

The influence of a scene preview on eye movement behavior in natural scenes

Nicola C. Anderson¹ · Mieke Donk¹ · Martijn Meeter²

Published online: 12 April 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Rich contextual and semantic information can be extracted from only a brief presentation of a natural scene. This is presumed to be activated quickly enough to guide initial eye movements into a scene. However, early, short-latency eye movements in natural scenes have been shown to be dependent on the salience distribution across the image (Anderson, Ort, Kruijne, Meeter, & Donk, 2015). In the present work, we manipulated the salience distribution across a natural scene by changing the global contrast. We showed participants a brief real or nonsense preview of the scene and examined the time-course of eye movement guidance. A real preview decreased the latency and increased the amplitude of initial saccades into the image, suggesting that the preview allowed observers to obtain additional contextual information that would otherwise not be available. However, the preview did not completely override the initial tendency for short-latency saccades to be guided by the underlying salience distribution of the image. We discuss these findings in the context of oculomotor selection based on the integration of contextual information and low-level features in a natural scene.

Keywords Attention · Eye movements · Salience · Natural scene viewing · Contextual information

✉ Nicola C. Anderson
nccanderson@gmail.com

¹ Department of Cognitive Psychology, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

² Department of Education Science, VU University Amsterdam, Amsterdam, The Netherlands

Introduction

A key question in research on oculomotor behavior in natural scenes is how such eye movements are controlled. It is generally accepted that our eyes and attention can be influenced by both the stimulus itself and by more cognitive factors, such as knowledge and task goals. In research utilizing simple displays, stimulus features capture attention even when participants have a strong top-down goal (Godijn & Theeuwes, 2002; Hunt, von Muhlenen, & Kingstone, 2007; Siebold, van Zoest, & Donk, 2011; Zehetleitner, Koch, Goschy, & Muller, 2013). In research utilizing natural scenes, such bottom-up effects on attention have been studied by comparing fixated locations to a salience map, which quantifies the relative conspicuity of individual features in the visual field (Foulsham & Underwood, 2008; Itti & Koch, 2000; Koch & Ullman, 1985; Mannan, Ruddock, & Wooding, 1996; Peters, Iyer, Itti, & Koch, 2005; Reinagel & Zador, 1999). Even though this represents a straightforward approach, the observed correlations also may be accounted for by more cognitive influences, such as the participants' task (Anderson et al., 2015; Castelhana, Mack, & Henderson, 2009; Einhauser, Rutishauser, & Koch, 2008; Yarus, 1967), the presence of a bias to look in the center of an image (Tatler, 2007), the meaning of the scene (Foulsham & Underwood, 2011), or the correspondence between objects and salience (Einhäuser, Spain, & Perona, 2008; Nuthmann & Henderson, 2010).

To estimate how salience affects natural scene viewing behavior, we recently performed a study in which we manipulated the salience distribution across an image (Anderson et al., 2015; see also Einhauser et al., 2008). We asked participants either to memorize a scene or to search for a bull's-eye-shaped target. Critically, half of the scene was either reduced or increased in contrast relative to the other half, changing the overall distribution of salient regions across the image without

the confounding factors that have bedevilled other attempts to correlate salience with eye movement behavior (Einhäuser, Rutishauser, & Koch, 2008). We found that this contrast manipulation influenced where participants attended. When they initiated their first saccade into the scene quickly, within approximately 300 ms after the onset of the image, they were more likely to land on the region of higher contrast. Beyond 300 ms after the presentation of the image, however, participants were almost equally likely to go to either side of the image. These results suggest that while long-latency and subsequent saccades might be based on more goal-driven or cognitive influences, short-latency eye movements are salience-driven. However, salience may not necessarily be the only driving force behind rapid initial selection.

Unlike synthetic, uniform displays that enforce tight control on salience and object placement, natural scenes have complex spatial arrangements that are rich in meaning. The visual system has much practice in extracting contextual information, semantics and objects from pictures of the real world. Indeed, a lot of information can be gleaned from just a brief glance (50–250 ms) at a scene. It can rapidly provide information about a scene's structure (Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007) and semantics (Greene & Oliva, 2009). It is enough to establish the gist of a scene (Oliva & Torralba, 2006), its consistency (Davenport & Potter, 2004), and the presence and identity of some objects, people and animals (Fei-Fei, Iyer, Koch, & Perona, 2007; Thorpe, Fize, & Marlot, 1996). What is less well known, however, is to what extent such information influences eye movements occurring rapidly after the presentation of an image.

Torralba, Oliva, Castelano, and Henderson (2006) proposed that both salience and contextual information are computed in parallel and integrated early on, before the occurrence of a first eye movement. Local (salience) and global (context) features are assumed to quickly converge into a contextually modulated salience map that may potentially affect even the fastest eye movements. Castelano and Henderson (2007) and Vo and Henderson (2010) demonstrated that a briefly presented scene preview (as short as 50 ms), in conjunction with a prolonged delay between the preview and final image, reduced the time and the number of eye movements before a target was found. Already the first saccade into the scene was altered by the preview: its latency was reduced and its amplitude increased compared with a control condition without preview. This suggests that the preview allowed observers to extract contextual information that would otherwise not be available, which would run counter to the idea of Torralba et al. (2006) that contextual information is rapidly available. However, the scene preview also may have provided time for target knowledge to be integrated with the contextual representation extracted from the preview allowing subsequent eye movements to be quickly guided towards the

target (Vo & Henderson, 2010). This would bring the results in line with the contextual guidance model of Torralba et al. (2006): a preview may not necessarily yield contextual information that would otherwise be missing but may allow a faster integration of that information with target knowledge. We sought to differentiate between these two possibilities by investigating whether a preview would also affect the first saccade into an image if there was no explicit top-down search goal and thus no need for any target-context integration.

In the present work, we presented participants with a brief preview of a scene that was either a normal preview of the upcoming image, or a nonsense image. After a delay, sufficiently long to establish a strong contextual representation, the final image was presented. Unlike the normal preview, the final image was manipulated to be reduced in contrast on one side of the image. The scene preview was shown long enough, and with a significant delay before the final scene onset to establish a strong representation of any gist or conceptual or semantic scene knowledge (Vo & Henderson, 2010). For convenience, we will refer to any information gleaned from the scene preview to be “contextual,” although see Wu, Wick, and Pomplun (2014) for an excellent discussion of the different forms of semantic information this definition might entail. The task for participants was to remember the scene for a later memory test to allow for relatively free exploration of the images.

If contextual information is available early enough to guide even the fastest first eye movements in a scene (Torralba et al., 2006), the preview type should not further affect initial selection behavior for there is no need for any target-context integration in the present set-up. Accordingly, the contrast distribution of the final image should affect initial eye movements equally in the normal relative to the nonsense preview condition. Alternatively, if contextual information is more gradually acquired, then a real preview should reduce the impact of the contrast distribution in the final image relative to a nonsense preview. This would be the case because a normal preview should then lead to a change in the contextually modulated salience map, such that the relative influence of context becomes larger at the expense of salience. This should lead to a reduced salience effect in the normal preview relative to the nonsense preview condition.

Methods

Participants

Sixteen participants (ages 18–28 years, $M = 21.9$ years, 94% female) were recruited from VU University Amsterdam and participated in this experiment for course credit or 9 Euros. All reported normal or corrected to normal vision and were naive to 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 experiment was 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-inch (diagonal) Samsung Syncmaster 2233RZ with a resolution $1,680 \times 1,050$ pixels and refresh rate of 120 Hz at a viewing distance of 75 cm. Eye position was recorded via a second computer at 1,000 Hz with a spatial resolution of 0.01° visual angle using a 9-point calibration and validation procedure. The eye with the best spatial accuracy as determined by the calibration procedure was chosen for tracking. The online saccade detector of the eye tracker was set to detect saccades with amplitude of at least 0.5° , using an acceleration threshold of $9,500^\circ/\text{s}^2$ and a velocity threshold of $35^\circ/\text{s}$. 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

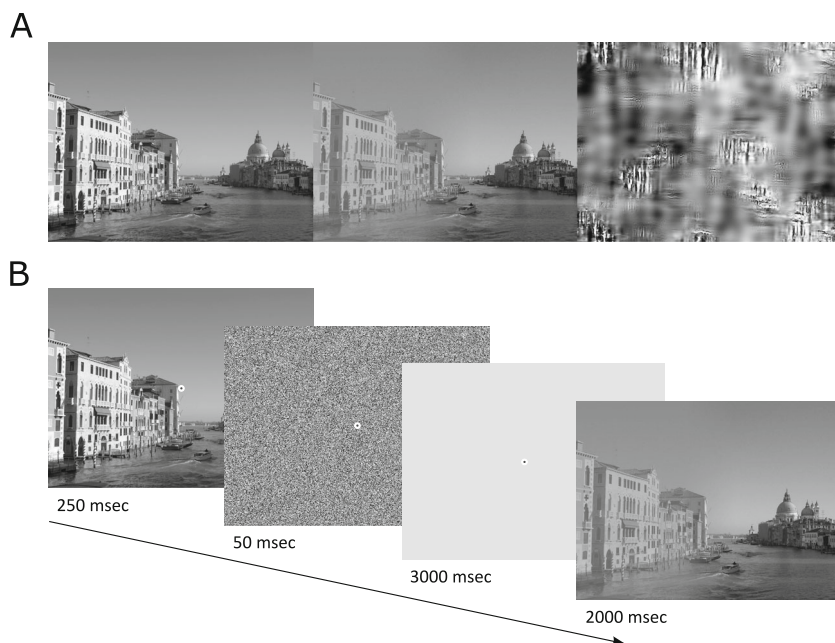
Images were selected from the SUN2012 Database (Xiao, Hays, Ehinger, Oliva, & Torralba, 2010) and from “Learning to Predict where Humans Look” (Judd, Ehinger, Durand, & Torralba,

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 $1,024 \times 768$ pixels and was converted to greyscale. Images were further selected such that their mean intensity values across the left and right side of the image conformed to a ratio of at most 3:4. Conforming to these selection criteria, 100 images were used from the SUN2012 Database and 91 images 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 degrees horizontal by 16 degrees vertical. The rest of the screen surrounding the image was gray.

Contrast adjustment

Image contrast adjustment was performed on all selected images using Matlab’s `imadjust` function (MATLAB, 2011). Intensity values on one side of the image were linearly remapped to a range spanning 40% of the original intensity range, in a way that left mean intensity unaffected. Only the variance in intensity was thus reduced. For each image, either the left or right 2/5 of the image width was manipulated in this fashion. For the center 1/5, contrast was reduced gradually, from full to reduced contrast, leaving the remaining 2/5 of the image width unmanipulated. Two versions of each image were created, either with the contrast gradually reduced from left to right or from right to left (Fig. 1a). Saliency maps were computed for each modified image using the Saliency Toolbox (Walther & Koch, 2006). Mean saliency was significantly lower for the reduced contrast side of each image ($M = 0.005$) compared with the original ($M = 0.03$), $t(381) = 38.59$, $p < 0.001$.

Fig. 1 **a** Example of an original grayscale image, the same image with reduced contrast on the left and the synthesized scrambled “nonsense” preview. **b** Schematic representation of an encoding trial with a normal preview



Scrambled images

A “scrambled” version of each image was created using a texture synthesis algorithm created by Portilla and Simoncelli (2000) but modified by Greene and Oliva (2009). The algorithm takes the image as input and calculates a number of image statistics, such as orientation and luminance information. It then coerces a noise stimulus to have the same properties. The resulting stimuli have similar low level perceptual features as the scene input stimulus, but without any object or spatial layout information (Greene & Oliva, 2009).

Procedure

Participants were seated with their head constrained in a chin rest and were given verbal and written instructions regarding the experimental procedure. Calibration and validation of their eye position was performed. In the first phase of the experiment, participants were instructed to explore the images carefully to remember the images for a later recognition task. Each trial began with a drift-correction screen in which participants were required to press the spacebar while fixating a centrally presented circular dot. Participants were then given either a preview of the image that they were about to see that was an unmanipulated version (i.e., without the contrast adjustment) of the final image, or a nonsense preview that was a scrambled version of the image (Fig. 1a). The preview or nonsense image was presented for 250 ms immediately followed by a 50-ms noise mask, then a 3-s uniform grey “integration” screen. The image was then presented for 2 s (Fig. 1b). The previews and integration screens contained a centrally presented circular fixation dot and participants were instructed to fixate on the dot for as long as it appeared onscreen. This dot disappeared when the final image was presented and participants were then free to move their eyes throughout the image. Figure 1b depicts a possible trial sequence.

We selected 150 images from the dataset, with the contrast adjusted on both the left and right. These images were presented in 300 trials with a mixed random order such that each participant saw the same image (with a different side adjusted) twice throughout the first phase of the experiment, with the restriction that participants could never see the second version of the image within the following three consecutive trials. The images were presented in 10 blocks and after each block participants were given feedback about their progress through the experiment.

At the end of the experiment, participants performed a recognition task on 20 images that had been presented in the initial phase and 20 new images (these images were randomly chosen from the image pool and were a mix of image categories). Each image was presented for 2 s, after which participants were asked to press the “z” key if they had seen the image before, or the “/” key if they had not seen the image before.

To familiarize participants with the task, participants performed 14 practice trials followed by 14 recognition trials (7

“old” images and 7 “new” images) before the experimental phase. Immediately after the practice trials, participants were given feedback on their performance in the recognition phase and had an opportunity to ask questions about the experimental procedure. At the conclusion of the experiment, participants were given feedback on their performance in the recognition task and then were asked to type in a field on the screen whether or not they had “noticed anything strange about the images.” This was done to check whether the contrast adjustment was noticed by the participants. The entire experiment lasted approximately 60 minutes.

Data processing

Fixations were removed if their duration was longer than 700 ms or faster than 120 ms or if they started outside a 65 pixel radius (approximately 1 degree visual angle) from the central fixation dot. These restrictions resulted in the removal of 2.6% of the trials.

The primary dependent measure was the proportion of initial saccades that landed in the higher contrast region. This region 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 that was gradually reduced in contrast at one side). A second dependent measure, saccade amplitude (in degrees visual angle) was calculated to assess the potential impact of the various preview types on saccadic targeting. A third measure analyzed was the latency of the initial saccade, calculated from the onset of the final image. We focus only on the first saccade into a scene for two reasons. First, it has been demonstrated previously that these eye movements are influenced by the contrast manipulation, whereas subsequent saccades are less affected by it (Anderson et al., 2015). Second, because we do not employ a gaze-contingent design when participants view the final image, we cannot make inferences about how any contextual representation built during the scene preview influences these eye movements.

Results

Performance

Seven out of the 16 participants indicated in the questionnaire that they had noticed the contrast manipulation.¹ Recognition accuracy was 91.6% ($SE = 1.74\%$).

¹ When submitted as a between subjects factor, whether people noticed the manipulation did not influence the proportion of eye movements that landed on the higher contrast region, $F < 1$, nor did it interact with preview, $F < 1$, bin, $F(1,14) = 1.54$, $MSE = .008$, $p = .234$, $\eta_p^2 = .099$, nor the preview by bin interaction, $F < 1$. This was also the case for saccade amplitudes, all F 's < 1 .

Proportion of eye movements to the higher contrast region

Figure 2 shows the proportion of saccades that landed on the higher contrast region of the image as a function of their saccade latency for both the normal and nonsense preview conditions. A 2 (preview type: nonsense vs. normal) by 4 (saccade latency bin) within-subjects analysis of variance was conducted on the proportion of first saccades that landed on the higher contrast region of the image, with the saccade latency factor treated as a linear contrast.

There was a main effect of preview type, $F(1, 15) = 30.91$, $MSE = 0.008$, $p < 0.001$, $\eta_p^2 = 0.673$, such that saccades landed more often on the higher contrast region when the preview was a nonsense image compared with when it was a normal image. There was a marginal main effect of saccade latency bin, $F(1, 15) = 3.85$, $MSE = 0.009$, $p = 0.069$, $\eta_p^2 = 0.204$, such that short-latency first saccades tended to land more often on the higher contrast side of the image than long-latency saccades. There was no interaction between preview type and saccade latency bin, $F(1, 15) = 0.34$, $MSE = 0.010$, $p = 0.569$, $\eta_p^2 = 0.022$.

Figure 2 suggests that when the preview was a normal image, selection of the higher contrast region may vary with latency in a non-linear fashion. To investigate the time course of selection performance across conditions, we performed a follow-up analysis of variance with saccade latency bin as a quadratic factor (Fig. 2). There was no quadratic main effect of bin, $F(1, 15) = 1.43$, $MSE = 0.004$, $p = 0.250$, $\eta_p^2 = 0.087$, but there was a significant interaction between preview type and latency bin, $F(1, 15) = 8.64$, $MSE = 0.006$, $p = 0.010$, $\eta_p^2 = 0.364$. This interaction likely results from the particular tendency, when the preview was a normal image, for the shortest-latency saccades (from the first saccade latency quartile; Fig. 2) more often to land on the higher contrast region of

the image than those from later latency quartiles. However, for each bin and across both normal and nonsense preview types, first saccades were more likely than chance to land on the higher contrast region, all $t > 4.87$, $p < 0.001$.

Saccadic amplitude

Figure 3 shows the amplitude of the first saccade as a function of its latency for both preview conditions. A 2 (preview type: nonsense vs. normal) by 4 (saccade latency bin) within-subjects analysis of variance was conducted on the first saccadic amplitude, with saccadic latency factor treated as a linear contrast.

There was a main effect of preview type, $F(1, 15) = 35.78$, $MSE = 0.222$, $p < 0.001$, $\eta_p^2 = 0.705$, such that saccade amplitude was significantly larger when the preview was a normal image, compared with when the preview was a nonsense image. There was no linear effect of saccade latency bin, $F(1, 15) = 3.02$, $MSE = 0.348$, $p = 0.013$, $\eta_p^2 = 0.168$, and no interaction between preview type and saccade latency bin, $F(1, 15) = 2.52$, $MSE = 0.192$, $p = 0.133$, $\eta_p^2 = 0.144$.

We also performed a follow-up analysis of variance with saccade latency bin as a quadratic factor on the first saccade amplitude. There was a quadratic main effect of latency bin, $F(1, 15) = 17.87$, $MSE = 0.060$, $p = 0.001$, $\eta_p^2 = 0.544$, but no interaction between preview type and saccade latency bin, $F(1, 15) = 2.74$, $MSE = 0.332$, $p = 0.119$, $\eta_p^2 = 0.154$.

Latency of the first saccade

The latency of the first saccade into the image was significantly shorter when the preview was a normal image ($M = 286$ ms; $SD = 25.21$ ms) than when the preview was a nonsense image ($M = 310$ ms; $SD = 26.59$ ms), $t(15) = 6.57$, $p < 0.001$.

Fig. 2 Proportion of saccades that landed in the higher contrast region of the image separately for each preview condition and saccade latency bin. Error bars in this and all subsequent figures represent standard error corrected for between-subjects variance (Cousineau, 2005; Morey, 2008)

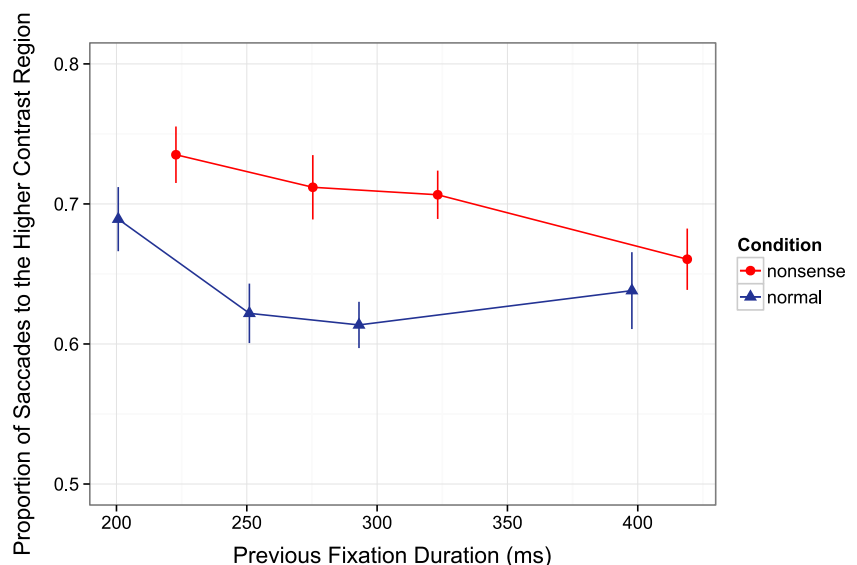
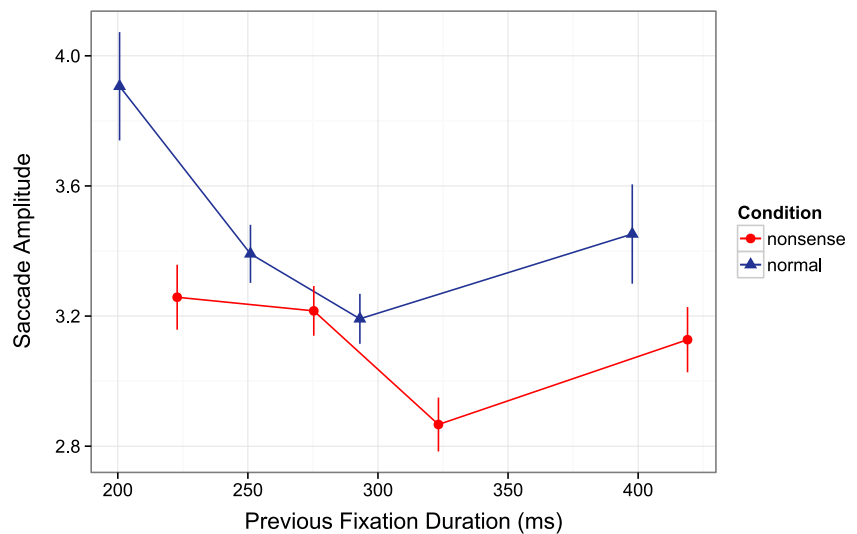


Fig. 3 Saccade amplitude separately for each preview condition and saccade latency bin



Discussion

The results of the present work revealed that a brief preview of an image was enough to influence saccadic programming to the extent that the tendency to move the eyes to the higher-contrast region was reduced relative to when the preview was a nonsense image. A preview of a natural scene significantly shortened the latency and increased the amplitude of the initial saccade. This finding extends previous work (Vo & Henderson, 2010) by demonstrating that a preview influences short-latency initial saccades into a scene. The present work additionally shows that this occurs even when there is no subsequent disruption to scene context (Vo & Henderson, 2010). More importantly, the results demonstrate that a preview affects initial saccades while observers were not engaged in any search task. This suggests that a preview effect may not only arise because the preview prolongs the available context-target integration time (Vo & Henderson, 2010) but also because it enhances the relative influence of context on oculomotor selection behavior (at the expense of the contribution of salience).

Nevertheless, the scene preview was not enough to override completely an early tendency to look toward the region of the image with a higher contrast, indicating that salience, to some extent, still affected oculomotor selection behavior. This finding contrasts with work suggesting that a contextual representation built during a 250-ms image preview should allow selection to be completely guided in a top-down fashion (Brockmole & Henderson, 2008; Vo & Henderson, 2010), and even when not provided with an image preview, much work suggests that goal-driven processes can rapidly influence eye movement behavior (Henderson, Malcolm, & Schandl, 2009; Neider & Zelinsky, 2006; Spotorno,

Malcolm, & Tatler, 2014; Torralba et al., 2006; Vo & Henderson, 2010).

Whereas the idea that stimulus salience may influence oculomotor control is against theories suggesting that selection behavior is driven predominantly by cognitive factors (Henderson, Brockmole, Castelhana, & Mack, 2007; Neider & Zelinsky, 2006; Spotorno, Malcolm, & Tatler, 2014), such earlier work has not investigated selection behavior on such a fine-grained timescale as in the present work (see also: Anderson et al., 2015; Mackay, Cerf, & Koch, 2012). Our results are in line with a view of oculomotor control that integrates both stimulus salience and knowledge structures (Navalpakkam & Itti, 2005; Torralba et al., 2006). Our findings here and in studies utilizing more simple displays (Donk & van Zoest, 2008; van Zoest & Donk, 2008; van Zoest, Donk, & Theeuwes, 2004) further constrain this relationship by suggesting that salience may be perceived as an emergent property of the speed at which individual objects are processed in the visual system. In this view, salience is coded in the temporal as well as spatial domain, where more conspicuous regions receive earlier activation than less conspicuous regions. This results in a selection bias for salient regions for early, fast responses, but eventually leads to a state of equivalence across locations that stand out from the background.

The present findings suggest that even though both salience and context determine selection behavior early on in scene viewing (Torralba et al., 2006), the relative contribution of context can be increased by the presentation of a real preview. This latter finding is in line with the idea that the contextual representation is acquired during a longer period of time than previously assumed (Torralba et al., 2006) and suggests that context, like salience exerts its influence through a dynamically changing representation in time.

Conclusions

We demonstrated that a brief preview of a natural scene provides some contextual guidance of eye movements. It can decrease the latency and increase the amplitude of the first saccade into a natural scene. In addition, it reduces the tendency for initial saccades to be guided by the saliency distribution of the scene. The contextual representation built during the preview, however, was not strong enough to completely override the influence of saliency, as saccades, particularly those with the shortest latency, were more likely than chance to land on the higher contrast side of the image.

Acknowledgments The authors thank Wouter Kruijne, Michelle Greene, and an anonymous reviewer for helpful comments on earlier versions of this manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Anderson, N. C., Ort, E., Kruijne, W., Meeter, M., & Donk, M. (2015). It depends on when you look at it: Saliency influences eye movements in natural scene viewing and search early in time. *Journal of Vision*, *15*(5), 9. doi:10.1167/15.5.9
- Brockmole, J. R., & Henderson, J. M. (2008). Prioritizing new objects for eye fixation in real-world scenes: Effects of object-scene consistency. *Visual Cognition*, *16*(2-3), 375–390. doi:10.1080/13506280701453623
- Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 753–763. doi:10.1037/0096-1523.33.4.753
- 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. doi:10.1167/9.3.6
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*(1), 42–45.
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. doi:10.1111/j.0956-7976.2004.00719.x
- Donk, M., & van Zoest, W. (2008). Effects of saliency are short-lived. *Psychological Science*, *19*(7), 733–739. doi:10.1111/j.1467-9280.2008.02149.x
- Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*(2), 2. doi:10.1167/8.2.2
- Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14), 18. doi:10.1167/8.14.18
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), 10. doi:10.1167/7.1.10
- Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2). doi:10.1167/8.2.6
- Foulsham, T., & Underwood, G. (2011). If Visual Saliency Predicts Search, Then Why? Evidence from Normal and Gaze-Contingent Search Tasks in Natural Scenes. *Cognitive Computation*, *3*(1), 48–63. doi:10.1007/s12559-010-9069-9
- Godijn, R., & Theeuwes, J. (2002). Programming of endogenous and exogenous saccades: Evidence for a competitive integration model. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1039–1054. doi:10.1037//0096-1523.28.5.1039
- Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, *20*(4), 464–472.
- Henderson, J. M., Brockmole, J. R., Castelano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford, UK: Elsevier
- Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, *16*(5), 850–856. doi:10.3758/PBR.16.5.850
- Hunt, A. R., von Muhlenen, A., & Kingstone, A. (2007). The time course of attentional and oculomotor capture reveals a common cause. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 271–284. doi:10.1037/0096-1523.33.2.271
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489–1506. doi:10.1016/S0042-6989(99)00163-7
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286–3297. doi:10.1016/j.visres.2007.09.013
- Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. *IEEE International Conference on Computer Vision (ICCV)*. doi:10.1109/ICCV.2009.5459462
- Koch, C., & Ullman, S. (1985). Shifts in Selective Visual-Attention - Towards the Underlying Neural Circuitry. *Human Neurobiology*, *4*(4), 219–227.
- Mackay, M., Cerf, M., & Koch, C. (2012). Evidence for two distinct mechanisms directing gaze in natural scenes. *Journal of Vision*, *12*(4), 9. doi:10.1167/12.4.9
- Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*(3), 165–188. doi:10.1163/156856896x00123
- Mathot, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. doi:10.3758/s13428-011-0168-7
- MATLAB. (2011). version 7.12.0 (R2011a). Natick, Massachusetts: The MathWorks Inc.
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*(2), 205–231. doi:10.1016/j.visres.2004.07.042
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621. doi:10.1016/j.visres.2005.08.025
- Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8), 20. doi:10.1167/10.8.20

- Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Visual Perception, Pt 2: Fundamentals of Awareness: Multi-Sensory Integration and High-Order Perception*, 155, 23–36. doi: [10.1016/S0079-6123\(06\)55002-2](https://doi.org/10.1016/S0079-6123(06)55002-2)
- Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, 45(18), 2397–2416. doi:[10.1016/j.visres.2005.03.019](https://doi.org/10.1016/j.visres.2005.03.019)
- Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1), 49–71. doi:[10.1023/A:1026553619983](https://doi.org/10.1023/A:1026553619983)
- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, 10(4), 341–350. doi:[10.1088/0954-898x/10/4/304](https://doi.org/10.1088/0954-898x/10/4/304)
- Siebold, A., van Zoest, W., & Donk, M. (2011). Oculomotor Evidence for Top-Down Control following the Initial Saccade. *Plos One*, 6(9). doi: [10.1371/journal.pone.0023552](https://doi.org/10.1371/journal.pone.0023552)
- Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision*, 14(2), 7. doi: [10.1167/14.2.7](https://doi.org/10.1167/14.2.7)
- Tatler, B. W. (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. doi:[10.1167/7.14.4](https://doi.org/10.1167/7.14.4)
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520–522. doi:[10.1038/381520a0](https://doi.org/10.1038/381520a0)
- 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(4), 766–786. doi:[10.1037/0033-295X.113.4.766](https://doi.org/10.1037/0033-295X.113.4.766)
- van Zoest, W., & Donk, M. (2008). Goal-driven modulation as a function of time in saccadic target selection. *Quarterly Journal of Experimental Psychology*, 61(10), 1553–1572. doi:[10.1080/17470210701595555](https://doi.org/10.1080/17470210701595555)
- van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 746–759. doi:[10.1037/0096-1523.30.4.749](https://doi.org/10.1037/0096-1523.30.4.749)
- Vo, M. L., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, 10(3), 14. doi:[10.1167/10.3.14](https://doi.org/10.1167/10.3.14)
- Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 19(9), 1395–1407. doi:[10.1016/j.neunet.2006.10.001](https://doi.org/10.1016/j.neunet.2006.10.001)
- Wu, C. C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, 5. doi: [10.3389/Fpsyg.2014.00054](https://doi.org/10.3389/Fpsyg.2014.00054)
- Xiao, J., Hays, K., Ehinger, A., Oliva, A., & Torralba, A. (2010). SUN Database: Large-scale scene recognition from Abbey to zoo. *IEEE Conference on Computer Vision and Pattern Recognition*. doi:[10.1109/CVPR.2010.5539970](https://doi.org/10.1109/CVPR.2010.5539970)
- Yarbus, A. L. (1967). *Eye Movements and Vision*. New York: Plenum Press.
- Zehetleitner, M., Koch, A. I., Goschy, H., & Muller, H. J. (2013). Salience-Based Selection: Attentional Capture by Distractors Less Salient Than the Target. *Plos One*, 8(1), e52595. doi:[10.1371/journal.pone.0052595](https://doi.org/10.1371/journal.pone.0052595)