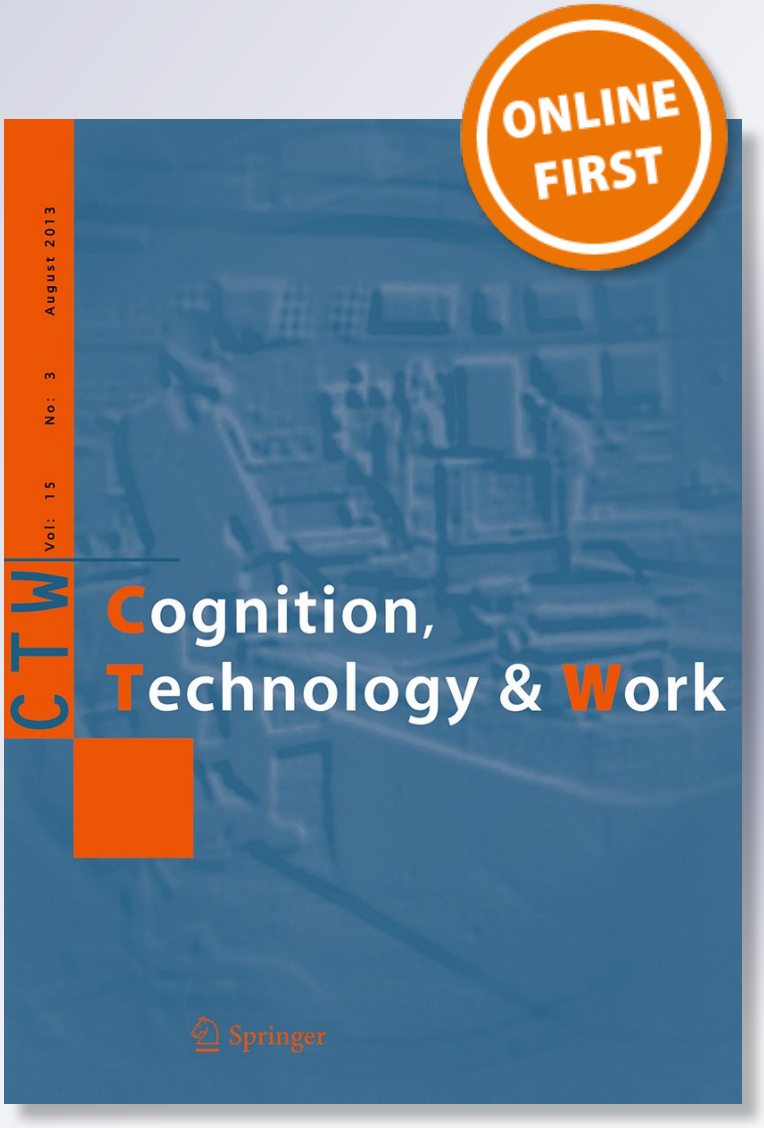
Web page attentional priority model

Jeremiah D. Still

Cognition, Technology & Work

ISSN 1435-5558

Cogn Tech Work
DOI 10.1007/s10111-017-0411-9



Springer

Cogn Tech Work
DOI 10.1007/s10111-017-0411-9

ORIGINAL ARTICLE

Web page attentional priority model

Jeremiah D. Still[1]

Received: 8 March 2017 / Accepted: 29 May 2017

Abstract Designing an interface that is both information rich and easy to search is challenging. Successfully finding a solution depends on understanding an interface's explicit and implicit influences. A cognitively inspired computational approach is taken to make the implicit influences apparent to designers. A saliency model has already been shown to predict the deployment of attention within web page interfaces. It predicts regions likely to be salient, based on local contrast stemming from the bottom-up channels (e.g., color, orientation). This research replicates these previous findings and extends the work by proposing a web page-specific attentional priority (AP) model. This AP model includes previous interaction experience history, manifested as conventions, within the already valuable saliency model. These sources of influence automatically nudge our attention to regions that usually contain useful visual information. This research shows that, by integrating spatial conventions with a saliency model, designers can better predict the deployment of attention within web page interfaces.

Keywords Computational model · Eye movements · Salience · Human–computer interaction · Design

✉ Jeremiah D. Still
jstill@odu.edu

[1] Department of Psychology, Old Dominion University, 250 Mills Godwin Building, Norfolk, VA 23529-0267, USA

1 Overview

Complex interface designs are imbedded in our everyday activities. We are dependent on quick and easy visual searches within these ubiquitous interfaces. Designing a successful interface depends on an understanding of how the interface will influence users. In this work, I focus on the cognitive influences that implicitly guide users' visual search. While designers are able to consciously perceive the reflective, explicit, influences—like the information architecture within a website—the implicit influences (like visual saliency) are difficult to recognize. The goal of this research is to understand how the visual elements within an interface implicitly guide attention and make this influence visually available to developers. Implicit processing plays a key role in allowing our cognitive system to overcome our physical and processing limitations in order to operate effectively within an overwhelming environment. Our ability to automatically recognize patterns, with no conscious insight into the process of our pattern recognition, is one example of this. Saliency is another implicit process; it nudges us to certain interface regions over others in an attempt to make visual searches more efficient. For example, we can easily and quickly recognize a visually salient object within a cluttered interface, but just as we are unable to describe how a pattern was recognized, we are unable to describe how we detected that object. Because implicit processes are not available for conscious introspection, we need a method that allows developers to quickly measure the saliency within complex displays. One possible solution is the employment of a computational model that can quantify display influences. A visual saliency model provides one possible method for predicting which visual elements will draw attention. Prior research suggests that a saliency model can predict fixations within

Published online: 06 June 2017

web pages (Still and Masciocchi 2010, 2012; Masciocchi and Still 2013). In this research, an attentional priority (AP) model that combines a computational saliency model prediction with web page-specific spatial biases (c.f., schema or convention) is proposed. Website-specific biases can manifest as conventions, learned from common design patterns, which implicitly guide users' expectations. Conventions are best elicited by observing users interactions with an interface (Still et al. 2015). Therefore, eye-tracking data were used to extract website-specific conventions. The current cognitive science literature suggests that integrating both salience (external factor) and expectations (internal factor) produces the best performing predictive models (Awh et al. 2012).

2 Background

Designers can produce easier-to-search interfaces by making common task elements higher in visual saliency. Within Cognitive Psychology, it has been shown that colocating task-relevant information with salient visual features reduces search times and often facilitates successful task completion (Wolfe 2007). This is accomplished by implicitly communicating to viewers where they ought to begin their visual search. Salient features are those that are visually unique, relative to their surroundings (Parkhurst et al. 2002). For example, text that is underlined amid non-underlined text is salient and implicitly attracts the reader's attention. However, many interfaces are rich with visual media (e.g., text, pictures, logos), which can make the determination of salient features a complicated task. Given this complexity, designers are often forced to make their best guesses about which regions within an interface are salient; this can lead to costly iterative design cycles and more effortful interactions. The following review of the role that attention plays in perception, the factors that guide it, and the historical approach to predicting hierarchical-based searches within an interface will provide background for understanding the unique and important contributions that the proposed AP model offers.

2.1 Selective attention and guiding influences

The human perceptual system is only able to process a very limited amount of the overwhelming amount of information being presented in a given environment. According to Cowan (2000), human working memory, our conscious workbench, is limited to the simultaneous processing of 4 ± 1 chunks of information. Thus, attention must act as a selective mechanism, controlling what information is held in working memory (Johnston and Dark 1986). In a visual search, only regions within a scene that are likely to support ongoing task needs are selected and long-term memory schemas are activated (Norman and Shallice 1986). Attention has to overcome physical restrictions, as well. Specifically, only information that lands on the fovea—the retinal region that contains a dense concentration of cone photoreceptors—receives high-resolution processing, and it only covers approximately two degrees of central vision. Because of these restrictions, attention is often described as a "spotlight" (Fernandez-Duque and Johnson 2002).

Our perception of the world around us as coherent and seamless is a reflection of working memory, attention, and long-term memory playing in concert. Situation-appropriate long-term memories are activated. Working memory then integrates and updates those long-term memories with new incoming visual information (Baddeley 1992; Cowan 1988). This combination of previous experiences and spotlight scanning tricks us into believing that our perceptions of the world are complete. For instance, the change blindness paradigm demonstrates that viewers miss critical information even within a single static image. In this paradigm, a viewer is asked to detect a major change between two pictures of the same scene (e.g., a missing tree). If our perception of a scene were complete and accurate, detecting the change between the pictures would be quick and effortless. However, the detection of the change is neither easy nor quick. We must systematically process only fragments of the picture and store them for comparison in working memory. An understanding of what guides the spotlight of attention is valuable for scientists interested in making products easier to search.

Our spotlight of attention is guided by two major low-level and early influences. The first is visual saliency (Itti et al. 1998), which impacts the general population uniformly (unless a user's visual processing is abnormal, e.g., in colorblindness). The second is a relevant search schema based on previous interaction experiences (Chun 2000; Malcolm and Henderson 2010). This influence is culturally based, often reflecting design conventions (Still and Dark 2010, 2013); it prioritizes the selection of regions based on the need for rapid recognition of a familiar context.

According to Desimone and Duncan's (1995) neurologically based biased competition model, the guidance of attention reflects an interaction between early processes (i.e., bottom-up) and later directed processes (top-down, e.g., defining features of the target). The probability that an object will be attended is determined by its level of activation relative to neighboring competitors. The object with the highest relevant activation level is selectively attended to. Their model assigns the amount of neural activity devoted to an object by combining the bottom-up and the

top-down information. According to Wickens and McCarley (2008), a "computational model guided solely by bottom-up salience calculation can do a reasonable job of simulating human search behavior (Itti and Koch 2000), but models incorporating a top-down component perform better (Navalpakkam and Itti 2005)" (p. 71). Directed top-down processes are tied to a user's ongoing task goals. Therefore, a drawback to models incorporating top-down processing is this: Optimal use requires calibration for each user.

Although both top-down and bottom-up processes combine to influence attention (Wolfe and Horowitz 2004), human–computer interaction literature has mainly focused on top-down processing. For example, users reading content on a web page often display a top-to-bottom and left-to-right pattern of fixation (Rayner 1998). Eye-tracking research has shown that changing the task also changes fixation patterns (Cutrell and Guan 2007). Even in studies intended to manipulate top-down influences, top-down and bottom-up processes interact, which can make it difficult to see their unique contributions (McCarthy et al. 2003). Therefore, research examining bottom-up influences should not be neglected. Based on the Cognitive Psychology visual search literature, saliency ought to carry a heavy influence, early in a visual search. This makes web page interface searches a strong platform for exploring bottom-up processes, since users typically spend only 4–9-s searching within a web page, skimming approximately 18 words (Chen et al. 2001; Nielsen 2008).

Interestingly, interface designers have been able to use bottom-up processes to override users' biases in web pages. For instance, users learn where ad banners are conventionally located and they learn to avoid them. This spatial avoidance bias is commonly referred to as *banner blindness* (Albert 2002). To thwart avoidance, designers introduce transient signals within an ad by flashing or shaking objects (Burke and Hornoff 2001). These signals are characterized as "exogenous," meaning that attention is reflexively drawn. Unlike visual saliency and conventional knowledge, which nudge the spotlight of attention, transient signals demand attention. Using these cues is a double-edge sword; they draw attention, but they also disrupt the ongoing task (Rensink 2002). Unless we are designing alerts, this type of design solution is experienced as unpleasant (Iqbal and Bailey 2008). Users explore interfaces with a wide variety of goals. It is unlikely that one information source, like a product banner ad, will facilitate a user's goal. Attentional control should not be taken away from users; rather, designs can implicitly guide users to potential interface elements of stakeholder interest.

2.2 Historical approach to predicting visual hierarchy

Traditionally, Web designers have sought an understanding of visual perception from the Gestalt perspective (Arnheim 1954). This psychological framework focuses on the whole being greater than the sum of its parts. This wider conceptual gaze led to a set of perceptual organization principles that attempt to describe the emergent experience of the "whole" scene (e.g., figure–ground, similarity, proximity). These Gestalt principles helped designers effectively create sets of Web elements that appear to belong together. In turn, these principles have been described as a key to offering intuitive designs (Flieder and Modritscher 2006). However, designers still need a visual hierarchy that guides viewers between groupings and to important groupings. This can be achieved by employing hierarchy tools like size, color, contrast, alignment, and repetition (Jones 2011). According to Bradley (2015), designers can create entry points by having one dominant element. This dominant element could be a large picture which carries the greatest visual weight. Then, the viewers can move through a series of focal points which are subdominant. These points carry lesser dominance, but still, they hold the user's attention. Focal points might be achieved by using color contrast, for instance, since a difference in local color calls attention. Finally, the other Web elements ought to be subordinate. They should fade into the background (e.g., the body of text). Therefore, this description of a visual hierarchy is represented by decreasing the level of visual dominance (i.e., from most important to less distinction: dominant, subdominant, subordinate). Unfortunately, these visual hierarchy proposals are neither well explored nor empirically justified. Grier et al. (2007), for instance, did not find that large images dominate attention within web pages. Clearly, there is a need for a computationally driven model which provides designers with a means to effectively predict the bottom-up guidance of attention within complex displays.

3 Classic visual saliency and the attentional priority model

A critical factor in whether a search will be fast and effortless is visual saliency. It guides attention toward regions containing locally unique features (Itti and Koch 2000). In the rare case that a display is uniform, salient objects appear to pop-out (e.g., a red apple among green apples). Research has shown that, under these conditions, the time it takes to detect the salient object is not modulated by the number of distractors within the display. The salient object is quickly

and easily found. Clearly, object saliency affects attention. Researchers have shown this to be the case, even when the saliency is associated with a distractor instead of a target. In other words, saliency will pull the spotlight of attention whether it is task-relevant or not. For instance, Theeuwes (1992) presented a salient object (circle) amid uniform distractors (diamonds) and found a pop-out effect. In some of the trials, one of the distractors was presented in a unique color making the irrelevant distractor salient. Interestingly, Theeuwes found that responses to targets were slower in the presence of a salient distractor. This suggests that attention is involuntarily captured by salient distractors despite their irrelevance to the task (i.e., the uniquely colored item was never presented as a target). Further, Pashler (1988) showed that simply *knowing* the location of uniquely colored distractors slowed target detection. Finally, Kim and Cave (1999) confirmed the influence of salient distractors by showing that target detection is faster when the target appears at a location that previously held a salient distractor. Notably, in some trials, participants' responses were just as quick to probes appearing at locations of salient distractors as to target location. As a whole, these studies suggest that saliency can drive attention, regardless of task demands (Theeuwes 2004).

Still and Masciocchi (2010) suggested that Itti et al. (1998) classic saliency model could be used to predict which regions of a web page would draw users' stimulus-driven attention. The model's predictions are void of any useful object meaning or spatial structure within a scene. Instead, the visual properties of a design contribute to the formation of regions with differing amounts of uniqueness, or salience, producing an initial stimulus-driven pre-attentive bias. Saliency models can account for this attentional bias in scenes. These models predict human behavior by assuming that certain low-level pre-attentive visual features implicitly influence overt attention independently of goal-directed processes.

Itti et al. (1998) saliency model produces a visual saliency map that is based only on local image contrast. According to Moraglia (1989), local contrast can impact search performance; for instance, participants were slower to respond to the orientation of a line segment when neighboring segments had a similar orientation. Saliency models are typically based on low-level channels and local processing, following the extraction of information from feature channels focused on image color, light intensity, and orientation. The separate channels are combined to produce a single saliency map. An image's corresponding saliency map provides predictions about where users will fixate. The highest values in the map identify where a viewer ought to fixate first, and then, they are likely to continue fixating on locations with descending values within the map.

3.1 Attentional priority model description

The classic saliency model's predictions can be improved by integrating spatial bias that reflects interaction conventions. The novel process of forming a convention map by extracting commonly fixated regions across a variety of web page types is explained. The convention map was generated using Masciocchi and Still's (2013) fixation database. The fixations were collected while participants viewed a variety of web page types. This web page-specific convention map, as expected, reflected a top-left to center spatial bias congruent with previous web page search findings (Buscher et al. 2009; Grier et al. 2007; Jana and Bhattacharya 2015). Notably, convention bias is a reflection of implicit expectations, not a specific property of the stimulus. Incorporating search biases ought to improve the predictive power of the saliency model within web pages. Others, computationally focused researchers (e.g., Parkhurst et al. 2002), include these fixation biases in their baseline for comparison (i.e., they remove it as a valuable contribution). This research examines the potential benefit of this inclusion and begins to explore the best map weighting. The following formal notation describes the creation of a convention map, and the integration of the convention and saliency maps (see Fig. 1).

The AP model is described in two steps. First, the convention map is created. The convention map was formed by using the Masciocchi and Still (2013) fixation database. It contained eighteen participants with both normal vision and extensive website experience. Participants freely viewed fifty web pages classified as mostly text, half picture and text, and mostly pictures. The first ten fixations ($_{fc}$) were extracted from the Masciocchi and Still (2013) database for each participant. Figure 1 provides the formal notation for calculating a convention map for a single participant's first ten fixations. The convention map employed in this study and shown in Fig. 2 included data from 18 participants. This is achieved by adding all fixations to the map matrix (e.g., 180 fixations) before the convention map is normalized by the max value. The fixations are scaled to reflect the stimuli display resolution of 1024 (r_y) $\times$ 768 (r_x). For consistency and subjective mapping, all the produced maps will reflect this resolution. At every fixation location, a normal three-dimensional Gaussian distribution is applied with a standard deviation of 27 pixels. This standard deviation (σ) represented the approximate eye tracker error. When Gaussian distributions overlapped, the values were summed ($\bar{M}'$). Then, the map was normalized by dividing all of the values by the max value within the map. This produced a convention map ($\bar{C}$) with an output similar to the saliency model (see Fig. 2). The second step is to combine the saliency ($\bar{S}$) and

Fig. 1 Panel A reflects the formal notation of creating a convention map, and panel B reflects the joining of the convention and saliency maps to produce the AP map

A. Create Convention Map

Let f_c be 10 for number of fixations; $f_c \in \mathbb{N}$

Let r_x be 768 and r_y be 1024 for display resolution; $r_x \in \mathbb{N}$, $r_y \in \mathbb{N}$

Let σ = 27

Let -3 σ ≤ l ≤ 3 σ and -3 σ ≤ k ≤ 3 σ

Given fixation matrix, $\overline{F} = (f_{ij}) \in \mathbb{N}^{f_c \times 2}$, where $1 \leq f_{i1} \leq r_x$ and $1 \leq f_{i2} \leq r_y$

$$v = f_{i1} + k$$

$$z = f_{i2} + l$$

Given map matrix, $\overline{M}_i = (M_{vz}) \in \mathbb{R}^{r_x \times r_y}$; $1 \leq v \leq r_x$ and $1 \leq z \leq r_y$

$$(M_{vz}) = e^{\frac{-(1-l)^2+(1-k)^2}{(2\sigma)^2}}$$

$$\overline{M}' = \sum_{i=1}^{f_c} \overline{M}_i$$

Given Convention Map matrix, $\overline{C} = \frac{1}{\max(M')} \cdot \overline{M}$

Given Saliency map matrix, $\overline{S} = \in \mathbb{R}^{r_x \times r_y}$

B. Combine Convention and Saliency Maps

Let weight for Convention map be $w_c = 0.4$ and Saliency map be $w_s = 0.6$

$$\overline{AP} = w_c \overline{C} + w_s \overline{S}$$

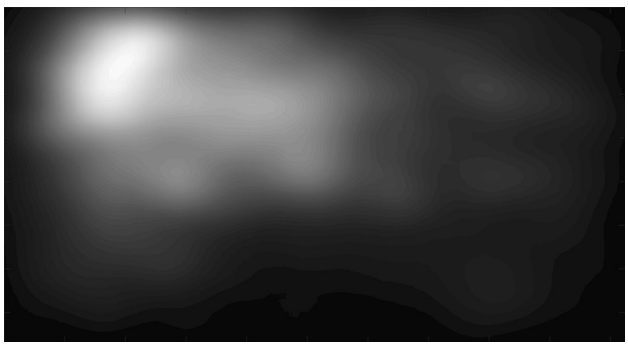


Fig. 2 Subjective visualization of the convention map

convention map to produce the AP map. The convention and saliency map values are first weighted (*w*). Then, the convention map is linearly combined with the saliency map producing the AP map ($\overline{AP}$).

3.2 Saliency, spatial conventions, and web page design

In Faraday's (2000) theoretical framework, designers use a list of guidelines to identify and rank the most salient locations in order. For instance, a moving Web element ought to be ranked first, followed by the largest design element, and followed next by the brightest object. Grier et al. (2007) empirically investigated this approach and only found support for motion attracting attention within web pages. Interestingly, spatial location within the display was found to predict the likelihood of initial fixation better than the features highlighted in Faraday's guidelines. Other researchers have also attempted to describe where users fixate, early in the search sequence. For instance, Buscher et al. (2009) found that viewers need to recognize the page and to perform information-foraging tasks. The initial page

recognition process clearly shows a search bias toward the top-left of the page. Perhaps this region is favored since it normally contains elements that facilitate site recognition (e.g., logos, titles). It appears that the traditional qualitative heuristics for how to guide attention through an interface do not have strong empirical support. This lack of support may stem from the heuristics being based on reflection, but designers cannot use reflection to identify selective pre-attentive processes. Still and Masciocchi (2012) recommend the use of a computational model to quantify the guiding stimulus properties within an interface.

The AP model is an empirical extension of this work focusing on integrating visual saliency with spatial conventions. The saliency map reflects the stimulus properties, while the conventional map reflects general expectations. The AP model ought to outperform the saliency model within both nature scenes and web pages given the consideration of both internal (expectations) and external (salience) factors. The increase in performance within nature scenes will result from the inclusion of central bias, which is well documented in vision science literature (Tatler 2007). Normally, this central bias is viewed as an experimental artifact that results from screen viewing conditions. However, accounting for central bias is viewed as valuable within the context of predicting attentional deployment within screen displayed web page interfaces.

The performance of this AP model's ability to predict fixations within web pages will be contrasted with the original saliency model. This will allow us to replicate a study (Masciocchi and Still 2013) that uses a saliency model to predict attentional deployment within web pages. Further, image type will be explored by showing participants several different stimuli: traditional nature scenes, mostly image web pages, equal parts image and text web pages, and mostly text web pages. The saliency model ought to be able to predict the fixation location within traditional nature scenes and web page images types. The inclusion of traditional nature scenes allows for replication of previous findings (Parkhurst et al. 2002) and provides a baseline comparison (i.e., will reflect central bias, but not web page conventions). In addition, the weighting of the conventional map with saliency will be explored.

4 Method

4.1 Participants

Twenty-one undergraduates (17 females; 20 right-handed; 20 English is Native Language) participated in this study for course research credit. The experiment lasted approximately 15 min. One participant was excluded from further analyses because the eye-tracking system was not able to track their fixations.

4.2 Stimuli and equipment

Images were screenshots of 72 websites and 24 traditional nature scenes. The web page images were composed of the following three types: mostly text, mixed images and text, and mostly images (24 images of each type). Similar to Masciocchi and Still (2013), these web page text-to-image portions were meant to reflect some of the diversity reflected in Web design. The Itti et al. (1998) original saliency model maps were generated by using Harel et al. (2006) MATLAB implementation. Each pixel within a saliency map ranged from 0 (black) to 1 (white). A value close to 1 indicated that the model strongly predicted that specific location to be high in visual salience. Further, a value of 0 indicated a strong prediction of no visual salience. None of the model's parameters were modified from their default settings. Figure 3 displays sample images and their associated saliency model maps.

Images were displayed on a 43 × 24 cm screen and at a viewing distance of 61 cm. The images encompassed the entire screen at a 1024 × 768 pixel resolution and subtended at approximately 38.8 × 22.3° of visual angle.

Fixations were captured and defined by the Tobii Pro X3-120 system. This eye tracker gathers binocular data and performs tracking using both dark and bright pupil data at a sampling rate of 120 Hz. The system's tracking accuracy was ($M = .68°$, SD = .21°) and successfully calibrated for more than 94% of the participants. Further, Tobii Studio (3.4.6) was used to present and control the displays.

4.3 Procedure

Participants were instructed to look at the images "as if they were normally surfing the Web." The experiment began by asking participants to fixate on each of the nine numbers within a 9-point calibration sequence. Then, participants fixated on an additional 9-point sequence, providing a means to record system tracking accuracy. Between each trial, participants were shown a fixation cross at the center of the display. The order of the web page and the nature images were random, and each was shown for 5 s.

5 Results

5.1 Replication of previous saliency model findings

To determine whether the saliency model could predict the deployment of attention above chance, two distributions were formed. The observed distribution was created by simply extracting values at fixation locations from saliency maps corresponding with the viewed stimuli. The shuffled

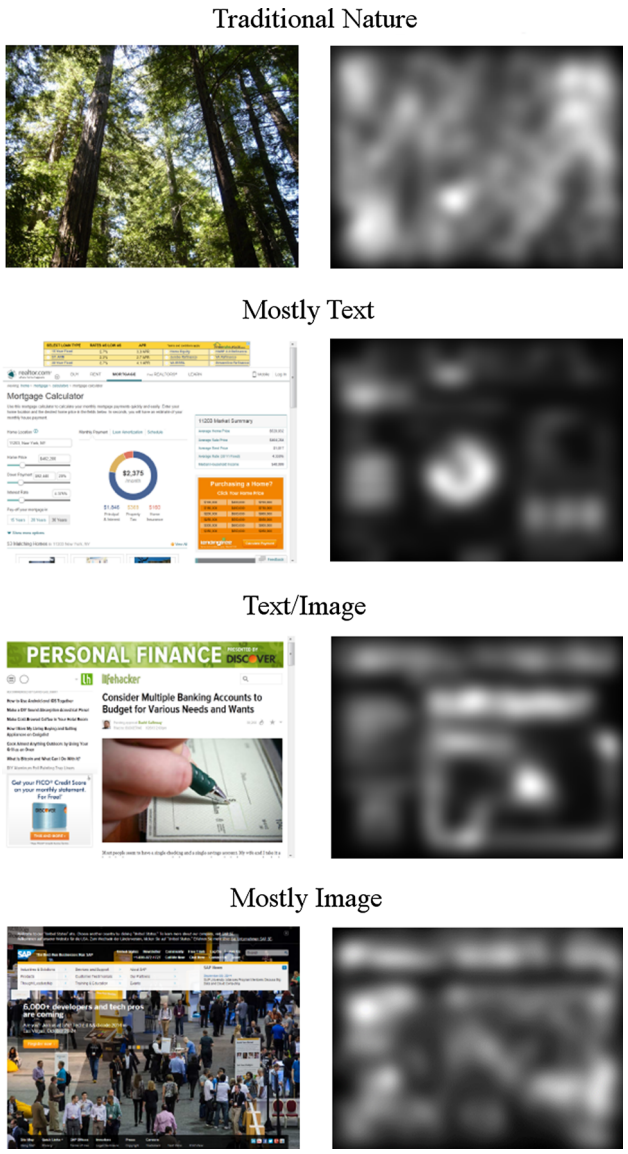


Fig. 3 Prototypical images and their associated Itti et al. (1998) saliency map

distribution was created by extracting saliency map values from fixation locations for each participant that did not correspond to the viewed stimuli. For example, participant one's fixations associated with the first image were used to extract values from all the other images' saliency maps. Then, a shuffle value was computed for each participant by taking the average across their shuffle database. This formed a very conservative chance, which included conventional and central biases.

The means for the observed and shuffled distributions by fixations are shown in Fig. 4. A 2 Distribution (Observed, Shuffled) × 3 Fixations (1 to 4, 5 to 8, 9 to 12) × 4 Image Type (Traditional Nature, Mostly Image, Image/Text, Mostly Text) repeated measures ANOVA was employed to determine whether the difference between the distributions

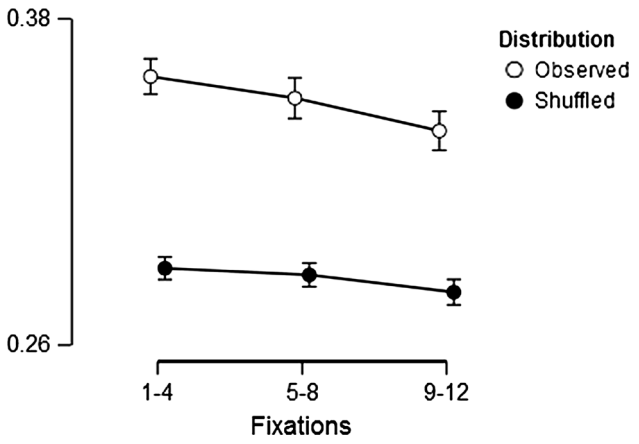


Fig. 4 Mean saliency values by fixations for the observed and shuffled data sets. The *error bars* represent 95% confidence intervals

varied by fixations or image type. Mauchly's test of sphericity indicated that the assumption of sphericity was violated ($p < .05$). Therefore, a Greenhouse–Geisser correction was used. The main effect of Distribution Type was significant, $F(1,19) = 356.83$, $p < .001$, $n_p^2 = .95$, with the observed distribution ($M = .35$, SEM = .004) being higher than the shuffle distribution ($M = .29$, SEM = .001). There was not a significant interaction between Distribution × Image Type $F(2, 35) = 2.9$ $p = .07$, $n_p^2 = .13$, Distribution × Fixations, $F(2, 32) = 3.2$, $p = .062$, $n_p^2 = .14$, or Distribution X Fixations X Image Type $F(5, 87) = .92$, $p = .466$, $n_p^2 = .05$. The difference in distributions supports previous claims (Parkhurst et al. 2002; Still and Masciocchi 2010) that a saliency model could be employed to predict the deployment of attention across a variety of image types.

The main effects and interactions associated with Fixations and Image Type were explored by considering the observed distribution only to prevent shuffle distribution contamination. A 3 Fixations (1 to 4, 5 to 8, 9 to 12) × 4 Image Type (Traditional Nature, Mostly Image, Image/Text, Mostly Text) repeated measures ANOVA was employed to determine whether differences in the observed values varied by fixations or image type. The main effect of Fixations was significant, $F(2, 38) = 10.85$, $p < .001$, $n_p^2 = .36$. The main effect of Image Type was not significant, $F(3, 57) = .83$, $p = .486$, $n_p^2 = .04$. There was a significant interaction between Fixations × Image Type, $F(6, 114) = 4.01$, $p < .001$, $n_p^2 = .17$ (see Fig. 5).

It is apparent the interaction is being driven by a difference between image types reflecting a good portion of text and those image types reflecting mostly images. A 3 Fixations (1 to 4, 5 to 8, 9 to 12) × 2 Image Type (Image/Text, Mostly Text) repeated measures ANOVA revealed no significant differences for Image Type, $F(1, 19) = .77$, $p = .39$, $n_p^2 = .04$. However, it did show significance for

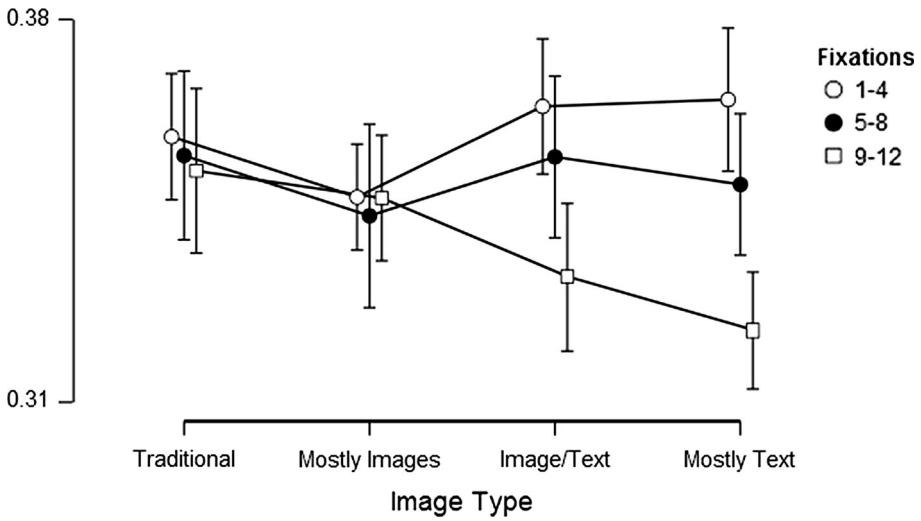


Fig. 5 Mean observed saliency values by fixations and image type. The *error bars* represent 95% confidence intervals

Fixations, $F(2, 38) = 19.94$, $p < .001$, $n_p^2 = .51$. A Bonferroni post hoc test showed that fixations 1 to 4 ($M = .36$, SD = .04, $p < .001$) reflected higher values than 9 to 12 ($M = .35$, SD = .04, $p < .001$), but was not different from 5 to 8 ($M = .33$, SD = .03 $p = .127$). The interaction between Fixations × Image Type was not significant, $F(2, 38) = .57$, $p = .57$, $n_p^2 = .03$. This suggests that the saliency model makes better predictions for earlier than later fixations within web pages containing a good portion of text.

5.2 Determine weighting between saliency map and convention map

The map weightings were achieved by multiplying each saliency map value by the targeted weighting value (.2, .4, .6, .8) and each convention map value by the complementary weighting (.8, .6, .4, .2). Then, the two maps were summed to produce the final AP model map reflecting each of the targeted weightings.

The area under the ROC curve was used as the dependent measure to capture the quality of agreement between a model's map and the associated fixation map. The procedure was implemented in MATLAB using Harel et al. (2006) ROC scripts that calculate values employing a binary classification process. A true positive reflects a pairing of both a fixation map value and a saliency map value above a threshold. The area under the ROC curve represents model performance. A value of 1 means that the model's map reflects the associated fixation map. However, a value of .5 means the model did not represent the associated fixation map.

A 4 Saliency Map Weighting (20, 40, 60, 80) × 4 Image Type (Traditional Nature, Mostly Image, Image/Text, Mostly Text) repeated measures ANOVA explored the predictive performance, as measured through the area under the ROC curve. Mauchly's test of sphericity indicates that the assumption of sphericity was violated ($p < .05$). Therefore, a Greenhouse–Geisser correction was used. The main effect of Saliency Map Weighting was significant, $F(1, 20) = 5.46$, $p = .029$, $n_p^2 = .22$. Bonferroni post hoc tests revealed that the only differences were between the weighting of 80 ($M = .69$, $SE = .006$) compared with 40 ($M = .71$, $SE = .009$) and 60 ($M = .71$, $SE = .008$), $ps < .05$. The main effect of Image Type was significant, $F(2, 37) = 28.21$, $p < .001$, $n_p^2 = .60$. Bonferroni post hoc tests revealed the traditional nature ($M = .68$, $SE = .008$) and mostly images ($M = .68$, $SE = .008$) were not different, $p = 1$. Also, the image/text ($M = .72$, $SE = .01$) and mostly text ($M = .74$, $SE = .01$) were not different, $p = .612$. However, all other comparisons were significantly different, $p < .05$. The interaction between saliency map Weighting × Image Type was significant, $F(2, 36) = 8.44$, $p = .001$, $n_p^2 = .31$ (see Fig. 6). It appears to reflect the decrease in difference between the image/text and mostly text pair and the traditional nature and mostly images pair, across the saliency map weightings. These findings support the selection of either the 40 or 60% saliency map weighting. The 60% saliency weighting was selected as it reflects a bias toward stimulus dependence.

5.3 Comparing the AP and saliency model across a variety of images

A 2 Model Type (Saliency, Saliency + Convention map) × 3 Fixations (1 to 4, 5 to 8, 9 to 12) × 4 Image Type (Traditional Nature, Mostly Image, Image/Text, Mostly Text) repeated measures ANOVA explored predictive performance, as measured through the area under the ROC curve. Mauchly's test of sphericity indicated that the assumption of sphericity was violated ($p < .05$).

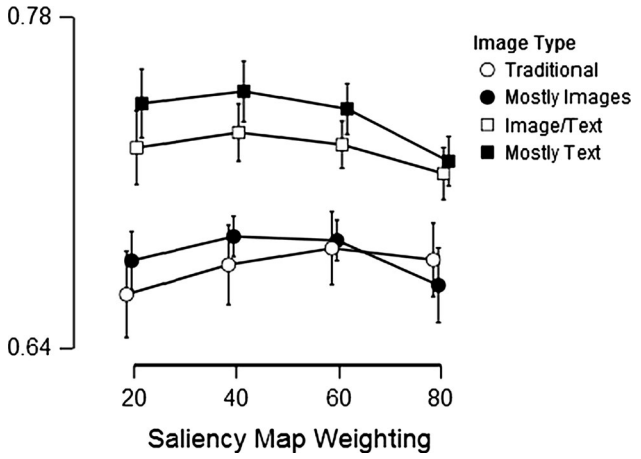


Fig. 6 Mean area under the ROC curve by saliency map weighting for image types. The *error bars* represent 95% confidence intervals

Therefore, a Greenhouse–Geisser correction was used. The main effect Model Type was significant, $F(1, 19) = 55.62$, $p < .001$, $n_p^2 = .75$, with the Saliency + Map model ($M = .71$, $SE = .007$) outperforming Saliency ($M = .67$, $SE = .005$.). The main effect of Fixations was significant, $F(1, 27) = 46.27$, $p < .001$, $n_p^2 = .71$. A Bonferroni post hoc test revealed all comparisons were significant with later fixations reflecting lower values; 1 to 4 ($M = .71$, $SE = .007$), 5 to 8 ($M = .69$, $SE = .005$), 9 to 12 ($M = .67$, $SE = .006$), $ps < .001$. The main effect of image type was significant, $F(3, 49) = 22.16$, $p < .001$, $n_p^2 = .54$. A Bonferroni post hoc test revealed that traditional nature ($M = .67$, $SE = .006$) and mostly images ($M = .66$, $SE = .006$) were not different, nor was image/text ($M = .71$, $SE = .008$) different from mostly text ($M = .71$, $SE = .008$). However, all other comparisons were significantly different, $ps < .001$. The interaction between Model Type × Fixations stems from differences at fixations 1 to 4 and 5 to 8, but not at 9 to 12, $F(2, 31) = 68.62$, $p < .001$, $n_p^2 = .78$. The interaction between Model × Image Type stems from a lack of difference between the traditional nature images by model type, $F(2, 36) = 12.11$, $p < .001$, $n_p^2 = .39$. This shows the lesser impact of the conventional map on the traditional nature image compared with the web page images. The interaction between Fixations × Image Type is significant, $F(5, 88) = 5.31$, $p < .001$, $n_p^2 = .22$, which appears to be from fixations 5 to 8 and 9 to 12 within the traditional nature and mostly images, not benefiting as much as images/text and mostly text. A three-way interaction between Model Type × Fixations, × Image Type was significant, $F(4,$

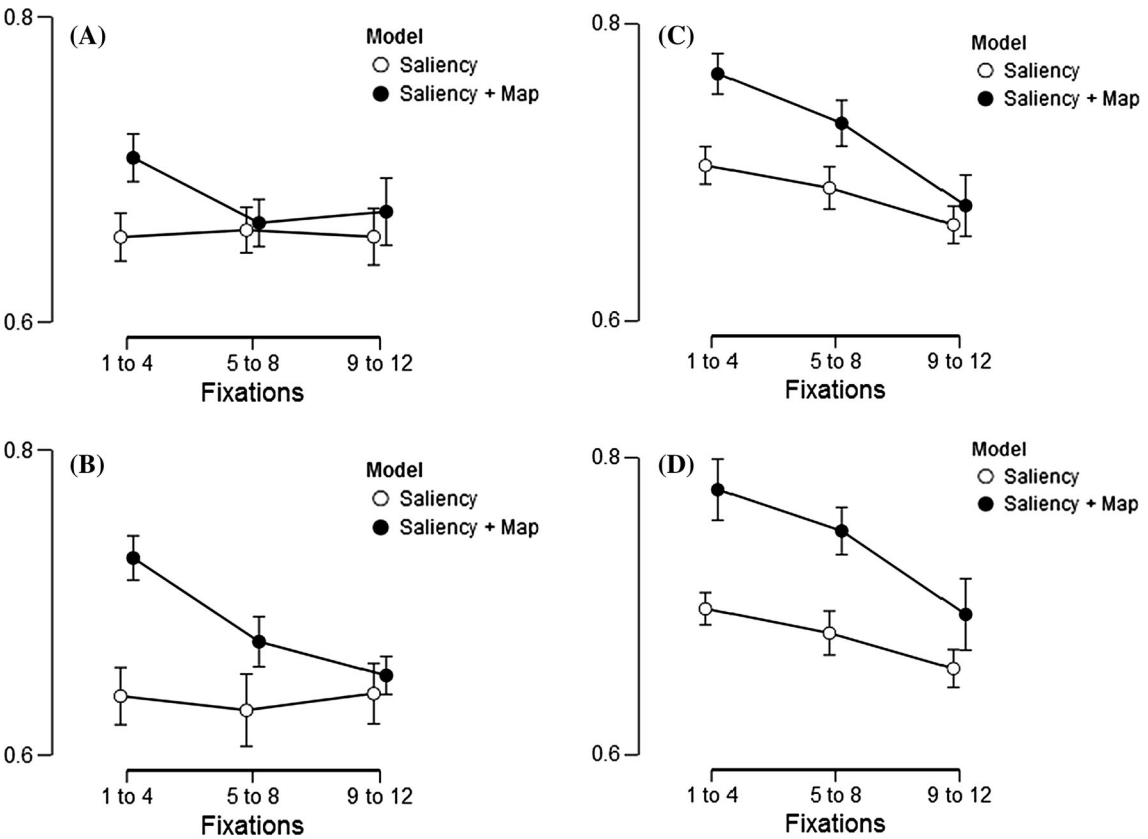


Fig. 7 Mean area under the ROC curve by Model Type × Fixations. *Each panel* represents a unique image type. The *error bars* represent 95% confidence intervals. **a** Traditional nature, **b** mostly images, **c** images/text, **d** mostly text

78) = 7.01, $p < .001$, $n_p^2 = .27$ (see Fig. 7). It appears that the interaction results from the Traditional nature image not differing at 5 to 8 fixations by Model Type. These findings support the use of a web page-specific AP model. Further, they show the best performance early in the visual search and for web pages which contain a good amount of text.

6 Conclusions

As interface designers, we strive to produce visual displays that are easy to search. Within a visually rich and cluttered display, it is not possible to verbally describe the varying degree of bottom-up influences which web page elements reflect. Designers attempt to follow simple heuristics to establish a visual hierarchy (e.g., Faraday 2000), but these rules are too simple to represent the complex processes occurring at an unconscious level (c.f., Treisman and Gelade 1980). Thus, anyone employing this heuristic approach will struggle to account for fixation differences within media-rich displays. However, a more sophisticated account is possible through the employment of a computational model. Previously, Itti et al.'s (1998) saliency model has been shown to predict attentional deployment within web pages (Masciocchi and Still 2013). The AP model is an extension of Itti et al.'s (1998) saliency model. It allows designers the ability to visualize the influence of the saliency and convention biases that can impact the efficiency of interface searches. The saliency map reflects a local contrast that is specific to the display being viewed. And the convention map shows users' implicit expectations, independent of the particular display being viewed. But does integrating a convention map with saliency provide a better account for the deployment of attention within a web page?

Yes, the data suggest that the proposed integration of a conventional map with a saliency model enhances predictive performance within web page stimuli. Further, it appears the AP model best predicts the deployment of attention within web pages which contain equal parts text and images or mostly text. Page layouts that are text-oriented reflect the most conventional influence as their layout structure follows similar design patterns. This work offers a clear extension to the previous saliency models with a focus on web page interfaces. Additionally, these data replicate previous findings that a saliency model can predict the deployment of within both nature scenes (Parkhurst et al. 2002) and web page interfaces (Still and Masciocchi 2010).

6.1 Discussion

User searches are guided to certain interface elements over others. Making these influences apparent to designers will help them create interface elements that are easier to find whether for safety or business reasons. Also, it will allow designers to determine when a non-task critical design element is salient (e.g., banner ad). The goal of this study was to create a conventional map and to combine it with a saliency map and then to perform basic predictive performance testing. Future research will need to explore the effectiveness and usefulness of the AP model from a designer decision making perspective.

Rosenholtz et al. (2011) explored whether using a computational model could actually facilitate design decisions. They found that a model can help, but that more development is needed to truly bridge the gap between research and practice. Others have also attempted to communicate implicit influences to designers without requiring the knowledge of how biologically inspired computational models operate. For example, Jana and Bhattacharya (2015) attempted to provide designers with a useful attention model that mainly employs the saliency toolbox (Walther and Koch 2006). They focused on making their model useful by breaking the web page into objects to provide rank ordering. Unfortunately, this segmentation assumes that objects do not span more than the identified three regions (left, right, or middle). This segmentation limits designers' fine gain ability to understand the nudging of attention within a web page. But clearly, there is a need for further model development with the goal of making implicit design influences apparent.

The AP model should be employed early within the formative develop processes. It can easily be executed quickly on a computer as many times as necessary to achieve design needs. This eases iterative development, because bottom-up attentional priority assessment does not require the time, effort, and cost associated with participant recruitment and testing. Designers could determine whether interface elements which receive higher attentional priority ought to receive it. Then, the users' actual behavior could be verified later in the development cycle, during the summative assessment.

However, future AP model development can focus on going beyond initial fixations and searching with a set goal in mind. Of course, this study allowed the measurement of the spotlight of attention as it moved through a variety of web pages. It is important to understand the influence that a design has on a visual scan, independent of additional factors. However, future research needs to investigate the interaction between explicit goals and the influences captured by the AP model. Based on previous basic visual search research findings (c.f., Theeuwes 1992), participants will be guided by bottom-up influences, regardless of task goals. But examining the interaction can reveal useful and interesting insight both for HCI and cognitive literature.

This research shows that, by integrating spatial conventions with a saliency model, researchers can better predict the deployment of attention. The proposed approach to creating a conventional bias map can be applied and empirically tested within other interface types (e.g., mobile devices or professional systems). Once created, these conventional maps can be shared across the HCI community. Of course, as interfaces change, so will users' expectations. This will require an update to the convention maps. However, the development of culture conventions takes time, so convention maps ought to remain stable for years.

References

Albert W (2002) Do web users actually look at ads? A case study of banner ads and eye tracking technology. In: Proceedings of usability professional association conference

Arnheim R (1954) Art and visual perception: a psychology of the creative eye. University of California Press, Berkeley

Awh E, Belopolsky AV, Theeuwes J (2012) Top-down versus bottom-up attentional control: a failed theoretical dichotomy. Trends Cognit Sci 16:437–443

Baddeley AD (1992) Working memory. Science 255:556–559

Bradley S (2015) Design principles: dominance, focal points and hierarchy. Smashing Mag. Retrieved from https://www.smashingmagazine.com/2015/02/design-principles-dominance-focal-points-hierarchy/

Burke M, Hornoff AJ (2001) The effects of animated banner advertisements on a visual search task. Computer and Information Science Report. University of Nantes, Nantes

Buscher G, Cutrell E, Morris MR (2009) What do you see when you're surfing? Using eye tracking to predict salient regions in web pages. In: Proceedings of the computer-human interaction conference, pp 21–30

Chen M, Anderson JR, Sohn M (2001) What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing. In: Proceedings of CHI: extended abstracts on human factors in computing systems, pp 281–282

Chun MM (2000) Contextual cueing of visual attention. Trends Cognit Sci 4:170–177

Cowan N (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. Psychol Bull 104:163–191

Cowan N (2000) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav Brain Sci 24:154–176

Cutrell E, Guan Z (2007) What are you looking for? An eye-tracking study of information usage in web search. In: Proceedings of CHI conference on human factors in computing systems, pp 407–416

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222

Faraday P (2000) Visually critiquing web pages. In: Proceedings of the 6th conference on human factors and the web, Austin, TX

Fernandez-Duque D, Johnson ML (2002) Cause and effect theories of attention: the role of conceptual metaphors. Rev Gen Psychol 6:153–165

Flieder K, Modritscher F (2006) Foundations of pattern language based on gestalt principles. In: CHI: works-in-process, pp 773–778

Grier R, Kortum P, Miller J (2007) How users view web pages: an exploration of cognitive and perceptual mechanisms. In: Zaphiris P, Kurniawan S (eds) Human computer interaction research in web design and evaluation. Information Science Reference, Hershey, pp 22–41

Harel J, Koch C, Perona P (2006) Graph-based visual saliency. In: Proceedings of neural information processing systems, pp 1–8

Iqbal ST, Bailey BP (2008) Effects of intelligent notification management on users and their tasks. In: Proceedings of the CHI conference on human factors in computing systems, pp 93–102

Itti L, Koch C (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res 40:1489–1506

Itti L, Koch C, Niebur E (1998) A model of saliency-based fast visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20:1254–1259

Jana A, Bhattacharya S (2015) Design and validation of an attention model of web page users. In: Advances in human-computer interaction, pp 1–14

Johnson WA, Dark VJ (1986) Selective attention. Annu Rev Psychol 37:43–75

Jones B (2011) Understanding visual hierarchy in web design. Envato. Retrieved from http://webdesign.tutsplus.com/articles/understanding-visual-hierarchy-in-web-design–webdesign-84

Kim MS, Cave KR (1999) Grouping effects on spatial attention in visual search. Gen Psychol 126:326–352

Malcolm GL, Henderson JM (2010) Combining top-down processes to guide eye movements during real-world scene search. J Vis 10:1–11

Masciocchi CM, Still JD (2013) Alternatives to eye tracking for predicting stimulus-driven attentional selection within interfaces. J Hum-Comput Inter 34:285–301

McCarthy JD, Sasse MA, Riegelsberger J (2003) Can I have the menu please? An eyetracking study of design conventions. In: Proceedings of human-computer interaction, pp 401–414

Moraglia G (1989) Display organization and the detection of horizontal line segments. Percept Psychophys 45:265–272

Navalpakkam V, Itti L (2005) Modeling the influence of task on attention. Vis Res 45:205–231

Nielsen J (2008) How little do users read? http://www.useit.com/altertbox/percent-text-read.html

Norman DA, Shallice T (1986) Attention to action: Willed and automatic control of behavior. In: Davidson RJ, Schwartz GE, Shapiro D (eds) Consciousness and self-regulation: advances in research and theory, vol 4. Plenum Press, New York, pp 1–18

Parkhurst D, Law K, Niebur E (2002) Modeling the role of salience in the allocation of overt visual attention. Vis Res 42:107–123

Pashler H (1988) Cross-dimensional interaction and texture segregation. Percept Psychophys 43:307–318

Rayer K (1998) Eye movements in reading and information processing: 20 years of research. Psychol Bull 124:372–422

Rensink RA (2002) Internal vs. external information in visual perception. In: Proceedings of the 2nd international symposium on smart graphics, pp 63–70

Rosenholtz R, Dorai A, Freeman R (2011) Do predictions of visual perception aid design? ACM Trans Appl Percept 8:1–20

Still JD, Dark VJ (2010) Examining working memory load and congruency effects on affordances and conventions. Int J Hum Comput Stud 68:561–571

Still JD, Dark VJ (2013) Cognitively describing and designing affordances. J Des Stud 13:285–301

Still JD, Masciocchi CM (2010) A saliency model predicts fixations in web interfaces. In: Proceedings of the 5th international workshop on model-driven development of advanced user interactions, pp 25–18. Atlanta, GA

Still JD, Masciocchi CM (2012) Considering the influence of visual saliency during interface searches. In: Alkhalifa EM, Gaid K (eds) Cognitively informed intelligent interfaces: system design and development. Information Science Reference, Hershey, pp 84–97

Still JD, Still ML, Grgic J (2015) Designing intuitive interactions: exploring performance and reflection measures. Interact Comput 27:271–286

Tatler BW (2007) The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor bases and image feature distributions. J Vis 14:1–17

Theeuwes J (1992) Perceptual selectivity for color and form. Percept Psychophys 51:599–606

Theeuwes J (2004) Top-down search strategies cannot override attentional capture. Psychon Bull Rev 11:65–70

Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cogn Psychol 12:97–136

Walther D, Koch C (2006) Modeling attention to salient proto-objects. Neural Netw 19:1395–1407

Wickens CD, McCarley JS (2008) Applied attention theory. CRC Press, Boca Raton

Wolfe JM (2007) Guided search 4.0: current progress with a model of visual search. In: Gray W (ed) Integrated models of cognitive systems, Oxford, New York, pp 99–119

Wolfe JM, Horowitz TS (2004) What attributes guide the deployment of visual attention and how do they do it? Nat Rev Neurosci 5:1–7. doi:10.1038/nrn1411