OBSERVATION

The Art of Gaze Guidance

Nida Latif
Queen’s University

Arlene Gehmacher
Royal Ontario Museum, Ontario, Ontario, Canada

Monica S. Castelhano and Kevin G. Munhall
Queen’s University

An ongoing challenge in scene perception is identifying the factors that influence how we explore our visual world. By using multiple versions of paintings as a tool to control for high-level influences, we show that variation in the visual details of a painting causes differences in observers’ gaze despite constant task and content. Further, we show that by switching locations of highly salient regions through textural manipulation, a corresponding switch in eye movement patterns is observed. Our results present the finding that salient regions and gaze behavior are not simply correlated; variation in saliency through textural differences causes an observer to direct their viewing accordingly. This work demonstrates the direct contribution of low-level factors in visual exploration by showing that examination of a scene, even for aesthetic purposes, can be easily manipulated by altering the low-level properties and hence, the saliency of the scene.

Keywords: saliency, gaze, art perception

Supplemental materials: http://dx.doi.org/10.1037/a0034932.supp

When we look at art in a museum, we find ourselves in what the writer Isaac Bashevis Singer called the chasm between the artist’s inner vision and its ultimate expression. Our work here shows that the fine details of the artist’s expression can guide visual exploration in this aesthetic task. Paintings and their visual examination thus may reveal the operating principles of the human visual system.

It is well known that the human visual system is attracted to some properties over others. Salient visual features such as local differences in contrast, orientation, color, and texture have shown to collectively predict gaze in natural scenes (Itti & Koch, 2001). In fact, the early primate visual system presents a common cross-species response to these visually prominent regions through a topographically coded saliency representation or map. Here we utilize the Itti and Koch model of saliency to analyze differences in images. This model combines a subset of low-level properties channels, explicitly highlighting regions that are locally different.

In dynamic tension with these bottom-up influences are the influences of the viewer’s priorities (see Henderson, Brockmole, Castelhano, & Mack, 2007, for review). These active aspects of visual attention allow the viewer to select areas of the scene that are consistent with ongoing goals. The interaction between bottom-up and top-down influences on visual activity are still not well understood. However, we know that visual attention is not affected by these influences independently1 (Adams & Mamassian, 2004).

One approach to examining determinants of gaze behavior is to hold the scene constant and vary the viewer’s task (Yarbus, 1967). Here we take the opposite approach: we hold the task and semantic content constant and vary saliency by manipulating the spatial frequency and hence the texture2 of an image. Although a correlational relationship between high-spatial frequency content and the probability of an individual directing their gaze to that region has been established (Mannan, Ruddock, & Wooding, 1997), causal relationship has not been adequately determined. Some studies that have demonstrated a causal relationship (Baddeley & Tatler, 2006; Einhäuser & König, 2003) have done so in an abstract manner by utilizing global or randomly located manipulation in natural scenes. Such studies allow us to understand the

---

1 For a review of computational saliency models that attempt to incorporate both high- and low-level influence gaze behavior; see Borji, Sihite, & Itti, 2012.

2 It has been shown that the perception of texture is heavily influenced by the spatial frequency components of an image. Similar channels within the saliency model have been implicated (i.e., luminance contrast: Mannan, Ruddock, & Wooding, 1997; Parkhurst, Law, & Neibur, 2002).
underlying mechanism of how the visual system may respond to saliency; however, further work is needed to incorporate the interaction between low-level and high-level influences. More specifically, expanding on findings that indicate how gaze may be directed toward salient regions that are also semantically informative (Enns & MacDonald, 2012) is important to understanding this interaction.

We examine the causal relationship between saliency, as indicated by the Itti and Koch (2001) model, and gaze by exploring naturally occurring saliency differences in multiple copies of the same painting (i.e., same semantic content). By maintaining semantic content, we demonstrate that how an observer viewed artwork was, in part, driven by saliency. Moreover, directly changing saliency of semantically informative regions by manipulating texture causes observers to redirect their viewing of the same piece of art.

**Experiment 1**

The stimulus used in this study is the historical painting *The Death of General Wolfe* (DoW), which depicts the moment of death of British General James Wolfe in 1759. This painting represents an event in Canadian history, a significant British victory over France in Quebec during the Seven Years’ War. Featuring many figures both historical and allegorical, the artist Benjamin West (1738–1820) dramatized the event with such artistic vision that his work became a public sensa tion. High demand for this painting resulted in West creating multiple copies, including one commissioned by the Royal Family (von Erfia & Staley, 1986). The multiple versions evidence variations, although the general historical theme remained consistent. This example of variation in reproduction provides an effective tool to examine whether saliency guides gaze while controlling the semantic content.

Using a viewing task with constant task demands, we varied the test image between three groups so that each group viewed a different version of the DoW painting. Fixations (i.e., maintenance of gaze at a single location) were examined to determine whether they reflected the physical differences between versions. Because saliency is most predictive of gaze in the early part of viewing an image (Underwood & Foulsham, 2006), we analyzed the first 2 s of viewing. There was an observable correspondence between the salient regions and fixation spread and duration (Figure 1C). We observed a difference in average fixation duration among the three versions of the DoW painting, $F(2, 45) = 4.41, p = .01, n_g^2 = .23$. Figure 2A shows that average fixations were shorter for the NG version ($M_{diff} = 55.0\text{ ms}, SE = 19.7, p < .01$). Because all three groups viewed the same filler paintings, we used them as a control and found no significant differences in the average fixation duration (Figure 2A).

We calculated the average saccadic length to examine spread of fixations. Results indicated that average saccadic length differed significantly between versions, $F(2, 45) = 5.27, p < .01, n_g^2 = .33$ (Figure 2B). Participants looking at the ROM version had a longer average saccadic length than the NG version ($M_{diff} = 1.63\text{°}, SE = 0.53, p < .01$). The fillers indicated no significant differences in average saccadic length. These behavioral differences did not influence participants’ familiarity and aesthetic preference, and no significant differences were observed in participants’ subjective ratings.

Because the semantic content was maintained, Experiment 1 demonstrated that the physical differences across versions were...
affecting gaze patterns in a manner that can be explained by saliency differences. Experiment 2 was conducted to directly manipulate an aspect of saliency (texture) and determine if gaze can be causally affected by this manipulation.

Experiment 2

In Experiment 2, we wanted to demonstrate that saliency variation was directly related to gaze differences seen in Experiment 1. Saliency maps demonstrated that all versions shared a common region near the main character, General Wolfe. However, the number and strength of additional regions was greater for the ROM version (Figure 1A). We predicted that these additional regions were causing participants’ gaze to be drawn to more areas of the painting for longer periods. We selected and manipulated the saliency maps of the ROM and the NG paintings because of their distinct saliency patterns and behavioral responses in Experiment 1. We switched the saliency pattern of each of the versions using sharpening and blurring filters to simulate artistic textural modification (DiPaola, Riebe, & Enns, 2010). We predicted that switching saliency patterns through textural manipulation would switch gaze patterns during viewing. Our logic was as follows: our textural manipulation produced a change in the output of the saliency model (Itti & Koch, 2001) and thus an assumed change to prioritized regions in observers’ visual systems. Gaze corresponding with these prioritized regions was predicted to align with saliency differences.

Methods

Participants. Sixty (49 women, 11 men, $M_{age} = 19.0$ years) undergraduates with normal or corrected-to-normal vision participated for course credit or for $10/hr$.

Stimuli and apparatus. The eye tracking apparatus was the same as that used in Experiment 1.

Figure 1. The three versions of The Death of General Wolfe (DoW) selected for presentation as stimuli in Experiment 1: National Gallery of Canada (NG), National Trust (NT), and Royal Ontario Museum (ROM). (A) The saliency maps for the three versions as calculated by the Itti and Koch (2001) computational model. (B) The three versions of DoW presented. (C) The gaze patterns for each of the three versions. The size of the circle represents the duration of each fixation. Fixations presented were only for the first 2 s of viewing. Images were printed with permission: (1) Benjamin West, The Death of General Wolfe, 1770, oil on canvas, 152.5 × 214.5 cm, Transfer from the Canadian War Memorials, 1921 (Gift of the 2nd Duke of Westminster, England, 1918), National Gallery of Canada, Ottawa, Photo © National Gallery of Canada; (2) The Death of General James Wolfe (1727–1759) by Benjamin West PRA (Swarthmore 1738–London 1820), at Ickworth, Suffolk, United Kingdom. © National Trust Images/J. Whitaker; and (3) The Death of General Wolfe, Benjamin West, 1776/1806 © Royal Ontario Museum.
Using the surface blurring filter and the unsharp masking filter in Adobe Photoshop CS5 to simulate texture, high-saliency regions were blurred and low-saliency regions were sharpened using a masking radius of 5–30 pixels at 20%. So the saliency map of regions were blurred and low-saliency regions were sharpened using an Adobe Photoshop CS5 to simulate texture. High-saliency regions were captured in saliency maps with the contrast and edge detection channels of the Itti and Koch (2000) saliency model. We presented the finding that saliency and gaze are causally related, a particularly striking result in an unconstrained, aesthetic viewing task. Our first study demonstrated that natural variation within versions of the same painting produced variation in partic-

Results

The results from Experiment 1 were replicated in Experiment 2. There was a significant difference in average fixation duration, $F(3, 56) = 3.02, p = .04, \eta^2_p = .12$, and average saccadic length, $F(3, 56) = 4.67, p < .01, \eta^2_p = .10$, between painting versions. Individuals demonstrated longer average fixation duration ($M_{\text{Diff}} = 56.0 \text{ ms}, SE = 23.1, p = .01$) and greater average saccadic length ($M_{\text{Diff}} = 2.36^\circ, SE = 0.95, p = .02$) for the Original ROM version compared with the Original NG version. As expected, the switch in saliency patterns caused a switch in gaze patterns. The Modified NG demonstrated greater average fixation duration ($M_{\text{Diff}} = 53.1 \text{ ms}, SE = 23.4, p = .02$) and average saccadic length ($M_{\text{Diff}} = 3.43^\circ, SE = 0.94, p < .01$) than the Original NG version but showed no difference from the original ROM version that it was manipulated to resemble. The Modified ROM showed a nonsignificant trend in fixation duration ($M_{\text{Diff}} = 34.5, SE = 24.3, p = .09$) compared with the Original ROM version but showed no difference from the Original NG version it was modified to resemble (Figure 4A and 4B).

A repeated-measures analysis of the ordinal saccadic length for the first five fixations was performed for a more detailed examination of the effect of our textural manipulation. This analysis demonstrated a significant effect of fixation number, $F(4, 45) = 5.059, p < .01$ (Figure 4C). By the fifth fixation, the Original ROM and the Modified NG demonstrated a greater saccadic length than the Original NG and the Modified ROM, and as viewing progressed, differences became more pronounced.

The results of Experiment 2 demonstrated that manipulating saliency through texture affected how participants viewed the DoW painting. Additional highly salient regions predicted the average saccadic length and the average fixation duration. Moreover, there was a visible correspondence between gaze and the relative locations of salient regions (see Supplemental Figure S1). Reversing the saliency between the two versions caused a reversal of the gaze patterns.

General Discussion

We presented the finding that saliency and gaze are causally related, a particularly striking result in an unconstrained, aesthetic viewing task. Our first study demonstrated that natural variation within versions of the same painting produced variation in partic-

---

7 Surface blurring was used to preserve edges while blurring the contents. Hence, blurring of highly detailed regions, such as facial regions, maintained their features. Unsharp masking is used to amplify high-frequency components of an image by suppressing the low-frequency components. Photographic blurring has been suggested to approximate high spatial frequency attenuation (Watson & Ahumada, 2011). Because it has already been demonstrated that spatial frequency content is related to texture, we infer that these methods would manipulate texture accordingly.

8 Texture is related to a higher spatial frequency content, which can be captured in saliency maps with the contrast and edge detection channels of the Itti and Koch (2000) saliency model.

9 The radius refers to the size of the edges to be enhanced or dampened. The range used accommodates both small- and large-scale manipulations. The amount listed as a percent controls the magnitude of contrast applied to the edges.

10 Gaussian blurring is not detected by the Itti and Koch (2000) saliency model and thus did not influence the saliency measures.

11 The results for this memory task will not be presented in this article.
Participants’ gaze patterns. Our second study showed that manipulation of an aspect of saliency caused participants’ gaze to change systematically.

Experiment 1 showed that gaze variations caused by physical differences in the paintings were consistent with the visual saliency hypothesis: our visual system is sensitive to differences in low-level dimensions such as color, intensity, and edge orientation (Itti & Koch, 2000), and these properties are predictive of saccadic movements. The behavior observed in Experiment 1 cannot be associated with the cognitive scene processing (e.g., Henderson et al., 2007) because the semantic content differences are subtle; a low-level, property-driven explanation is more appropriate. We further supported this hypothesis by using the Itti and Koch (2000) saliency model to demonstrate that the versions were different in saliency content and suggested that gaze differences were a result of the variation.

In Experiment 2, we used textural manipulation to explore whether image saliency directly affected how participants viewed the paintings. Textural manipulation was selected because high-texture areas (where the spatial frequency is relatively greater than surrounding regions) have shown to direct gaze in natural scenes (Baddeley & Tatler, 2006). In addition, many artists have indicated their awareness of using texture to guide viewers’ gaze toward the center of focus, an effective tool because highly textured areas are often associated with more meaningful aspects of a scene (Henderson et al., 2007). We demonstrated this association by showing that naturally occurring salient regions, and hence, our manipulations were not random; they corresponded with semantically informative regions such as various characters that may be more meaningful. This is consistent with the finding that high spatial frequency regions are prioritized when perceiving semantically meaningful information (Emms & MacDonald, 2012).

Figure 3. Original and modified saliency maps for the National Gallery of Canada (NG) and the Royal Ontario Museum (ROM) versions along with corresponding gaze patterns. (A) Gaze patterns and saliency patterns for the original NG and ROM versions of the painting. Size of circles indicate duration of fixations. (B) Gaze patterns and saliency maps for the version of NG modified to look like the ROM and the ROM modified to look like the NG version. The size of circles indicate duration of fixation. Note the switch in patterns with textural modification. The data displayed for all four groups were for the first five fixations of the experiment.
The textural component of an image is also involved in the recognition of stimuli; texture and clarity may direct attention to familiar objects. This confluence of high texture and importance may result in an aesthetic preference and attention to these regions (Zeki, 1999). In terms of art perception, we may prioritize such regions because they signal where an observer ought to look from the perspective of the artist (Enns & MacDonald, 2012). Accordingly, results of Experiment 2 confirmed the causal relationship between gaze and texture. By switching the saliency patterns between versions, we were able to switch the gaze behavior.

As a caveat, although we demonstrate causality, at no point do we suggest that bottom-up influences are completely independent. Instead, we focused our investigation based on our current knowledge of high-level influences in scene perception, that is, by using the highly cognitive task of aesthetic evaluation (Zeki, 1999) and using semantically coupled saliency manipulations (Enns & MacDonald, 2012). Further investigation of the interaction between these influences is required. Returning to the results, these data beg the question of why the paintings vary in saliency. Artists are able to present the visual world in a manner that exploits the human visual system. However, our results do not allow us to confirm West’s intentionality in introducing differences in these particular paintings. In addition to artistic composition, there may be other explanations for appearance and saliency alteration: aging under different conditions and multiple instances of restoration/preservation by both the artist and subsequent caretakers (e.g., removal/addition of varnish, cleaning and paint-loss filling; Daly-Hartin & Vogel, 2003). Although the source of differences is unclear, the differences in gaze patterns caused by textural modification are apparent. Further historical research is needed to understand the artist’s true intent, but based on our data, we suggest that artists can use artistic techniques to guide art viewing much like our experimental manipulation.

Our studies demonstrated how an artist could exploit the operating principles of the visual system by contrasting artistic techniques used to modify saliency with the high-level features of a scene’s meaning. Artistic compositions provide an opportunity to study scene perception in a manner naturally unbiased by explicit predetermined tasks, though not excluding implicit goals of aesthetic appreciation and scrutiny. Even in the relatively high-level cognitive task of art observation, visual attention is influenced by the low-level properties of saliency. Our demonstrated preference for such salient features may intuitively compel artists to use techniques to direct the gaze of their audience.

Figure 4. (A) Average fixation duration for the two original and two modified versions of the The Death of General Wolfe (DoW) paintings. Error bars = SE. (B) Average saccadic length for the two original and two modified versions of the DoW paintings. Error bars = SE. (C) Average ordinal saccadic length for the first five fixations in the original and modified versions of the DoW paintings. Error bars = SE.

The textural component of an image is also involved in the recognition of stimuli; texture and clarity may direct attention to familiar objects. This confluence of high texture and importance may result in an aesthetic preference and attention to these regions (Zeki, 1999). In terms of art perception, we may prioritize such regions because they signal where an observer ought to look from the perspective of the artist (Enns & MacDonald, 2012). Accordingly, results of Experiment 2 confirmed the causal relationship between gaze and texture. By switching the saliency patterns between versions, we were able to switch the gaze behavior.

As a caveat, although we demonstrate causality, at no point do we suggest that bottom-up influences are completely independent. Instead, we focused our investigation based on our current knowledge of high-level influences in scene perception, that is, by using the highly cognitive task of aesthetic evaluation (Zeki, 1999) and using semantically coupled saliency manipulations (Enns & MacDonald, 2012). Further investigation of the interaction between these influences is required. Returning to the results, these data beg the question of why the paintings vary in saliency. Artists are able to present the visual world in a manner that exploits the human visual system. However, our results do not allow us to confirm West’s intentionality in introducing differences in these particular paintings. In addition to artistic composition, there may be other explanations for appearance and saliency alteration: aging under different conditions and multiple instances of restoration/preservation by both the artist and subsequent caretakers (e.g., removal/addition of varnish, cleaning and paint-loss filling; Daly-Hartin & Vogel, 2003). Although the source of differences is unclear, the differences in gaze patterns caused by textural modification are apparent. Further historical research is needed to understand the artist’s true intent, but based on our data, we suggest that artists can use artistic techniques to guide art viewing much like our experimental manipulation.

Our studies demonstrated how an artist could exploit the operating principles of the visual system by contrasting artistic techniques used to modify saliency with the high-level features of a scene’s meaning. Artistic compositions provide an opportunity to study scene perception in a manner naturally unbiased by explicit predetermined tasks, though not excluding implicit goals of aesthetic appreciation and scrutiny. Even in the relatively high-level cognitive task of art observation, visual attention is influenced by the low-level properties of saliency. Our demonstrated preference for such salient features may intuitively compel artists to use techniques to direct the gaze of their audience.

Figure 4. (A) Average fixation duration for the two original and two modified versions of the The Death of General Wolfe (DoW) paintings. Error bars = SE. (B) Average saccadic length for the two original and two modified versions of the DoW paintings. Error bars = SE. (C) Average ordinal saccadic length for the first five fixations in the original and modified versions of the DoW paintings. Error bars = SE.

The textural component of an image is also involved in the recognition of stimuli; texture and clarity may direct attention to familiar objects. This confluence of high texture and importance may result in an aesthetic preference and attention to these regions (Zeki, 1999). In terms of art perception, we may prioritize such regions because they signal where an observer ought to look from the perspective of the artist (Enns & MacDonald, 2012). Accordingly, results of Experiment 2 confirmed the causal relationship between gaze and texture. By switching the saliency patterns between versions, we were able to switch the gaze behavior.

As a caveat, although we demonstrate causality, at no point do we suggest that bottom-up influences are completely independent. Instead, we focused our investigation based on our current knowledge of high-level influences in scene perception, that is, by using the highly cognitive task of aesthetic evaluation (Zeki, 1999) and using semantically coupled saliency manipulations (Enns & MacDonald, 2012). Further investigation of the interaction between these influences is required. Returning to the results, these data beg the question of why the paintings vary in saliency. Artists are able to present the visual world in a manner that exploits the human visual system. However, our results do not allow us to confirm West’s intentionality in introducing differences in these particular paintings. In addition to artistic composition, there may be other explanations for appearance and saliency alteration: aging under different conditions and multiple instances of restoration/preservation by both the artist and subsequent caretakers (e.g., removal/addition of varnish, cleaning and paint-loss filling; Daly-Hartin & Vogel, 2003). Although the source of differences is unclear, the differences in gaze patterns caused by textural modification are apparent. Further historical research is needed to understand the artist’s true intent, but based on our data, we suggest that artists can use artistic techniques to guide art viewing much like our experimental manipulation.

Our studies demonstrated how an artist could exploit the operating principles of the visual system by contrasting artistic techniques used to modify saliency with the high-level features of a scene’s meaning. Artistic compositions provide an opportunity to study scene perception in a manner naturally unbiased by explicit predetermined tasks, though not excluding implicit goals of aesthetic appreciation and scrutiny. Even in the relatively high-level cognitive task of art observation, visual attention is influenced by the low-level properties of saliency. Our demonstrated preference for such salient features may intuitively compel artists to use techniques to direct the gaze of their audience.

References


ART OF GAZE GUIDANCE


Received May 8, 2013
Revision received September 11, 2013
Accepted September 13, 2013