target was somewhat easier to find in reduced-contrast images ($M = 87.9\%, SD = 2.8\%$) than in increased-contrast images ($M = 80.2\%, SD = 3.5\%$), $r(11) = 8.16, p < 0.001$. The visibility of the target was manipulated both relative to the immediately surrounding pixels and in a staircase manner (see Methods section), in order to encourage active exploration before finding the target. This was achieved in that on average, four eye movements were made before the target was found. Only one of the 12 participants indicated noticing the contrast manipulation in the follow-up questionnaire.

Reduced-contrast images

Proportion to higher contrast region: Figure 5A shows the proportion of saccades that landed in the normal-(higher) contrast region of the image in relation to the latency of the preceding fixation for reduced-contrast images. In order to investigate to what extent eye movements were directed toward the higher contrast side of the image, a 7 (saccade number) × 4 (fixation latency bin) within-subjects analysis of variance was conducted on this proportion. Planned contrasts were conducted on saccade number, where the first saccade was compared to the average of the later saccades and fixation latency bin was treated as linear, as in Experiments 1a and 1b. There was a main effect of saccade number, where first saccades were more likely to land on the higher contrast region of the image than later saccades, $F(1, 11) = 16.362$, $MSE = 0.506, p < 0.005$, $\eta_p^2 = 0.598$. In addition, this first eye movement, in general, was more likely than chance to land on the
higher contrast region in a one-sample t test against 0.5, \( t(11) = 5.625, p < 0.001 \). There was a main effect of latency bin, where short-latency saccades were more likely to land on a higher contrast region than long-latency saccades, \( F(1, 11) = 5.905, MSE = 0.093, \eta^2_p = 0.349 \). These main effects were qualified by an interaction between saccade number and latency bin, \( F(1, 11) = 5.905, MSE = 0.093, \eta^2_p = 0.349 \), such that short latency first saccades were more likely than long latency and subsequent saccades to land on the higher contrast region (see Figure 5A).

### Normalized local luminance contrast at fixation

Figure 6A shows the normalized local contrast at fixation across latency bins following each saccade (described in the Data processing section of Experiment 1) for the reduced-contrast images. A 7 (saccade number) \( \times 4 \) (fixation latency bin) within-subjects analysis of variance showed no main effect of latency bin, \( F < 1 \).

There was an interaction of saccade number and latency bin, \( F(1, 11) = 10.548, MSE = 0.043, p = 0.008, \eta^2_p = 0.490 \), such that shorter-latency first saccades were more likely to land on a region of higher local contrast than longer-latency first saccades, whereas subsequent saccades did not show this pattern.

### Saccade bias

Figure 7A and B shows the likelihood of switching image sides across latency bin for the second to seventh saccades for the reduced-contrast images. In general, participants were more likely to stay on the side where they were than to make a switch to the other side; the likelihood of a side switch was below 50% for switches both from the normal- to the reduced-contrast region (\( M = 0.31 \), \( t(11) = 16.915, p < 0.001 \), and from the reduced- to the normal-contrast region (\( M = 0.35 \), \( t(11) = 9.478, p < 0.001 \). To investigate any systematic influences of the contrast side where the saccade was initiated, saccade number, and fixation latency bin, a 2 (initial contrast side: normal or reduced) \( \times 6 \) (saccade number: 2–7) \( \times 4 \) (fixation latency bin) within-subjects analysis of variance was conducted on the proportion of saccades that switched contrast sides. Contrasts were planned as linear for saccade number and latency bin.

There was an effect of initial contrast side, \( F(1, 11) = 11.514, MSE = 0.016, \eta^2_p = 0.511 \), such that switches from reduced- to normal-contrast were more likely than switches from normal- to reduced-contrast.

There was an effect of saccade number, \( F(1, 11) = 3.528, MSE = 0.043, p = 0.008, \eta^2_p = 0.243 \), where there was a tendency for earlier saccades to switch contrast sides more often than later saccades, potentially reflecting the eventual localization of the target (even though explicitly target-directed eye movements were excluded from the analysis). There was no main effect of latency bin, \( F(1, 11) = 1.778, MSE = 0.068, p = 0.171, \eta^2_p = 0.139 \), nor was there an interaction between initial contrast side and saccade number, \( F(1, 11) = 1.675, MSE = 0.032, p = 0.156, \eta^2_p = 0.132 \).
contrast side and latency bin, $F(1, 11) = 2.331$, $MSE = 0.032$, $p = 0.092$, $\eta^2_p = 0.175$, or saccade number and latency bin, $F(1, 11) = 1.273$, $MSE = 0.036$, $p = 0.224$, $\eta^2_p = 0.104$, nor between initial contrast side, saccade number, and latency bin, $F < 1$.

**Increased-contrast images**

**Proportion to higher contrast region:** Figure 5B shows the proportion of saccades that landed in the increased-(higher) contrast region of the image in relation to the latency of the preceding fixation for the increased-contrast images. A 7 (saccade number) × 4 (fixation latency bin) within-subjects analysis of variance was conducted on this proportion. There were no effects of saccade or latency bin, nor was there an interaction between these two, all $F$s < 1. The first eye movement was only marginally more likely than chance to land on the higher contrast region in a one-sample $t$ test against 0.5, $t(11) = 1.919$, $p = 0.081$.

**Normalized local luminance contrast at fixation:** Figure 6B shows the normalized local contrast at fixation following each saccade across latency bins. A 7 (saccade number) × 4 (fixation latency bin) within-subjects analysis of variance showed a main effect of latency bin, $F(1, 11) = 7.512$, $MSE = 0.032$, $p = 0.019$, $\eta^2_p = 0.406$, such that, in general, short-latency saccades were more likely to land on a region lower in local contrast compared to long-latency saccades. However, there was an interaction between saccade and latency bin, $F(1, 11) = 4.939$, $MSE = 0.130$, $p = 0.048$, $\eta^2_p = 0.310$, such that short-latency first saccades were more likely to land on a region higher in local contrast than long-latency first saccades, compared to subsequent saccades, which show the opposite pattern.

**Saccade bias:** Figure 7C and D shows the likelihood of switching image sides from higher or lower contrast sides to lower or higher contrast sides, respectively, across latency bin for the second to seventh saccades for the increased-contrast images. Again, the likelihood of a side switch was less than 50%, both from the increased- to the normal-contrast region ($M = 0.30$), $t(11) = 14.853$, $p < 0.001$, and from the normal- to the increased-contrast region ($M = 0.35$), $t(11) = 7.945$, $p < 0.001$. To investigate any systematic influences of the contrast side where the saccade was initiated, saccade number, and fixation latency bin, a 2 (initial contrast side: increased or normal) × 6 (saccade number: 2–7) × 4 (fixation latency bin) within-subjects analysis of variance was conducted on the proportion of saccades...
that switched sides. Contrasts were planned as linear for saccade number and latency bin. There was an effect of initial contrast side, $F(1, 11) = 25.941, MSE = 0.012, p < 0.002, \eta^2_p = 0.702$, such that switches from the normal- to the increased-contrast side of the image were more likely than switches from the increased- to the normal-contrast side of the image. There was an effect of saccade number, $F(1, 11) = 8.457, MSE = 0.027, p < 0.001, \eta^2_p = 0.435$, where there was a tendency for earlier saccades to switch contrast sides more often than later saccades. There was also an effect of latency bin, $F(1, 11) = 6.364, MSE = 0.036, p = 0.002, \eta^2_p = 0.366$, where short-latency eye movements were more likely to switch image sides than long-latency eye movements. There was no interaction between initial contrast side and saccade number, $F < 1$, initial contrast side and latency bin, $F < 1$, saccade number and latency bin, $F(1, 11) = 1.707, MSE = 0.021, p = 0.054, \eta^2_p = 0.134$, nor between initial contrast side, saccade number, and latency bin, $F(1, 11) = 1.116, MSE = 0.023, p = 0.346, \eta^2_p = 0.092$.

**Discussion**

In Experiment 2 we examined the extent to which eye movements in natural scenes are affected by reducing or increasing the contrast on one side of an image in a visual search task. Despite the strong top-down goal in this template visual search task, initial saccades were more likely than subsequent ones to land on a region of normal contrast than on a region of reduced contrast. This was particularly the case for short-latency first eye movements. The contrast increase manipulation did not affect performance to the same extent. Initial saccades were not more likely to land on the increased-contrast region of the image, and this did not vary with the latency to initiate a saccade. Moreover, normalized
local luminance contrast varied with the latency of the preceding fixation, such that short-latency first saccades tended to land on a higher local contrast region than long-latency saccades. This was the case for both reduced-contrast and increased-contrast images. Throughout scene viewing, when an eye movement traversed across the image, it was more likely to switch from the lower contrast side of the image into the higher contrast side. Taken together, these results suggest that even when they are instructed to search for a target, participants’ gaze selection is influenced by the salience distribution across an image. Moreover, the results suggest that this is particularly the case for short-latency initial saccades.

These results stand in apparent contrast to previous work using a very similar template target search paradigm where task demands overrode any image salience effects (Einhäuser, Rutishauser et al., 2008); however, there are several differences between that and the present work. First, we manipulated target visibility in order to avoid direct or early target acquisition and to encourage exploration. Second, although this did not change the pattern of our results, we excluded eye movements that were directed toward and eventually acquired the target, as these eye movements were ambiguous as to whether they were influenced by salience or top-down goals. Third, we introduced a modest time constraint such that participant responses were more speeded. For this reason alone it may be the case that saccades were initiated faster than in previous work, allowing us to find a salience effect in the fastest saccades.

It is also possible that a similar effect was present in the experiment by Einhäuser, Rutishauser et al. (2008) but only for short-latency initial saccades. We requested, and were kindly granted, their data in order to confirm our suspicions. In the template search (bull’s-eye) condition, the bias toward the high-contrast region of the image for the first saccade was linearly dependent on the latency of the preceding fixation, in both the condition in which the target was predominantly placed in the low-contrast region and the one in which it was distributed evenly across regions: a linear main effect of latency bin, $F(1, 4) = 10.24, MSE = 5624.18, p < 0.01, \eta_p^2 = 0.72$. This result is consistent with our findings and suggests that even with a different experimental design, the extent to which early eye movements may be influenced by salience is dependent on the latency of the first saccade.

An interesting pattern emerged in the normalized local contrast at fixation, where the average local luminance contrast (displayed in the top left of each panel in Figure 6) was elevated following the first saccade relative to subsequent saccades. One account of this finding is that it reflects the central bias in scene viewing, where early saccades into a scene typically fall close to the center of the image (Tatler, 2007). Interestingly, the average local contrast was elevated per saccade to a similar extent in both the reduced-contrast and increased-contrast images, suggesting that indeed the central bias is fairly systematic in natural scene viewing. It is possible that feature contrasts were elevated in the center of the image. In our images, we did not control for where the majority of features were, beyond restricting the ratio of mean intensity differences across the two sides of the image. By normalizing the local luminance contrast per saccade, we were able to factor out any effects of the central bias and focus only on how the local contrast varied with the preceding saccade latency. When the first saccade was initiated quickly, it tended to land on a region of higher local contrast than when it was initiated more slowly. This suggests that these initial eye movements are indeed sensitive to the underlying salience of the image.

Interestingly, beyond the first saccade there appears to be some influence of the contrast manipulation on where participants look in this visual search task. In general, participants were more likely to switch from the lower to the higher contrast side of an image than vice versa, particularly on early saccades beyond the first one. Switches in image side reduce as time in the trial progresses, likely reflecting either an early bias to explore the gross outlay of the scene or a progressing localization of the target. This influence of the contrast manipulation beyond the first eye movement is surprising and may have to do with particular correspondences between the contrast manipulation and a given participant’s search strategy. An intriguing alternative explanation is that as the central bias in scene viewing decreases over saccade number, a bias toward the higher contrast side of the image emerges. This is in line with previous work demonstrating that in visual search tasks where the target is defined by luminance contrast, fixation placement later in scene viewing follows the distribution of features in the image (Tatler, 2007).

**General discussion**

Previous research has established that saccades in simple artificial displays are affected by bottom-up salience, even in the presence of a strong top-down goal such as in visual target search (Siebold et al., 2011; van Zoest et al., 2004a). By including both a contrast reduction and a contrast increase manipulation, and by looking at the time course of selection behavior, we were able to demonstrate that salience indeed influences short-latency and early saccades in natural scenes even when participants are given a strong top-down goal such as template search, at least in the case of contrast-
reduced images. These results seem to contradict previous work showing little to no role for bottom-up salience in the control of saccades in natural scenes, but they are in line with current thinking on saccade control in simpler displays and to some extent in natural scenes (Donk & van Zoest, 2008; Einhäuser, Rutishauser et al., 2008; Einhäuser, Spain et al., 2008; Foulsham & Underwood, 2008; Henderson et al., 2007; Tatler et al., 2005; van Zoest et al., 2004a).

The influence of salience was stronger and the latency effects clear in Experiments 1a and 1b, in which participants were free to explore the image, than in Experiment 2, in which they engaged in a search task. This suggests that the presence of a top-down goal affects oculomotor selection, even early on. It is possible that the observers in Experiment 2 preprogrammed their initial saccade in line with their expectations concerning the possible target locations (Rao, Zelinsky, Hayhoe, & Ballard, 2002). As a result, initial eye movements might have been less affected by salience than those in Experiments 1a and 1b. This finding points to the idea that there are different degrees of top-down involvement. The view known as strategic divergence posits that low-level image features continuously influence saccadic targeting and it is the top-down influence that varies over time. Tatler et al. (2005) suggest that the visual representation required for such a high-level system may be an intermediate-level representation that allows for discrimination between informative and noninformative locations. This representation would also be insensitive to differences in the real world, such as illumination, but would preserve, for example, contrast and edge content (which have been previously shown to be highly discriminatory at fixated, compared to nonfixated, locations; Parkhurst et al., 2002; Tatler et al., 2005).

A more radical theory is that natural scene viewing is under full cognitive, knowledge-based control (Henderson et al., 2007). Saccade targets in this view are ranked and selected purely on the relevance to the current task and based on a pre-existing knowledge base of episodic and semantic memory. According to this view, the correspondence between salience and fixations is merely accidental, because regions of high salience tend to be where relevant objects are located (Nuthmann & Henderson, 2010; Walther & Koch, 2006). Support for this view lies in the idea that such episodic and semantic scene knowledge, or gist, can be extracted very quickly (Oliva & Torralba, 2006), and albeit in conjunction with the salience model, such a knowledge base predicts fixation locations well (Torralba, Oliva, Castelhano, & Henderson, 2006). In our work, targets tagged by such a knowledge system would presumably have existed on both the normal and contrast-manipulated sides of the image, and selection subsequent to the first saccade is therefore not fully biased to the higher contrast region. However, to the extent that the initial saccade was most likely into a higher contrast region, this suggests that a knowledge-based representation is either built on or influenced by, at least initially, bottom-up salience information.

In both Experiments 1a and 1b and Experiment 2, we found differential effects with reduced-contrast and increased-contrast images. Whereas reductions in contrast strongly affected the first saccade, the effects of increased contrast were smaller and less dependent on the latency of the initial saccade. One potential explanation of these differences is that the salience distribution was remarkably different between the two types of contrast manipulation (see Figure 1). In the contrast-reduced images, the number of salient regions decreased dramatically. In these images, the normal-contrast side usually contained the most salient regions, with one of these regions typically winning the competition for selection. The increased-contrast images contained many salient regions in both the manipulated and unmanipulated regions, resulting in less of an advantage for the increased-contrast side. Such an explanation is predicated on the notion that the target of an eye movement is chosen after competition for selection in some salience map (Koch & Ullman, 1985; Wolfe, 1994). An alternative explanation would be that less salient objects simply take longer to process. In our reduced-contrast manipulation, early initial saccades might be biased toward the normal side because processing of the reduced-contrast side of the image would not be complete yet. Slower initial saccades and later saccades would then be made when processing of the reduced-contrast side was complete and would therefore not be biased toward the normal-contrast side to the same extent. Increasing contrast may speed up processing relative to normal contrast, leading to smaller effects of contrast increase relative to contrast decrease and less variation with latency. Similar time-based explanations have been suggested for the effects of salience in artificial displays (Donk & van Zoest, 2008). It is still an open question whether such explanations can account for all effects of salience on eye-movement behavior (Siebold, van Zoest, Meeter, & Donk, 2013).

In the present work, we computed local contrast at fixation following each saccade by normalizing each value relative to the corresponding averaged value per subject obtained for that ordinal saccade number. This was done to correct for any potential influence of a central bias, which is known to decrease with saccade number (Tatler, 2007). The averaged contrast values dropped significantly over the first few eye movements (see insets of Figures 3 and 6). This decrease may be related to an initial bias to select center locations, which tend to have a higher contrast value than more peripheral locations. However, the central bias is
thought to be intrinsically invoked by the observer rather than being related to any image features. Regardless, any potential central-bias confound was removed in the normalized contrast at fixation. These values show that short-latency first saccades tended to land on a region higher in local contrast than long-latency first saccades, irrespective of the central bias.

In both experiments, salience had its strongest effects on the initial saccade into an image. These results are in line with theories based on work using simpler displays (e.g., Siebold et al., 2011; Theeuwes & Godijn, 2002; van Zoest et al., 2004a; Zehetleitner et al., 2013). One current theory is that stimulus-driven salience is only briefly represented in the visual system. Relative salience effects are short-lived (within the first few hundred milliseconds of a display onset). To explain these results, Donk and Soesman (2010, 2011) have proposed that relative salience might dissipate early, while information concerning the locations of distinct objects might be preserved over a longer period of time. This latter information might serve as a basis for further goal-driven selection in that it provides information about the locations of potentially interesting targets. The salience map in this sense is assumed to evolve over time from a representation containing information concerning the relative salience of different locations in the visual field to a representation containing information concerning the locations of (potentially interesting) objects (Donk & Soesman, 2010, 2011; Donk & van Zoest, 2008).

The conjecture that relative salience is replaced by an object-like representation (Donk & van Zoest, 2008) is mirrored by work in natural scenes showing that attentional selection critically depends on the objects in those natural scenes (Einhäuser, Spain et al., 2008; Nuthmann & Henderson, 2010; ’t Hart, Schmidt, Roth, & Einhauser, 2013). In our work, every image was presented to every participant with the contrast adjustment on the left and on the right, precluding any systematic relation between a particular object’s location and the contrast adjustment. Even if the early saccades were directed toward objects, not simply to highly salient regions, the salience distribution across the image influenced which objects were fixated first. This is consistent with the view of Donk and Soesman (2010, 2011) that the top-down component may rely on a representation of the scene that was built on initial salience and that this initial salience may have early effects on object parsing. Indeed, ’t Hart et al. (2013) have recently shown that if a rarely mentioned object (i.e., semantically unimportant) in a scene was increased in contrast relative to the underlying scene, the number of mentions of, and saccades to, that object increased. The authors suggest that an object’s importance is weighted relative to its expected importance given the underlying scene statistics of the image rather than the scene semantics alone. The extent to which underlying image salience may influence “objectness” or any intermediate-level representations serving top-down control (Donk & van Zoest, 2008; Henderson et al., 2007; Nuthmann & Henderson, 2010; Tatler et al., 2005) is a fruitful avenue for future research.

Regardless of the development and subsequent fate of such an intermediate-level representation, our results agree with the proposal that bottom-up and top-down control act on different timescales, where only short-latency, early eye movements are stimulus driven and top-down components take time to emerge (Donk & van Zoest, 2008; van Zoest et al., 2004a). This would indeed explain why, in our natural scenes, salience played a strong role early on. It also explains, to some extent, why reports of early salience effects have been mixed (Einhäuser, Rutishauser et al., 2008; Foulsham & Underwood, 2008; Nystrom & Holmqvist, 2008; Parkhurst et al., 2002; Tatler et al., 2005). Saccade latencies can depend critically on the difficulty of visual and cognitive processing (Nuthmann, Smith, Engbert, & Henderson, 2010). It is not yet clear under what exact conditions the initial saccade latency in a natural scene varies; however, it stands to reason that it may depend critically on task demands or stimulus complexity. If salience effects are strongest at the first moments of scene presentation (in our case, for the first two or three fixation latency bins), this may allow for a wide variety of differing reports on its actual influence when results are averaged over the millisecond-level time course. This was indeed the case for Einhäuser, Rutishauser et al. (2008), with salience effects only becoming apparent once this time course was uncovered.

The pattern of results reported here clearly indicates that the initial eye movement is influenced by the salience distribution across the images. But what can be said about any subsequent saccades? Participants generally spent more time fixating the higher contrast region of the image throughout the trials, both in the encoding and recognition and visual search experiments, and continued to be more likely to switch from the low-contrast to the high-contrast side than vice versa. This indicates that salience continues to affect later saccades; however, the mechanism is yet unclear. It is possible that the contrast manipulation could have influenced an intermediate-level representation of scene objects and their locations. Subsequently, the objects on the higher contrast side of the image maintained a sort of privileged status throughout scene viewing due to their higher initial salience (Donk & Soesman, 2011). Top-down mechanisms could therefore use such an intermediate-level map that was only initially influenced by salience. However, the lack of any dependence on latency in salience effects suggests that subsequent saccades are guided in a different way from the initial
saccade into an image. Previous work in both simple search displays and natural scenes suggests at least that these later eye movements are predominantly guided by top-down, cognitive control (e.g., Henderson et al., 2007; Siebold et al., 2011). The present results are not able to clearly distinguish between these accounts.

The analysis of eye-movement behavior in natural scenes is an important way of linking eye-movement behavior studied in simpler displays to the increasing need to study eye movements in the natural environment (Foulsham, Walker, & Kingstone, 2011; Hayhoe & Ballard, 2005; Kingstone, Smilek, & Eastwood, 2008; Tatler et al., 2011). Aside from practical reasons, the study of natural scenes in a laboratory setting is a compromise between stimulus control and naturalness. The onset of a natural scene may appear to be something that does not typically happen outside of the laboratory or away from our computers. However, it is a situation in which contextual and other forms of top-down cognitive information are not available to guide eye movements. Participants are uncertain as to the exact nature of the stimulus they are about to see. For this reason, the investigation of an initial eye movement into a natural scene allows us to better understand responses that occur in the absence of prior information about a scene or environment. Although as yet unknown, this may occur in situations encountered in our everyday lives, such as walking into an unfamiliar room, reorienting attention across a large space (turning the head or body), or viewing a scene change or cut in a movie or television show. We have shown that bottom-up stimulus features play a large role in guiding our early eye movements, perhaps because it allows for faster, more efficient object localization and may correlate with sudden onsets, motion, or unexpected objects. Eye movements in response to the sudden onset of a natural scene can help to reveal how the visual system deals with the interaction between the meaning of a scene and its bottom-up visual features. The present results therefore have important implications for theories and models of eye-movement behavior that are based on natural scene viewing and visual search.

**Conclusion**

In the present work, we have demonstrated that salience plays a role in the early deployment of attention into a natural scene. By decreasing or increasing the contrast in a gradient across an image, we manipulated the salience distribution. We have shown that fast first saccades are more likely to be directed toward a region of the image that contains more physically salient information. We have demonstrated that this occurs even when participants are given a strong top-down goal, such as searching for a bull’s-eye in an image. These results help to resolve some debate over the extent of salience effects in natural scene viewing, while helping to bridge the gap between research in more basic displays and those employing natural scenes.

**Keywords:** attention, eye movements, salience, natural scene viewing

**Acknowledgments**

We would like to thank Jeroen Silvis, Ben Tatler, and an anonymous reviewer for helpful comments on earlier versions of this manuscript. This work benefited from an NWO VIDI grant (452-09-007) to Martijn Meeter.

Commercial relationships: none.

Corresponding author: Nicola C. Anderson.

Email: n.c.c.anderson@vu.nl.

Address: Department of Cognitive Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.

**Footnotes**

1 For clarity and ease of computation, the pixel intensity values were converted to a scale from 0 to 1 instead of 0 to 255. After the application of the target, the image was converted back to the range 0 to 255.

2 We opted not to offset the fixation dot prior to image presentation as was done in Experiments 1a and 1b. Since participants were no longer free viewing the images but actively searching for a target, we no longer considered the prior offset of the fixation dot necessary to provoke fast initial saccades.

3 In our own work, we find significantly slower initial latencies in Experiment 1a, compared to 1b, but no difference in initial latencies between Experiments 1a and 1b on the one hand and 2 on the other.

**References**


