Directing User Attention via Visual Flow on Web Designs
Supplementary Material

Xufang Pang*   Ying Cao*   Rynson W.H. Lau   Antoni B. Chan

Department of Computer Science, City University of Kong Kong

S1   Statistics of the web design dataset

Our dataset includes 254 webpages from 6 different popular categories: email (40 pages), file sharing (36 pages), job searching (44 pages), product promotion (54 pages), shopping (40 pages) and social networking (40 pages). They were all crawled from the web. The screenshots of the webpages were captured in full-screen mode from a Chrome browser on a 17-inch monitor with a screen resolution of 1280×1024. To make our dataset as representative in design style as possible, for each category, we selected the webpages with varying design complexities in terms of the number, size and spatial layout of page components. In addition, we selected less popular webpages that the general public are not familiar with, so as to reduce personal bias that may arise in our eye tracking experiment (Section S2) due to the participants' familiarities with popular webpages. We show the statistics of our dataset in terms of the number and size of major page components (text, input field, button, and image) in Figures 1, 2 and 3. From these plots, we can see that the webpages in our dataset have various design styles, from designs with heavy text, to designs with one or several large images and little text, to designs with a balanced mix of text and images.

Figure 4 plots the spatial distributions of various page components and white space for webpages in different categories. The components tend to be uniformly distributed over the entire page, especially for the text, button, and image components. This demonstrates the diversity of our dataset in terms of spatial layout of page components. These figures show that the webpages in our dataset, while not exhaustive, span a wide range of design variations in terms of the number, size and spatial layout of page components. In Figure 5, we show some sample pages from each category of our dataset.
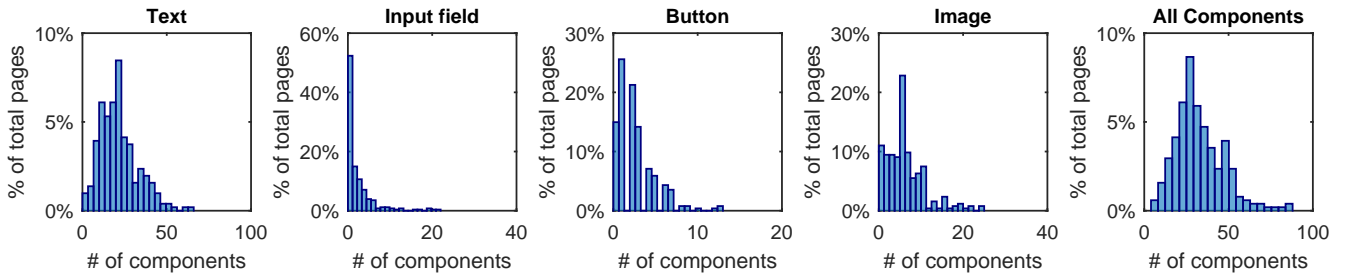
---

*Joint first authors

Figure 1: Distribution of the number of page components of each particular type on a single page. The vertical axis indicates the percentage of pages in the dataset.

S2   Eye tracking experiment

We performed an eye tracking experiment to collect users' eye gaze data on the webpages in our dataset under both task-driven and free-viewing conditions, using a Tobii T120 eye tracker. Our experiment involved 80 participants, who were recruited from a university's participant pool, with ages ranging from 18 to 40. All of the participants have normal and corrected-to-normal vision, and are experienced web users.

In each viewing session under the task-driven condition, each participant was asked to complete a sequence of tasks, each containing a task instruction, followed by one or two webpages to view. Our tasks fall into two categories: comparison tasks and shopping tasks. In a comparison task (as shown in Figure 6(a)), the participants were first presented a goal, which was relevant to the webpage category, to be achieved by comparing two webpages. Then, the participants viewed the two webpages one by one, and were asked to perform a selection action at the end of the task. In a shopping task (as shown in Figure 6(b)), the participants were first shown the goal of buying a favorite product item within a given cash budget. Participants then viewed a single shopping webpage to select the product item that they wanted to buy. The two tasks represent two common scenarios for browsing webpages: comparing services from different webpages, and comparing products on a single webpage. While more specific tasks depend on the webpage type, we believe our comparison and shopping tasks are representative enough to put participants into task-driven context in our study.

Each participant completed 14 comparison tasks and 5 shopping tasks in a viewing session, which took about 3-6 minutes. The presentation order was randomized for each participant. In particular, we randomized the order of comparison tasks, the order of shopping tasks, and the order of webpages shown in each comparison task. For each webpage, we collected the eye gaze data from 10 different participants under the task-driven condition.

In each viewing session under the free-viewing condition, each participant was shown a sequence of about 30 webpages randomly selected from our dataset, and was asked to view the pages casually without any
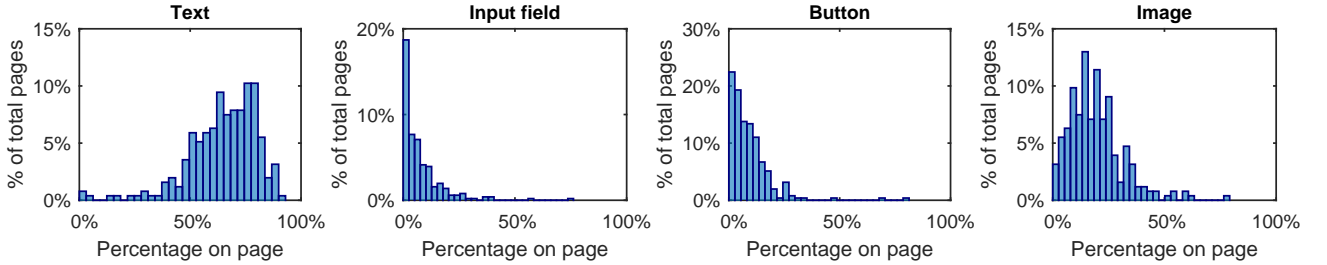

Figure 2: Distribution of the percentage of page components of each particular type on a single page. The vertical axis indicates the percentage of pages in the dataset.
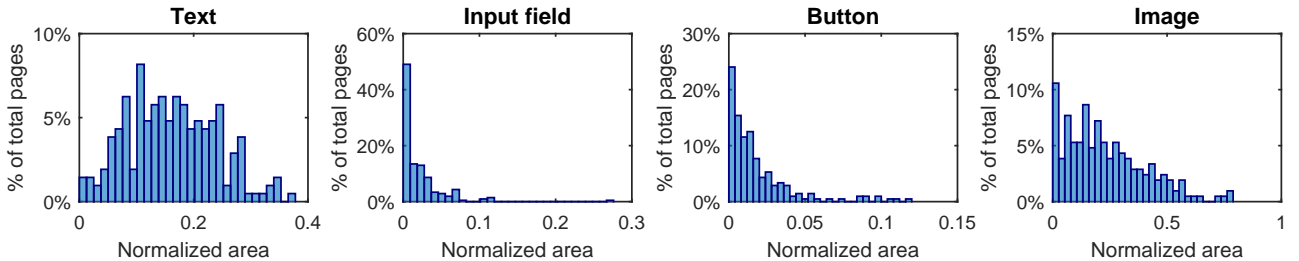

Figure 3: Distribution of the normalized area of page components of each particular type on a single page. The area is normalized with respect to the entire page area. The vertical axis indicates the percentage of pages in the dataset.

2

constraint, which took about 5 minutes on average. The presentation order of webpages was randomized for each participant. We collected the eye gaze data from 10 different participants under the free-viewing condition. The same participants could attend multiple viewing sessions under both task-driven and free-viewing conditions, but did not view the same webpage more than once. For both conditions, we did not limit the time spent by the participants on viewing each page, allowing them to manually advance their own progress via mouse-clicks.

The screenshots of webpages were displayed in their original resolutions on a 17-inch monitor with a screen resolution of 1280 × 1024. The participants were asked to sit in front of the monitor with their heads fixed on a chin-rest at a distance of 65cm from the screen. At the beginning of each viewing session, we used the 9-points grid calibration method to calibrate the eye tracker for each participant. After each task, we corrected the drift by asking the participants to fixate at the center of the screen. Raw eye gaze points were captured at a sampling rate of 120Hz and were grouped into fixations using Tobii's default I-VT fixation filter, which allocates two raw gaze points to the same fixation if they possess a speed below a certain threshold.

S3 Data Analysis for Context Discrepancy

To verify if there exists discrepancies in user behaviors between different contexts, we perform a correlation analysis to investigate how well users' eye gaze data is correlated between task-driven and free-viewing conditions. We describe users' eye gaze data using three representations: saliency map $\mathcal{M}^1$, local attention
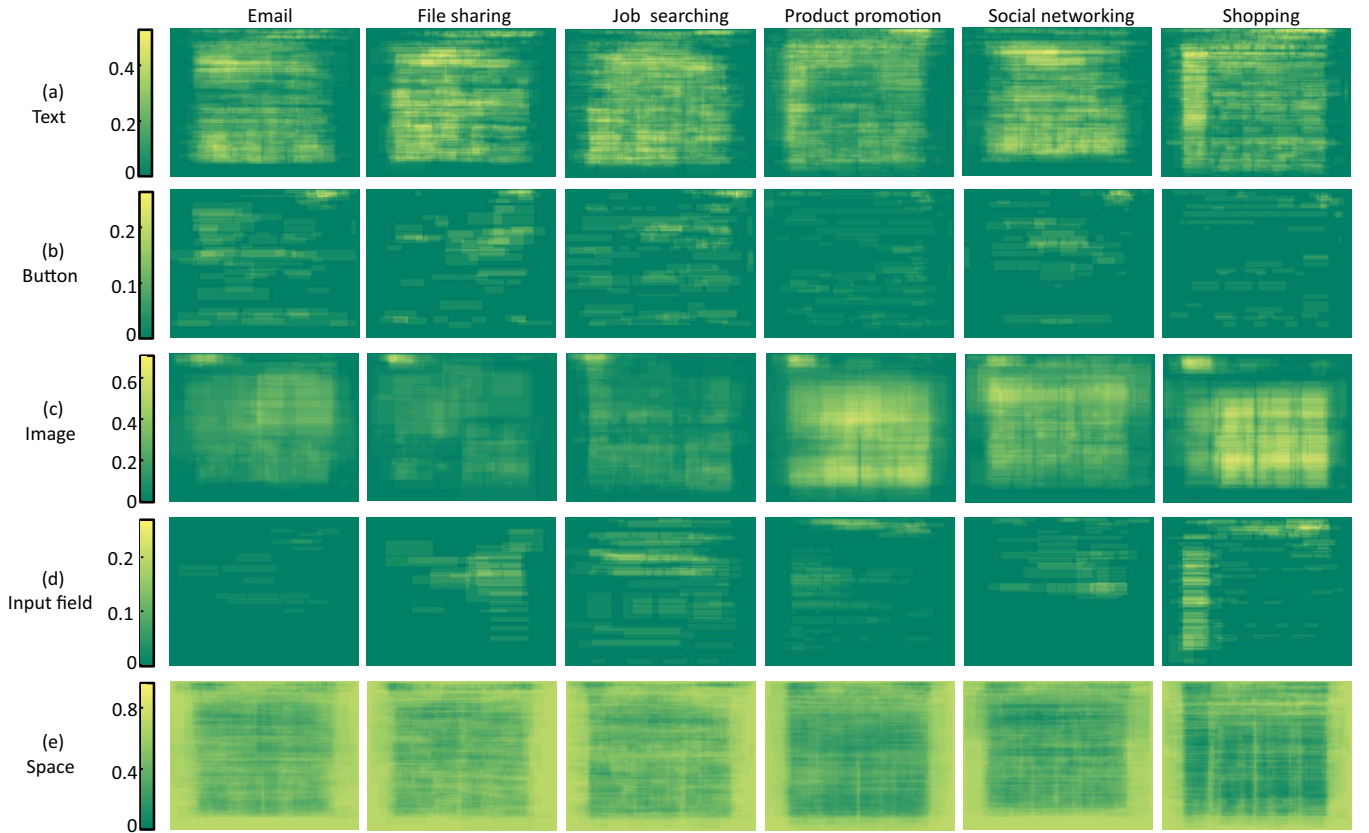

Figure 4: Spatial distributions of page components of different types (a-d) and white space (e) over the entire page for webpages of different categories. The color of a pixel indicates the probability of a page component at the pixel.
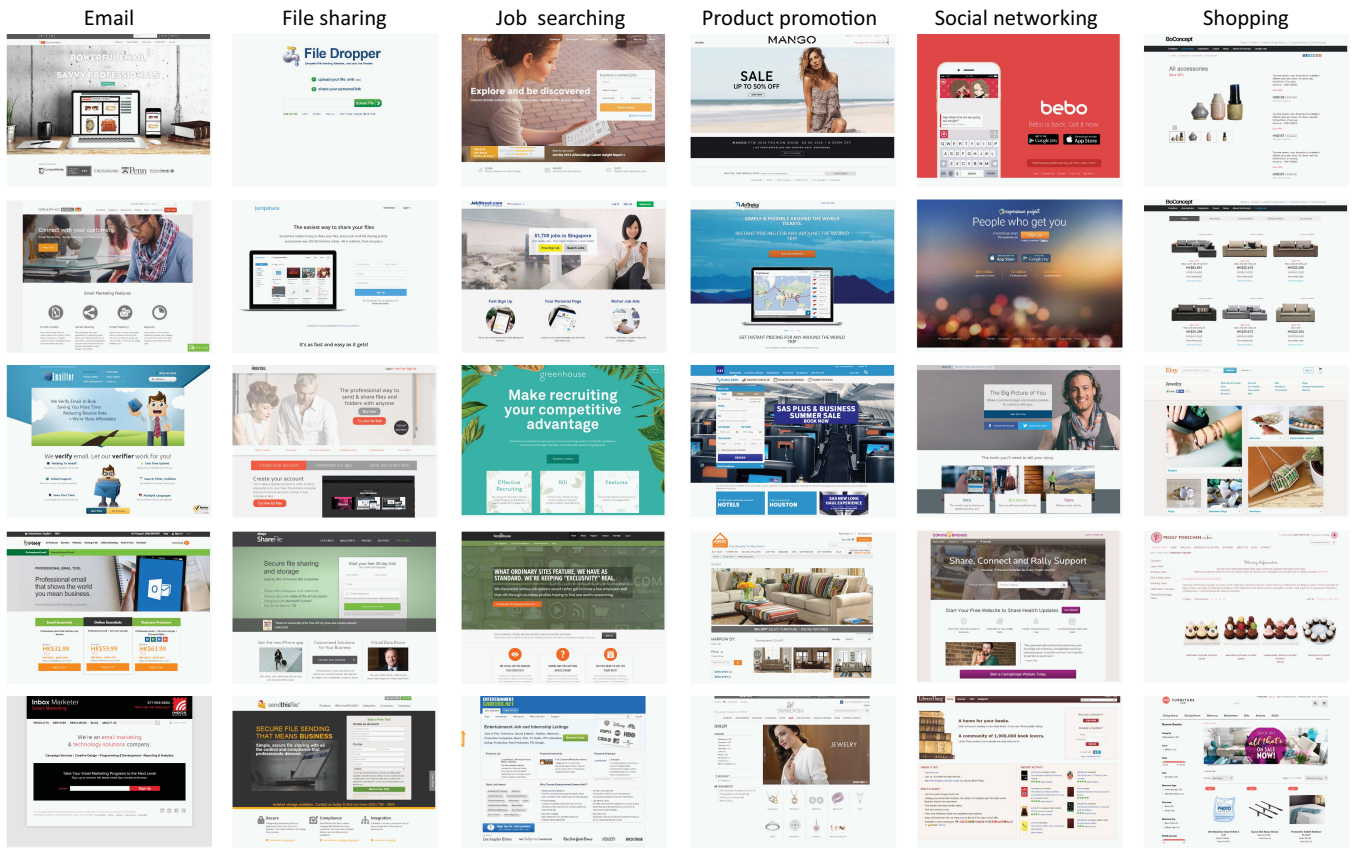
3

Email	File sharing	Job searching	Product promotion	Social networking	Shopping

Figure 5: Sample webpages from our dataset. Each column includes several webpages from a particular category.

Figure 6: Each viewing session contains two types of tasks: comparison task (a) and shopping task (b). Participants move forward in each task via mouse-clicks.

transition $\mathcal{M}^2$ and global eye gaze path $\mathcal{M}^3$. $\mathcal{M}^1$ is a spatial representation that indicates where users look, and is computed as an intensity map by convolving the users' fixation locations on a web design with a Gaussian kernel [JEDT09]. A position in $\mathcal{M}^1$ with a higher value is more visually important. $\mathcal{M}^2$ is a temporal representation that indicates how users move their eyes between each pair of page components in a web design. We construct an attention transition graph from users' eye gaze data, and treat the edge weight matrix of the graph as $\mathcal{M}^2$. $\mathcal{M}^3$ is a spatial and temporal representation that indicates how users move their eyes across the whole web design. For each viewer, $\mathcal{M}^3$ is calculated as a vector formed by concatenating eye fixation positions in temporal order. For each pair of users (k, j) on each web design in our dataset, we calculate the Pearson's correlation coefficient between their corresponding eye gaze representations, i.e., $r_{k,j}^i = corr(\mathcal{M}_k^i, \mathcal{M}_j^i)$. By aggregating across all the web designs in our dataset and across all pairs of users from the same condition, we obtain two *within-condition* correlation coefficients, one for the task-driven condition and another for the free-viewing condition. Similarly, we also calculate a *across-condition* coefficient by aggregating across all pairs of users from different conditions.[1] [SD87].

Figure 7 shows the average correlation coefficients for the three eye gaze representations. For all three representations, the eye gaze data of users from different conditions (T-F) has lower correlation, as compared with those from the same condition (T-T and F-F). This suggests that user eye movements on web designs under the task-driven condition are different from those under the free-viewing condition. With a task in hand, users tend to directly gravitate to the components that are likely to be useful for completing their goals, rather than looking at all information in the free-viewing condition.

S4   Distributions of input paths

The spatial distributions of the input paths of different lengths are shown in a normalized page space in Figure 12. In particular, we represent each input path using the coordinates of its page components'

---

[1]For aggregating correlation coefficients, we convert computed r coefficients to Fisher's z coefficients, average z coefficients, and convert the average back to r coefficients.
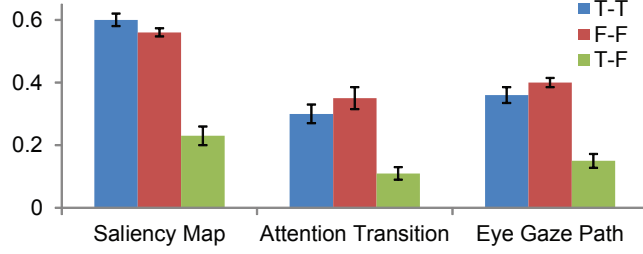
Figure 7: Correlation of eye gaze behaviors of different users under the task-driven condition (T) and the free-viewing condition (F) for three different quantities, i.e., saliency map, attention transition and eye gaze path. For all the representations, the correlation values for T-T and F-F are uncharacteristically similar (independent t-test, $p > 0.05$ for all the representations),while both T-T and F-F have significantly higher correlation values than T-F (independent t-test, $p < 0.001$ for all the representations).

centroids normalized w.r.t. its corresponding page, and plot all the input paths in a normalized page space $[0, 1] \times [0, 1]$. As can be seen, our input paths have a spatially uniform coverage over the entire page space. In addition, in Figure 12, the path segments of all the input paths are plotted in a length-orientation space, where a path segment is represented by its normalized length w.r.t its corresponding page and its orientation. We can see that our input path segments have various length and orientation. This suggests that some of our input paths deviate from basic Z and F paths, where the orientations of path segments are mainly from left to right and top to bottom.

S5 More results

More results generated by our method are shown in Figures 13 and 14.
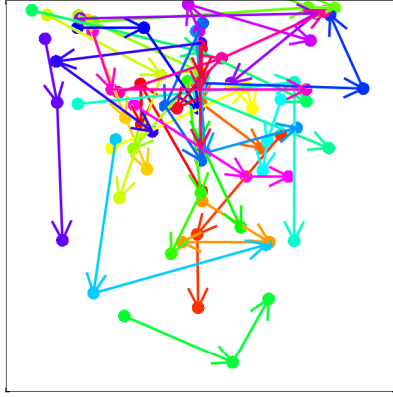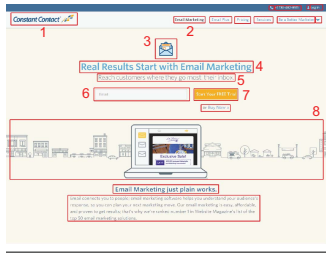
(a) Input paths of length 3

(b) Input paths of length 4
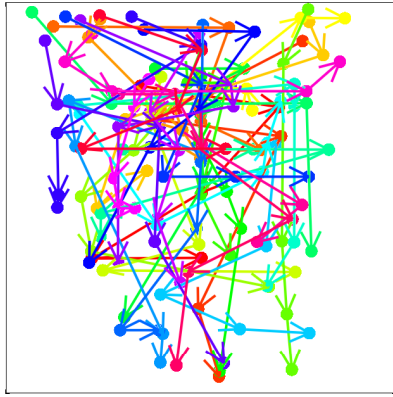
(c) Input paths of length 5
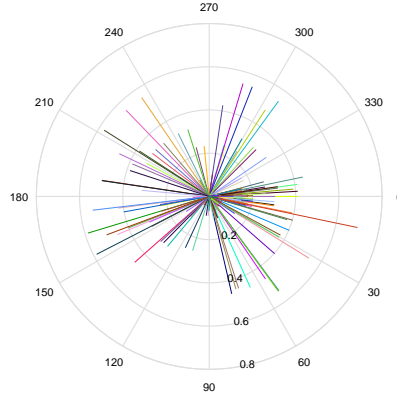
(d) Path segments of all input paths

Figure 12: Input paths of lengths 3, 4 and 5 in a normalized page space ((a), (b), (c)). Each color represents a single input path. Path segments of all input paths in a length-orientation space (d). Each color represents a single path segment.
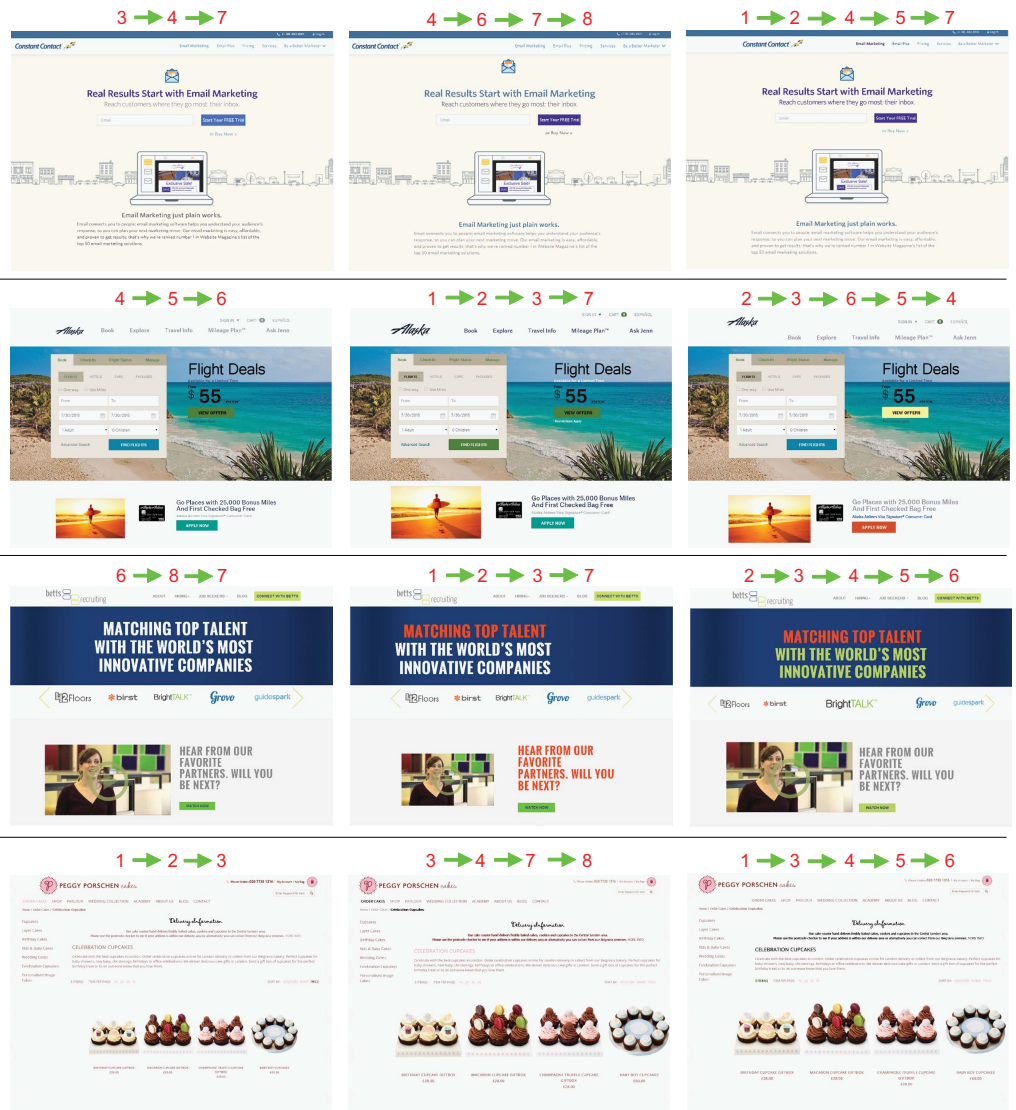
Figure 13: Web designs generated by our method. For each of the input designs (first column), three optimized web designs (second, third and fourth columns) are generated using three different input paths shown on top of the optimized designs.
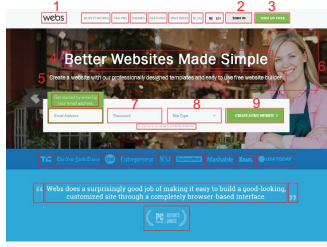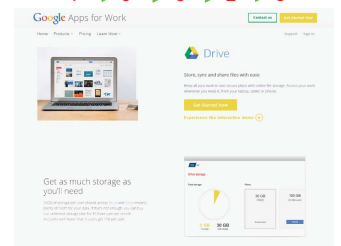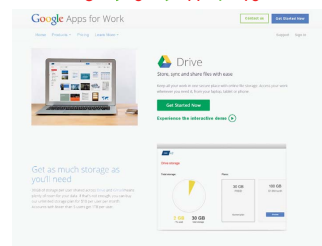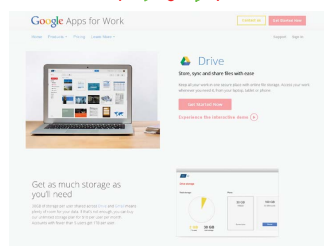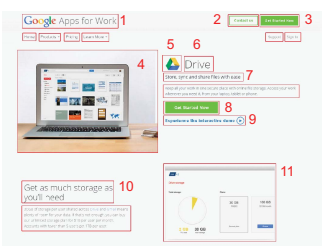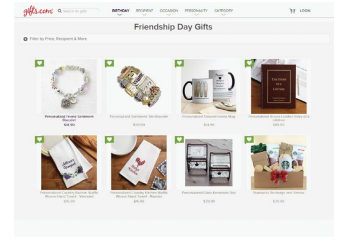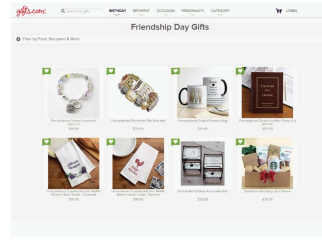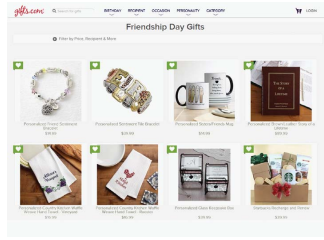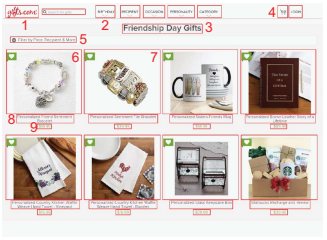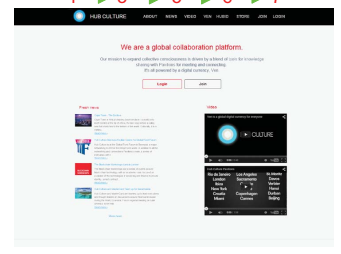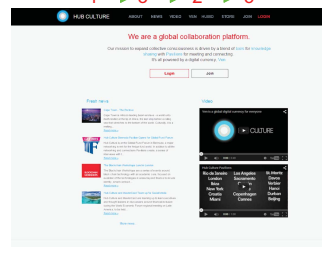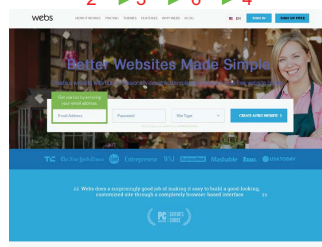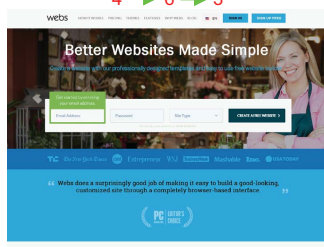
Figure 14: Web designs generated by our method. For each of the input designs (first column), three optimized web designs (second, third and fourth columns) are generated using three different input paths shown on top of the optimized designs.

S6 First-fixation Term

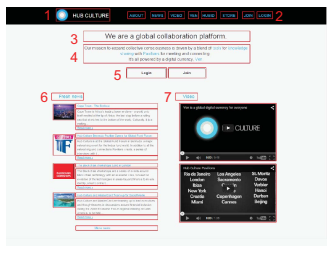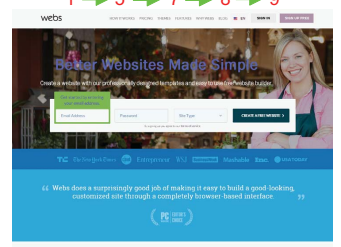The web designs generated by our method with and without the first-fixation term $\mathbf{E}_{att}^F$ are shown in Figure 15.

Figure 15: Web designs generated by our method without and with the first-fixation term $\mathbf{E}_{att}^F$ in our optimizer. With the first-fixation term, users are expected to look at the starting component of an input path first when browsing the web design.

S7   Comparison to Professional Web Designers

The eye gaze path matching rates of web designs generated by the web designers and our method are plotted in Figure 16. The comparisons of web designs generated by web designers and our method are shown in Figures 17 and 18.
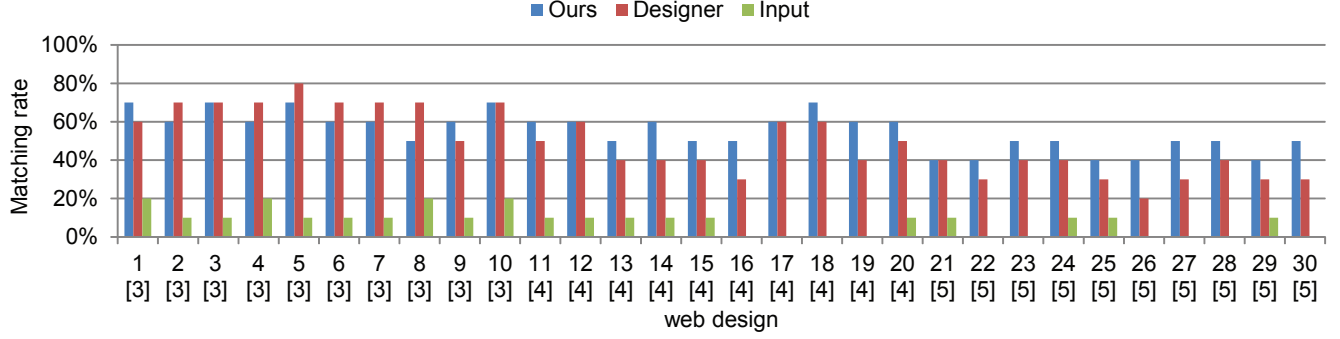
Figure 16: Eye gaze path matching rates of the original input web designs (green), the web designs generated by professional web designers (red), and those generated by our method (blue). The numbers in the square brackets indicate the input path lengths.

| Web design with input path | Designer | Ours |

Figure 17: Web designs generated by professional web designers and our method.

Web design with input path	Designer	Ours

Figure 18: Web designs generated by professional web designers and our method.

S8 A/B Test - Can visual flow affect user behavior?

Creating a visual flow to guide users through a webpage has already been a common practice for web designers to convey information efficiently and help users t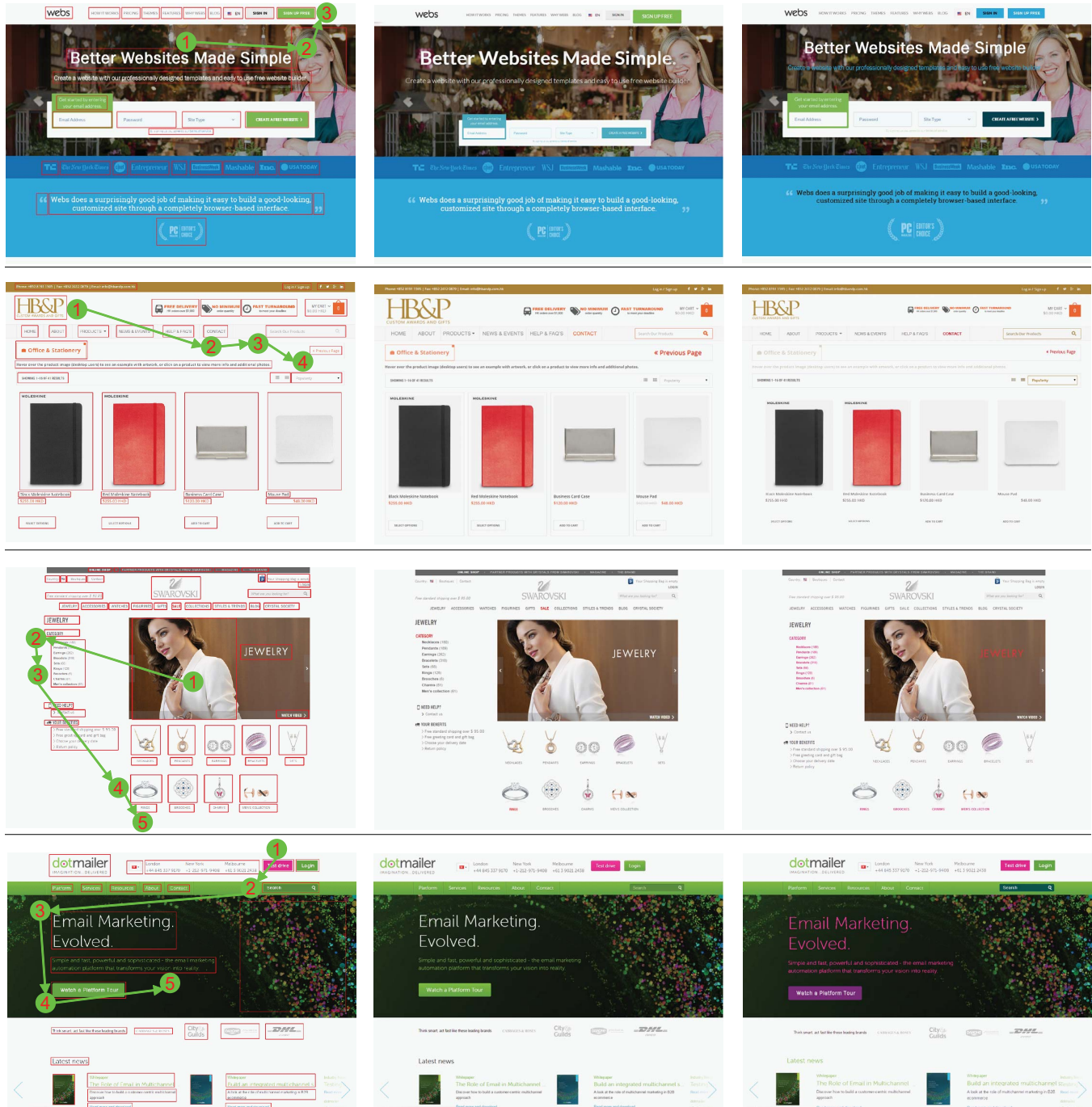o reach their ultimate goals easily and quickly [Cox11, Bra13]. We perform a preliminary A/B test to study whether following a designer-specified path would affect user behaviors in real-world scenarios. We take 2 web designs randomly from each of the 6 categories in our dataset, forming a set of 12 original web designs (original set). We optimize each of the original web designs using our method, according to an input path. This produces a set of 12 optimized web designs (optimized set). For each web design, the input path is set according to one of the goals that the design is expected to achieve (e.g., convince visitors to register an account). More specifically, we set an input path to pass through the regions that carry important information for persuading visitors to make certain decisions and to end with a call-to-action button or a product item to buy. In this way, each input path is intended to guide visitors to take a specific action, which is referred to as a *target action* in our context (e.g., to sign up an email account or buy a particular product item). Figure 19 shows some example input paths for the web designs used in our study.
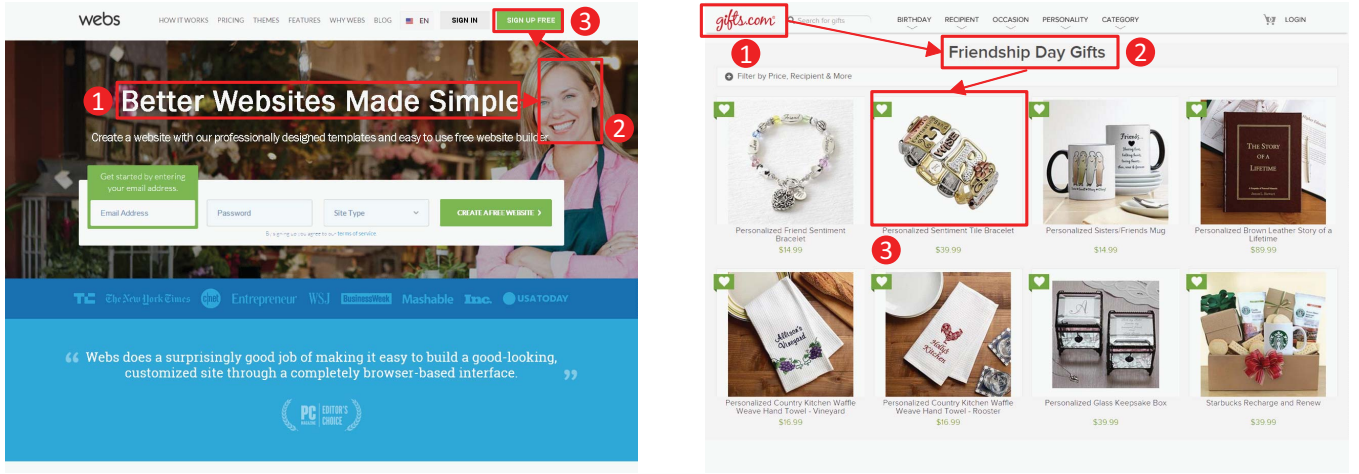

Figure 19: Visual flow (input paths) for web designs used in our A/B test. Each input path passes through the regions with important information and ends with a call-to-action button (left) or an product item to buy (right), and is thus used to guide visitors to take a particular action, e.g., sign up (left) or buy this item (right).

Our hypothesis in this study is that, by moving their eyes along an input path on a webpage, visitors would have a higher chance of performing a target action. We define the percent of visitors taking a target action as the *action rate*. If guiding users along a specific path on a webpage can effectively increase the chance of the target actions being taken, we should observe a higher action rate on the optimized version of a webpage than its original version.

Our study involves 160 participants including 100 male and 60 female, whose age ranges from 20 to 50. All participants have more than 5 years of experience in browsing webpages. Each participant was asked to view 12 different web designs one by one, half of which is randomly selected from the original set and another half from the optimized set. A participant did not view the original and optimized versions of the same web design. The presentation order of the web designs was randomized for each participant. For each web design, a participant was first presented a scenario where he or she was supposed to browse the web design to complete certain objectives (e.g., to find an email service to use, or to buy a gift for their friends). Then, the participant was presented the web design to view, which was displayed in their original resolutions on a 17-inch monitor with a screen resolution of 1280×1024, as in Section S2. Since we would like to test whether the participants could be guided through important page components to make decisions efficiently,

14

we constrained the viewing time of each web design to be within 15 seconds, to avoid some participants exhaustively searching over the whole web design for a long time. Finally, the participant was asked to answer a question about his or her intention to take a target action (e.g., whether they would like to try the free trial or which item they would like to buy). Two examples of the workflow are shown in Figure 20. In the end, we collected the answers from 80 different participants for each of the 24 web designs (12 original designs and 12 optimized designs).
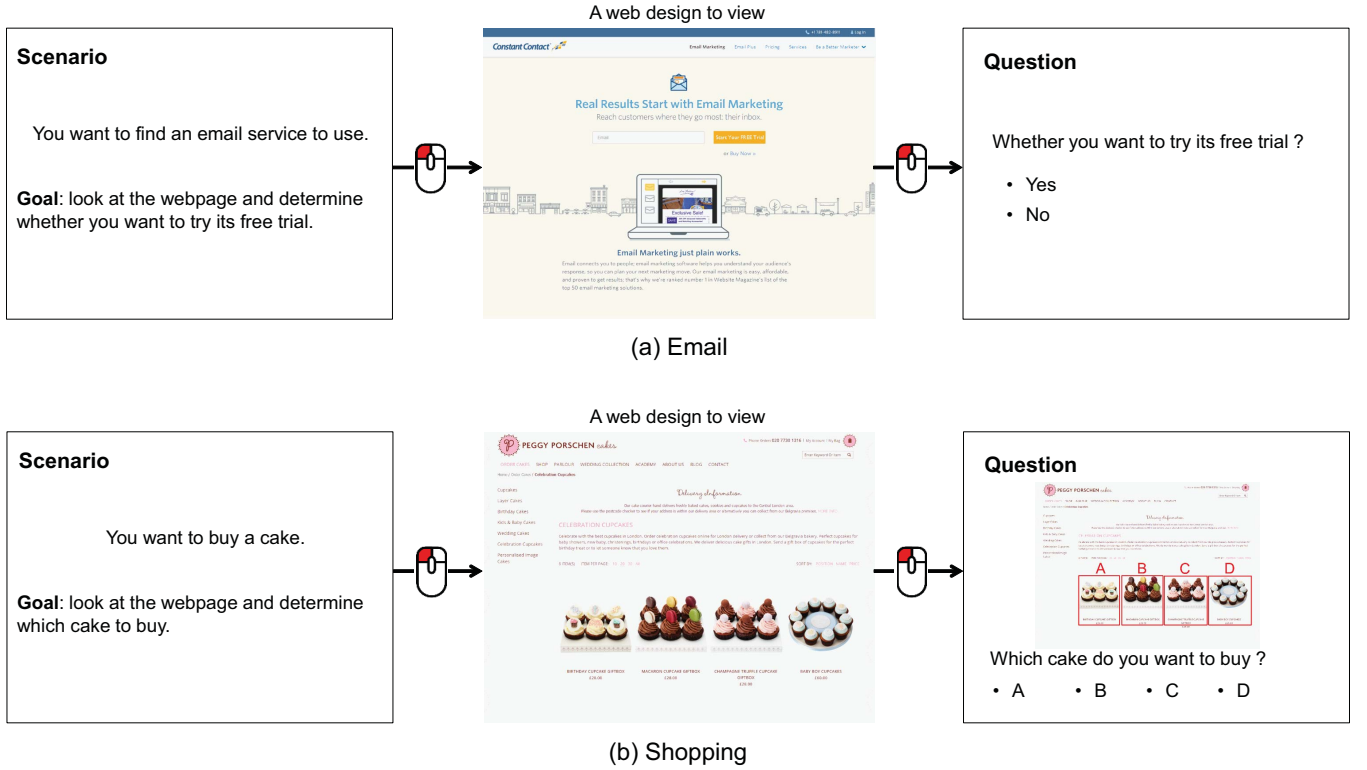

(a) Email

(b) Shopping

Figure 20: Examples of viewing web designs from email (a) and shopping (b) categories and answering related questions in the A/B test.

The results are shown in Figure 21. In general, our optimized web designs achieves significantly higher action rates than the original web designs (independent t-test with $p < 0.001$). However, for the "social" category, the increase in action rate is not statistically significant. Perhaps, this is due to the reason that some original web designs have already been optimized by their creators for similar purposes (e.g., to guide visitors to sign up). Hence, our method cannot produce substantial gains on these designs. Another possible reason is that, in order to impress visitors, social websites tend to highlight the unique taste and features of their communities by using many fancy images and personal photos, which could predominantly grasp user attention and thus make it very difficult to direct user attention along the given paths. We acknowledge that many factors beyond the design (e.g., personal interest and occupation) could also influence visitors' decision making process. A formal A/B test for web designs often requires a large number of visitors to randomize these factors properly. Therefore, while we have shown promising results in our rudimentary test, in future works we would like to perform a large-scale and well-controlled A/B test to more thoroughly evaluate our method as well as how guiding users along visual flows affects their behaviors.
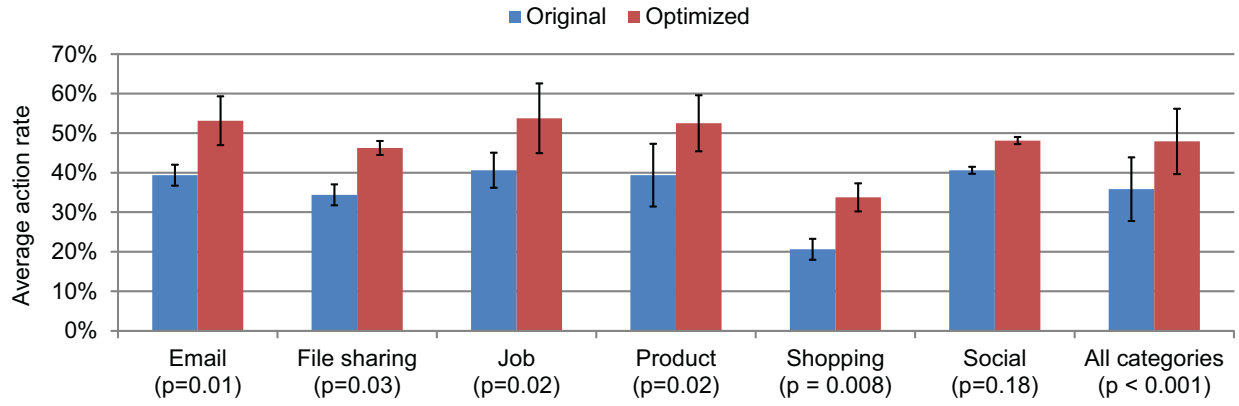
Figure 21: Average action rates of the original and optimized versions for each category and across all categories. The p values of independent t-test are displayed below the labels on the horizontal axis.

References

[Bra13]   S. Bradley. `http://vanseodesign.com/downloads/learn-design-fundamentals/`, 2013.

[Cox11]   P. Cox. `http://tympanus.net/codrops/2011/01/14/visual-movement-flow/`, 2011.

[JEDT09] T. Judd, K. Ehinger, F. Durand, and A. Torralba. learning to predict where humans look. In *ICCV 2009*, 2009.

[SD87]   N. Silver and W. Dunlap. Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology*, 72(1), 1987.