Examining the Effects of Clutter and Target Salience in an E-Commerce Visual Search Task

John M. Hicks and Jeremiah D. Still
Old Dominion University

For web page designers it is important to consider how the visual components of a page affect how easy it is to use. Visual salience and clutter are two bottom-up factors of stimuli that have been shown to affect attentional guidance. Visual salience is a measure of how much a given item or region in the visual field stands out relative to its surroundings, and clutter is a measure of how much visual information is present and how well it is organized. In this study, we examined the effects of visual salience and clutter in a visual search task in e-commerce pages. Clutter was manipulated by adding grids of varying densities to the background of stimuli. On each trial, participants searched for an item that was either the most or least salient of the items on the page as determined by a computational model of visual salience (Itti, Koch, & Niebur, 1998). The results showed that the high salient targets were found faster than the low salient targets and search times also increased as clutter increased, but these two factors did not interact. We conclude that designers should consider both factors when possible.

INTRODUCTION

E-commerce is becoming an ever more prevalent way in which individuals choose to do their shopping. In 2018, e-commerce accounted for 14.3% of all retail sales, up from just 6.4% in 2010, and is increasing at a faster rate (15%) than brick and mortar retail sales (5%) (U.S. Department of Commerce, 2018). As e-commerce shopping continues to grow, competition will increase. One way e-commerce site designers could help their site stand out compared to competitors is by making it easier to use by helping customers find products quickly and effectively. These components are important because users are more likely to return to webpages they find easy to use (Menon & Kahn, 2002). The visual clutter on a webpage is one factor that determines how easy it is to use and how quickly users can find what they are looking for. Additionally, retailers and product designers may want to guide customers to specific products over others. One way designers could guide users' visual searches effectively is by using visual salience.

Visual salience is how much a given object or region in a visual display stands out relative to its surroundings. Attention is guided automatically toward regions and objects associated with higher salience (Masciocchi & Still, 2013; Still & Masciocchi, 2012). Salient items are found faster than items with lower salience (Still & Still, 2019), and the presence of salient distractors can increase how long it takes to find a target (Theeuwes, 1992, 2004). Clutter also has an effect on search efficiency with more cluttered displays leading to longer search times (e.g., Beck, Lohrenz, & Trafton, 2010; Neider & Zelinsky, 2011). But are clutter and visual salience separate influences? The current study examined how search efficiency is affected by clutter and salience in e-commerce pages.

Selective Attention and Visual Salience

Our visual fields contain too much information to process efficiently at once so we rely on a mechanism known as selective attention in order to guide attention to information that may be important (Johnston & Dark, 1986). Selective attention is guided by the features of items in the visual field (stimulus-driven attention), by an individual's goals and expectations (goal-directed attention), and previous exposure to specific stimuli (selection history) (Awh, Belopolsky, & Theeuwes, 2012). For this study, we focused on stimulus-driven factors, specifically visual salience.

Salience is primarily determined by basic feature channels such as color, brightness, and orientation (Itti & Koch, 2001). For example, a green apple would be visually salient among red apples but not among other green apples. When an item in the visual field can be defined by a single feature it pops-out of the display and is noticed immediately regardless of how many other items are in the display (Treisman & Gelade, 1980). This finding suggests that certain aspects of objects, like basic features, can be processed preattentively. Theories of visual search posit that images are first processed preattentively in order to create a salience map. Attention is then automatically guided to the most salient point and then to subsequent points in order of most to least salient until the search is terminated (Wolfe, 2007). When searching for a target defined by a salient feature, the presence of an additional salient distractor increases search times, even though it is irrelevant to the current task (Theeuwes, 1992, 2004).

However, for real-world displays, items and targets are rarely defined by just one feature. Additionally, it can be difficult, if not impossible, to count the exact number of items in real-world images as there is no precise definition as to what does or does not constitute an object. Further, it can be difficult to determine which areas of a real-world image are visually salient just by looking at it. Thus, for real-world images and displays researchers have often relied on computational models that calculate the visual salience of an image based on its visual properties. The model developed by Itti, Koch, and Niebur (1998) has been widely used. This model has been shown to predict attentional guidance in a variety of types of images such as nature scenes, fractals, urban scenes, (Parkhurst, Law, & Niebur, 2002), mobile interfaces (Still, Hicks, Cain, & Billman, 2017), and web pages (Hicks, Cain, & Still, 2017; Still & Masciocchi, 2010) among others.

In e-commerce pages, visually salient products are found faster and more efficiently than those of lesser salience (Still & Still, 2019). In Still and Still (2019), individuals searched e-commerce pages for a target item and reported that item's price. The target was either the most salient item in the display or the least salient. In experiment 1, individuals were given a verbal description of the target prior to initiating the search phase. Their results showed that the high salient targets were fixated faster than low salient targets. However, these results may have been due to the relatively weak top-down information provided by a verbal cue. Past studies have shown that providing the exact visual target prior to search makes search more efficient by providing a much stronger top-down target representation (Malcolm & Henderson, 2009; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004). In experiment 2, Still and Still showed the exact visual target prior to each search instead of giving a verbal description of the target. Even with the stronger top-down representation in this experiment they found faster search times were associated with high salient targets. These results suggest that visual salience has an effect on search times even in directed search tasks with strong top-down target representations. For designers, it seems like visual salience is an essential factor that should be considered, but visual clutter may also be important.

Visual Clutter

The number of items and their organization in a display has a profound effect on visual search efficiency. In basic visual search displays, when the target does not immediately pop-out based on its visual features, the number of non-target items in the display affects how long it takes to find the target. As noted previously, however, it can be difficult to determine what exactly constitutes an object. Visual clutter has been proposed as an analogous measure of set size for complex and real-world displays. It is defined as the amount of visual information and its organization that leads to a detriment in search efficiency (Rosenholtz, Li, & Nakano, 2007). Imagine an office workers desk. The more documents, pens, and other items that are on the desk the more difficult it may be to find something specific. However, if all of those things are well organized the search may be relatively easy even though the same number of actual objects are present.

Clutter has been shown to affect search efficiency in a variety of types of displays. An increase in clutter, as defined by the density of a virtual city, is associated with decreased search efficiency (Neider & Zelinsky, 2011). On maps, targets become challenging to find as the overall level of clutter increased, especially when the target was in a location of high local clutter (Beck et al., 2010). However, the relationship between clutter and search efficiency may not be as simple as these findings might suggest. Similarly, Neider and Zelinsky (2008) manipulated the number of trees in a display to examine how it would affect visual search. Like other studies, they found that increasing the number of trees increased search times, but once a certain number of trees was reached the pattern reversed. At that point, adding more trees actually decreased search times. The authors attributed this to perceptual grouping. As more trees were added they became less sparse and were spaced closer together. This then allowed for trees to be visually grouped together. With few objects present, the trees were viewed as foreground objects and the rest of the display as the background, but once the trees encompassed a majority of the display, it seemed as if those roles were reversed. This pattern of results suggests that examining or manipulating clutter may not be as simple as adding more items to a display.

For web pages, clutter is an important factor in design because it impacts subjective ratings of aesthetics (Lavie & Tractinsky, 2004). An essential component of aesthetics is the ratio of the organization of items in a display to the total number present, which apparently shares a good degree of overlap with the concept of clutter. This type of simplicity and minimal clutter design is associated with increased accessibility (Hoehl & Lewis, 2011), increased revisits to web pages (Rosen & Purinton, 2004), and an increase in purchases online (Karvonen, 2000). Clutter is an important component to web page design, but is it different from the influence of visual salience? Previous work has examined the role of visual salience in web pages and found that computational models could be an effective tool for designers to use to predict where users will look. The current study directly manipulated clutter of e-commerce pages to examine how it would affect search efficiency for targets near regions with higher or lower salience.

METHOD

Participants

Forty undergraduate students (30 female, M = 19.30 years, SD = 2.02; 38 native English speakers; 38 right-handed) from a large southeastern university in the United States participated and were compensated with partial course research credit.

Stimuli

Ninety-six e-commerce pages were selected for testing. The stimuli were screenshots taken from a variety of popular e-commerce sites (e.g., Amazon, Target, Walmart) at a resolution of 1024 x 768. E-commerce pages were selected for testing based on the following criteria: the page must contain six item displayed in a grid format of two rows by three columns, each item must be visually unique, each item must not appear on more than one page, among the six items on a page one must be of clear highest visual salience and one must be of clear lowest salience, and finally the high and low salient items must have different prices. Salience was determined using the Itti et al. (1998) computational model (Figure 1). Images were initially selected qualitatively by examining each page's salience map. Then we examined the salience of these stimuli and confirmed quantitatively by sampling the salience

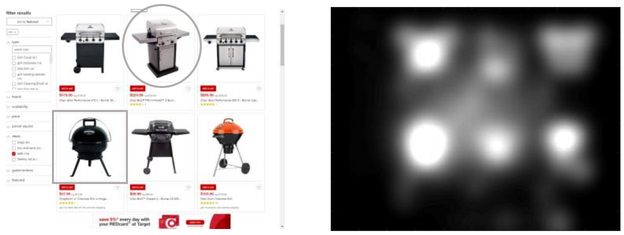

Figure 1. A sample stimulus with its corresponding salience map. The square indicates the high salient item and the circle the low salient item. These shapes are for illustrative purposes only and were not present during testing.

map values of the item locations. To do this, we normalized the salience values to a scale of 0-100 then sampled the salience values of each pixel within 100 pixels from the center of mass of the target and averaged them. The average difference between the high and low salient targets was about 40 points across all levels of clutter (Table 1).

Visual clutter was manipulated by placing grids of varying sizes in the background of each e-commerce page (Figure 2). We manipulated clutter in a way which did not change the general semantics or gist of an image while also keeping the effective set size constant. For the medium clutter condition, dark gray lines 1 pixel thick were placed every 60 pixels both vertically and horizontally. For the high clutter condition, the lines were placed every 20 pixels. This was repeated for all 96 stimuli. The low clutter condition presented the e-commerce pages as is without any manipulation. To check whether this manipulation affected clutter we tested it with two different computational models developed by Rosenholz et al. (2007): feature congestion and subband entropy. Separate one-way repeated measures ANOVAs were run for each model with clutter level as the independent variable (low, medium, high) and the model value as the dependent measure (see table 2 for clutter values). These results showed that our manipulation was effective: feature congestion, $F(2, 190) = 8554$, $p < .001$, $\eta^2_p = .989$, with the high clutter condition showing significantly higher scores than the medium $t(190) = 94.61$, $p < .001$, and low conditions, $t(190) = 93.09$, $p < .001$, and the medium condition showing higher scores than the low condition, $t(190) = 87.87$, $p < .001$. The pattern of results was the same for subband entropy. The clutter values for our stimuli fall slightly on the higher end of the distribution for web page clutter, which previous research has shown to have a mean of 5.70 and SD of 1.70 for feature congestion (Lafleur & Rummel, 2011).

Table 1
Average Target Salience Values (SD in parentheses)

Target Salience	Clutter Low	 Medium	 High
High	56.43 (11.91)	56.23 (11.46)	55.40 (11.31)
Low	14.78 (8.01)	15.17 (7.84)	14.96 (7.41)

Note: Salience was calculated using Harrell, Koch, & Perona's (2006) implementation of the Itti et al. (1998) model.

Figure 2. Example of the clutter manipulation for the low clutter (top), medium clutter (bottom left), and high clutter (bottom right) conditions.

Equipment

Participants' eye movements were recorded using a Tobii X3-120 eye-tracker, which sampled eye movements at a rate of 120 fixations per second. The experiment was programmed and run using the Tobii Studio software. The average tracking error across participants was 0.65° with a standard deviation of 0.19°. The monitor on which the stimuli were presented was 43.18 cm wide by 24.13 cm in height set to a resolution of 1024 x 768. The monitor occupied 38.83° of visual angle by 22.26°. Participants' head movements were not restricted, and they were seated so their eyes were approximately 60 cm from the screen.

Procedure

Participants provided informed written consent prior to participating. The experiment began with two 9-point calibration sequences to configure the eye tracking system and record tracking performance. Our experimental procedure was adopted from Still and Still (2019). Each trial began with a central fixation cross for one second. Then the target item replaced the central fixation cross for two seconds. The target item was either the most or least salient item from its respective e-commerce page. Then a perceptual mask replaced the target for one second. The mask consisted of all possible targets layered on top of each other at 1% opacity. After the mask, another fixation cross appeared for two seconds. This fixation was located directly between the most and least

Table 2
Average Clutter Values (SD in parentheses)

Clutter Model	Clutter Low	 Medium	 High
Feature Congestion	4.98 (0.63)	7.24 (0.54)	9.68 (0.55)
Subband Entropy	2.98 (0.28)	3.74 (0.15)	4.10 (0.13)

salient items from the subsequent e-commerce page. Then the full e-commerce page would appear and remain on screen until a response was made. Participants' task was to locate the target and verbally report its price "as quickly and as accurately as possible." The experimenter recorded the verbal responses and pressed a key to proceed to the next trial once a response was given. Participants completed a total of 96 trials. On half of the trials they searched for the high salience item, and on the other half, they searched for the low salience item. Of the 48 high and low target salience trials, there were 16 each of low, medium, and high clutter. No item or e-commerce page was shown to a single participant more than once and the target salience and clutter levels were counterbalanced such that each combination would appear an equal number of times across participants. The order of item presentation was random for each participant. The entire session lasted approximately 30 minutes.

RESULTS

Primary Analysis

Two participants (5%) were excluded from analysis because of eye tracking errors of greater than 1° of visual angle. The primary dependent measure used was the total search time. This was defined in accordance with Malcolm and Henderson's (2009) description of visual searches. Searches begin with an initiation phase, defined as the duration of the first fixation, followed by the search phase defined as the time from the second fixation until fixating in the area of interest, and finally the target verification stage, defined as the duration of the first fixation in the area of interest. Thus, to calculate how long it took to find the target, we summed the fixation durations beginning with the second fixation up to and including the first fixation in the area of interest.

Data were analyzed using a 2 (target salience: low, high) x 3 (clutter: low, medium, high) repeated measures ANOVA with reaction time as the dependent measure (Figure 3). The analysis revealed a significant main effect of target salience, $F(1, 37) = 14.18$, $p < .001$, $\eta^2_p = .277$, with high salient targets ($M = 824.29$ ms, $SD = 232.87$) being found faster than low salient targets ($M = 904.14$ ms, $SD = 287.08$). The main effect of clutter was also significant, $F(2, 74) = 5.83$, $p = .004$, $\eta^2_p = .136$. The two-way interaction between target salience and clutter was not significant, $F(2, 74) = 0.094$, $p = .911$, $\eta^2_p = .003$. To further examine the main effect of clutter we conducted pairwise comparisons using Bonferroni corrections. These results showed significant differences between the high and low clutter conditions, $t(74) = 3.35$, $p = .004$, but no significant differences between the high clutter and medium clutter conditions, $t(74) = 2.38$, $p = .060$ or between the medium and low clutter conditions, $t(74) = 1.12$, $p = .802$. Further, a trend analysis showed a significant linear trend for clutter, $t(37) = 3.36$, $p = .001$, with RTs increasing as clutter increased.

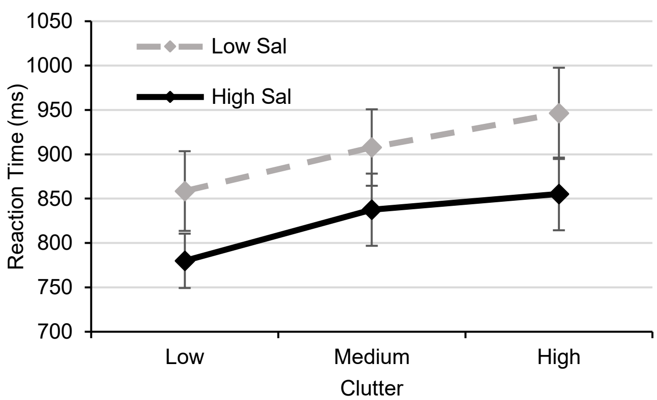

Figure 3. Mean RTs for the primary analysis. Error bars represent standard error of the mean.

Item Analysis

For designers, it is important to know how probable it is that a salience or clutter computational model is actually predictive across a variety of interface designs. To do this, instead of averaging across images and examining the effect at the participant level we averaged across participants and examined the effects at the image level, and also the proportion of images on which the expected pattern of results was found. Two images were not included in the analyses for having missing data on too many trials. We ran a 2 (target salience: low, high) x 3 (clutter: low, medium, high) repeated measures ANOVA with reaction time as the dependent measure. The analysis revealed the same pattern of results as the primary analysis with a significant main effect of target salience, $F(1, 93) = 7.03$, $p = .009$, $\eta^2_p = .070$, a significant effect of clutter, $F(2, 186) = 5.36$, $p = .005$, $\eta^2_p = .054$, and no significant two-way interaction, $F(2, 186) = 0.12$, $p = .884$, $\eta^2_p = .001$. Again, there was a significant linear trend for clutter, $t(93) = 3.21$, $p = .002$, with RTs increasing as the clutter level increased. For salience, 63 out of 94 (65%) stimuli showed faster RTs for high salient targets compared to low salient targets. For clutter, the high clutter condition produced longer RTs than the low clutter condition on 52 out of 94 stimuli (55%).

DISCUSSION

This study examined the effects of clutter and salience on visual searches in e-commerce displays. Consistent with previous literature, increasing clutter produced longer search times (e.g., Beck et al., 2010; Neider & Zelinsky, 2011). Additionally, products near higher salience were found faster than products near lower salience replicating the findings of Still and Still (2019). Interestingly, no interaction was revealed. It appears clutter and salience are independent influences. On each trial, participants were shown the target precisely as it would appear during the search task, which should create a very strong target representation. This is consistent with Still and Still's (2019) findings using a similar experimental paradigm, but it conflicts with findings from research on basic displays showing that a strong top-down target representation can override the effects of attentional capture by salient stimuli (e.g., Dowd & Mitroff, 2013; Folk, Remington & Johnston, 1992). For complex real-world displays, it may be that individuals rely more on subtle differences in visual salience compared to simple laboratory displays that tend to use stimuli that tend to produce a strong pop-out effect.

Our findings regarding the effects of clutter and target salience are consistent with previous research. However, in the design literature, clutter has been emphasized as an important factor much more often than visual salience. In terms of search efficiency, the current study's findings suggest that both are important components in visual search in e-commerce pages. The item analysis revealed that these effects are generally consistent across stimuli but they are not 100% consistent. However, the finding that salience is effective in guided search is consistent with past findings examining visual salience in web pages (e.g., Hicks et al., 2017; Still, 2017) and e-commerce pages (Still & Still, 2019). More research is needed to determine how robust the effect of clutter is at determining search efficiency in web pages. Our findings also provide support for the use of computational models in determining the salience and clutter of web pages. Using a computational model of visual salience or clutter is inexpensive, simple to implement, and can be done during initial design stages.

CONCLUSION

In this study, we directly manipulated the clutter of e-commerce pages to examine how quickly targets of either high or low salience were fixated. We found that as clutter increased so did reaction times. For salience, we found that targets of high salience were fixated faster than targets of low salience. However, these effects did not interact. The effect of target salience was similar at all three levels of clutter that we tested. We recommend that web designers should try to minimize clutter as much as possible while also considering the relative visual salience of web elements.

REFERENCES

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*, 437-443.

Beck, M. R., Lohrenz, M. C., & Trafton, J. G. (2010). Measuring search efficiency in complex visual search tasks: Global and local clutter. *Journal of Experimental Psychology: Applied, 16,* 238-250.

Dowd, E. W., & Mitroff, S. R. (2013). Attentional guidance by working memory overrides salience cues in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 1786-1796.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology Human Perception and Performance, 18,* 1030-1044.

Hicks, J. M., Cain, A. A., & Still, J. D. (2017, September). Visual saliency predicts fixations in low clutter web pages. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 61, No. 1, pp. 1114-1118). Sage CA: Los Angeles, CA: SAGE Publications.

Hoehl, J., & Lewis, C. (2011, October). Mobile web on the desktop: simpler web browsing. In *Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 263-264). ACM.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 194-204.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene-analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254-1259.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology*, *37*, 43-75.

Karvonen, K. (2000, November). The beauty of simplicity. In *Proceedings on the 2000 Conference on Universal Usability* (pp. 85-90). ACM.

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies, 60,* 269-298.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*, 1-13.

Masciocchi, C. M., & Still, J. D. (2013). Alternatives to eye tracking for predicting stimulus-driven attentional selection within interfaces. *Human-Computer Interaction, 28,* 418-441.

Menon, S., & Kahn, B. (2002). Cross-category effects of induced arousal and pleasure on the internet shopping experience. *Journal of Retailing*, *78*, 31-40.

Neider, M. B., & Zelinsky, G. J. (2008). Exploring set size effects in scenes: Identifying the objects of search. *Visual Cognition, 16,* 1-10.

Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision, 11,* 1-16.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*, 107-123.

Rosen, D. E., & Purinton, E. (2004). Website design: Viewing the web as a cognitive landscape. *Journal of Business Research, 57*, 787-794.

Rosenholz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7,* 1-22.

Still, J. D. (2017). Web page attentional priority model. *Cognition, Technology & Work*, *19*, 363-374.

Still, J. D., Hicks, J., Cain, A., & Billman, D. (2017, July). Predicting stimulus-driven attentional selection within mobile interfaces. In *International Conference on Applied Human Factors and Ergonomics* (pp. 255-261). Springer, Cham.

Still, J. D., & Masciocchi, C. M. (2010, April). A saliency model predicts fixations in web interfaces. In *5th International Workshop on Model Driven Development of Advanced User Interfaces (MDDAUI 2010)* (p. 25).

Still, J. D. & Masciocchi, C. M. (2012). Considering the influence of visual saliency during interface searches. In Alkhalifa, E. M. & Gaid, K. (Eds.), *Cognitively Informed Intelligent Interfaces: Systems Design and Development* (pp. 84-97). Hershey, PA: Information Science Reference.

Still, J. D. & Still, M. L. (2019). Influence of visual salience on webpage product searches. *Journal of ACM Transactions on Applied Perception.*

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51*, 599-606.

Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. *Psychonomic Bulletin & Review*, *11*, 65-70.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97-136.

U.S. Department of Commerce (2018). *Quarterly retail e-commerce sales*. Retrieved from https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf

Wolfe, J. M. (2007). Guided Search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 99–119). New York, NY: Oxford.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research*, *44*, 1411-1426.