Keep it Simple: How Visual Complexity and Preferences Impact Search Efficiency on Websites

Amanda Baughan[1], **Tal August**[1], **Naomi Yamashita**[2], **Katharina Reinecke**[1]
[1]Paul G. Allen School of Computer Science, University of Washington, Seattle, Washington
[2]NTT Communication Science Laboratories, Kyoto, Japan
{baughan, taugust, reinecke}@cs.uw.edu, naomiy@acm.org

ABSTRACT
Past research has shown that people prefer different levels of visual complexity in websites: While some prefer simple websites with little text and few images, others prefer highly complex websites with many colors, images, and text. We investigated whether users' visual preferences reflect which website complexity they can work with most efficiently. We conducted an online study with 165 participants in which we tested their search efficiency and information recall. We confirm that the visual complexity of a website has a significant negative effect on search efficiency and information recall. However, the search efficiency of those who preferred simple websites was more negatively affected by highly complex websites than those who preferred high visual complexity. Our results suggest that diverse visual preferences need to be accounted for when assessing search response time and information recall in HCI experiments, testing software, or A/B tests.

Author Keywords
Visual appeal; visual complexity; user interface; design; usability; search efficiency; information recall

INTRODUCTION
Designing visually appealing websites has been increasingly recognized as a key requirement for user acceptance and retention [9, 10, 35]. Websites that people perceive as visually appealing are also perceived as more trustworthy, usable, and are thought to increase purchase intentions [10, 13, 14].

In determining appeal, research has shown that visual complexity (sometimes referred to as clutter) is potentially the single most influential factor in what users find appealing [17, 27, 36, 37, 41]. It is also a strong predictor of performance, such as of search efficiency or memory [3, 11, 29, 36], with many findings suggesting that simple or medium complexity websites are more beneficial to users than highly complex websites [4, 11, 29, 33, 36]. However, prior work does not take diverse visual preferences into account.

CHI'20, April 25–30, 2020, Honolulu, HI, USA

ISBN 978-1-4503-6708-0/20/04...$15.00
DOI: https://doi.org/10.1145/3313831.3376849

Demographics such as age, gender, nationality, and education, as well as personality and cognitive styles, all impact whether someone prefers simple over highly complex websites [9, 24, 25]. What is unknown is how diverse visual preferences impact people's search efficiency and information recall when interacting with websites of varying complexity.

We answered this question by conducting an online study with 165 participants with diverse visual preferences. The experiment evaluated whether participants' search efficiency (as measured by their search response time and error rate) and their information recall (as a measure of familiarization with a website) was impacted by their preference for simple, medium, or highly complex websites. Unlike previous research studying the relationship between complexity and performance (e.g., [11, 29, 36]), we designed a study to evaluate the performance of people with diverse visual preferences across websites of varying complexity.

We found that people searched most efficiently and recalled information best when working with simple websites, and worst when working with highly complex websites. This suggests that simple websites are indeed better for users' search efficiency and information recall, even if they have a visual preference for highly complex websites. However, our findings also show that visual preferences do matter: participants who preferred simple websites were more negatively affected by highly complex websites than those who preferred high visual complexity. Given that many website categories, such as shopping or news websites, are necessarily visually complex due to their purpose of displaying much information at once [21], it is important to consider the effects of such website layouts.

Our contributions can be summarized as follows:

1. Our results are the first to show that users' visual preferences impact performance on websites of different visual complexity.

2. We show that one size does not fit all when it comes to the relationship between visual complexity and performance: people who prefer simple websites can find information significantly faster with simple websites than those who prefer highly complex websites. However, a preference for simple websites also means that their performance is more negatively affected by highly complex websites than those who prefer high visual complexity. With websites of medium complexity, people who prefer simple websites

remember significantly less information than those who prefer high visual complexity.

3. We provide a discussion of design implications that caution researchers and designers not to rely on speed of search performance as a metric of usability unless they are aware of or can control for the visual preferences of their users. Additionally, we discuss the permanence of judgements of visual appeal and introduce the idea of an adaptive user interface that could support future research.

BACKGROUND AND RELATED WORK

Visual appeal influences whether people think websites are usable, trustworthy, and affects their overall satisfaction [8, 13, 14]. A strong predictor of what users find appealing is the visual complexity of a website [2, 17, 27, 37, 41], which has been shown to be the stronger predictor of appeal between colorfulness and complexity in Reinecke et al. [27]. Researchers have extended Berylne's [1] aesthetic theory to visual complexity, in which a moderate amount of complexity is thought to be most pleasurable. Several studies based on this theory have shown a relationship between visual complexity and perception. Geissler et al. [7] found that consumers respond more favorably towards websites that fall within a moderate range of perceived complexity. There is also evidence of a negative correlation between complexity and pleasure in website perception [23]. Additionally, visual complexity may influence attitudes and intentions to use a website, and simple or medium complexity websites may be best received by users [4, 33]. However, these studies do not account for variations in perceptions of appeal across people from different demographic backgrounds, as shown in Reinecke and Gajos [25].

Several researchers have cautioned against making one-size-fits-all assumptions about visual complexity and have instead suggested that what users perceive as good design is subject to demographic and psychological differences [9, 24, 25, 27]. Gender, personality, education, nationality, and age all affect the level of visual complexity people perceive as most appealing [25, 27]. Past work has also shown that people's cognitive styles impact how they navigate websites and what website layouts they prefer. For example, those who are more impulsive and visual prefer to have less content and text, but more images. In contrast, those who are more deliberate and verbal tend to prefer more details and text [9]. Adapting website interfaces accordingly has been shown to lead to higher user engagement [9].

Past research studying the relationship between visual appeal, complexity, and performance has typically been measured as task completion time and information recall. The results have been partially contradictory. Some studies have suggested that visual appeal increases time spent on appealing stimuli, as "prolongation of a joyful experience" [31]. In contrast, other studies have shown that highly appealing aesthetics lead to decreased time spent on task completion, as "increased motivation" [18, 30, 32]. Follow-up work has further refined this idea, indicating that aesthetic appeal facilitates faster response time only under duress [28]. Additionally, some studies have found no correlation between appeal and task completion time [8, 34]. Regarding information recall, prior work suggests that recall is higher in more aesthetically pleasing websites, especially if the website is "prototypical," or within expectations of what a website should look like [6]. Additionally, low visual complexity websites were better remembered by Swiss participants in Tuch et al. [36], and high visual complexity led to poor search and memory performance. In contrast, research in the field of visualization has shown that highly complex visualizations were more memorable [3].

However, none of the prior work has evaluated the complexity level that users prefer across a large heterogeneous sample of stimuli and incorporated diverse preferences in their analysis of performance. This work therefore extends prior work by evaluating performance across websites of varying complexity with participants with diverse visual preferences.

METHODS

We conducted an online experiment to determine whether preference for a certain level of visual complexity in websites impacts search efficiency and information recall.

Hypotheses

To evaluate the relationship, we first established a set of hypotheses based on our research question and related work.

First, much work has shown that visual complexity in websites negatively affects different measures of usability (e.g., [4, 7, 33, 36]). For example, Tuch et al. [36] showed that website visual complexity increases reaction times in a visual search task and decreases recognition rates [36]. We therefore expect that our participants will perform most efficiently (i.e., take the least amount of time and misclicks) with simple websites.

[H.1] The visual complexity of a website has a significant negative effect on search efficiency, independent of participants' diverse preferences.

Our second hypothesis is based on the finding that people's visual preferences for websites vary depending on a variety of factors, such as age, education, gender, or country of origin [25]. While a website's visual complexity is most predictive of an overall preference for certain websites [27], not everyone likes the same level of visual complexity [25]. Hence, a diverse sample of participants should result in different preferred levels of visual complexity. If so, then we expect that participants' search efficiency will be positively affected by websites that correspond to a participant's preferred level of visual complexity. In other words, participants who prefer simple websites will perform better with simple websites than with visually complex websites, and vice versa.

[H.2] There is a significant interaction effect between visual preference and website complexity on search efficiency.

We also expect the same to be true for recognizing information provided by websites. If participants are engaged in the primary task of finding information, visually complex websites have been shown to negatively influence recognition rates [36].

[H.3] The visual complexity of websites has a significant negative effect on information recall, independent of participants' diverse preferences.

Because we expect varying preferences for different levels of website complexity, our fourth hypothesis focuses on the impact of visual preferences on information recall:

[H.4] There is a significant interaction effect between visual preference and website complexity on information recall.

Procedure

To test our hypotheses, we designed a within-subjects online experiment with two parts: Part 1 assessed participants' preferred visual complexity level, and Part 2 evaluated their performance in a visual search task.

The experiment began with a brief overview of the study, a consent form, and a demographics questionnaire. To assess participants' preferred visual complexity level (low, medium, high), Part 1 asked participants to rate the visual appeal of 12 website screenshots (out of 23 screenshots total), plus one practice screenshot of medium complexity at the start. Recall that visual complexity is the main predictor of visual appeal [27]; we therefore chose to ask people to rate websites on visual appeal to indirectly assess a preference for website complexity. Following the procedures in [15] and [27], participants were only shown each website screenshot for 500ms and subsequently asked to rate the screenshot's visual appeal on a Likert scale of 1 (very unappealing) to 9 (very appealing).

After the practice trial, each participant rated four website screenshots each of low, medium, and high complexity, presented in random order. The levels of website complexity were determined using a computational model of perceived website complexity, as detailed in the Materials section.

Part 2 of the experiment focused on assessing participants' search efficiency and information recall and consisted of six trials (showing two websites each of low, medium, and high complexity) in random order. Each trial presented participants with a scenario, such as "You want to buy office furniture. Click the link to filing cabinets." These scenarios were designed to represent common search activities on the web. In addition, the targets, such as "filing cabinets" in the example above, were chosen to represent various locations across the website where participants may usually find information. Once a participant had read through the scenario, they could proceed to the next page, which showed a website screenshot along with the scenario as a reminder. Participants were asked to find and click on the target as fast as possible. Clicks were accepted as correct if they were made within a 5 pixel boundary of the target. If participants clicked outside of this boundary, a red line framed the screenshot and a message to try again was shown below the screenshot. For the first 90 participants (55%), a correct click was required to proceed through the study. We later added an option to skip a trial.

Once participants clicked on the target, they were presented with two questions per trial to assess information recall. One of these questions focused on a detail within the main content of a webpage ("What is the name of the online store?"), and the other focused on peripheral information on the webpage ("What is the store's free delivery radius?"). Participants were given three answer options for each question (the correct answer, an incorrect answer, and "I don't know"). The order of the questions and the order of the correct/incorrect answers were randomized. Participants were then given the opportunity to report any technical difficulties and provide comments or questions. The final page showed their average task completion time compared to others.

To ensure a consistent viewing time of 500ms and a consistent time measurement for the search task, all screenshots were preloaded to minimize any effect of Internet bandwidth. The study was presented in English (including all website screenshots) and took 6 minutes on average to complete.

Materials

The experiment included two sets of 12 <u>website screenshots</u> for Part 1 of the experiment (four at low, medium, and high complexity) which participants were asked to rate on visual appeal. The second part of the experiment asked people to find information on six websites (two at low, medium, and high complexity), which were drawn from a larger set of 12 website screenshots. We also included four practice screenshots. We used an existing dataset of 80,901 website URLs [21] to select websites that represented a variety of topics, were written in English, and had not received wide public exposure, which was evaluated as having a global and in-country rating of over 100 per Alexa's top sites ranking. Website URLs were selected from finance, journalism, education, e-commerce, software as a service, travel, and entertainment. For each URL, we took a website screenshots at 1024 x 768 pixel resolution. We used "real-world" websites (instead of systematically changing the complexity level for one design) to avoid learning effects and increase external validity.

To ensure that the final set of websites included the same number of websites with a low, medium, and high complexity, we computed their perceived visual complexity using the computational model introduced in [27]. We then visually verified the computed visual complexity score before deciding on a final set. Across the three sets of low, medium, and high complexity, websites were chosen to be as consistent as possible in their colorfulness, to minimize the impact of another main predictor of appeal [27]. Figure 1 shows examples of low, medium, and high complexity websites used in the experiment.

The second part of our experiment included <u>scenarios</u> and <u>questions</u> for each of the 12 website screenshots (of which a participant viewed only 6). The scenarios were created to mimic real world search scenarios across the different website domains. Questions were separated into one regarding a detail in the main content of the website, and another regarding a detail in the periphery content of the website. Each question had the options of a correct answer, an incorrect answer, and "I don't know."

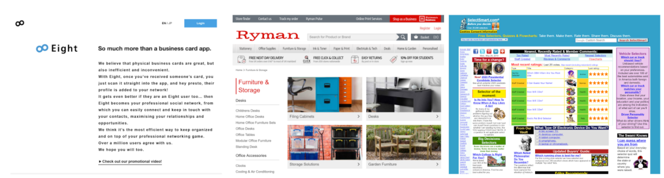


Figure 1. Left to right: Low, medium, and high complexity websites used in the visual search task.

Metrics

We recorded the following metrics:

Website complexity: The level of complexity of each website stimulus, encoded as an ordinal value (low < medium < high).

Preferred complexity level: A participant's preference for a certain level of visual complexity. We followed prior literature's methods to assess appeal across a variety of websites [15, 27] and then inferred preferred complexity level. We aggregated the Likert scale appeal ratings from Part 1 of the experiment using the median rating of the four websites for each level of complexity (low, medium, high). This resulted in one median appeal score per complexity level for each participant. The highest of these scores was assigned as their preferred complexity level. If two or three median appeal scores were the same for a participant, we assumed that the participant did not have a clearly preferred complexity level and excluded them from further analysis. The preferred complexity level was encoded as an ordinal value.

We used three independent sample's t-tests to confirm that these preference groups were age-balanced, as age has been shown to influence speed-based task performance [12, 38]. Additionally, we found that English proficiency had a significant effect on search response time (detailed in the Analysis section) .We therefore tested for significant differences between preference groups by conducting three Wilcoxon signed-rank tests across preference groups. English proficiency was not significantly different between preference groups.

Search efficiency: Similarly to Henderson et al.'s [11] definition of search efficiency, we evaluate both *search response time* and *errors*. *Search response time* was the time between when the screenshot was displayed and the participants' clicked on the target. In the analysis to follow, residuals of a fitted linear mixed effects model for *time* were not normally distributed, so log transformed values were used for statistical analysis. *Errors* were recorded as the number of incorrect clicks on the screenshot. Skipping the task was recorded as a true or false value, and skipped tasks were removed from analysis.

Information Recall: The questions in Part 2 of our experiment were intended to evaluate information recall. Note that this is different from "recognition" as studied by Tuch et al. [36], who asked participants to judge whether they had seen a website in an experiment that they participated in a week before. Instead, we define information recall as the peripheral memory that helps us assess to what degree participants have familiarized themselves with a website and can recall specific website elements. Since there were two questions per trial, participants could get 0, 1, or 2 questions correct per trial. This was recorded as a continuous variable and aggregated into sums of correct answers for low, medium, and high complexity websites by grouping the number of correct answers per person per complexity level.

We additionally recorded participant demographics that have been shown in prior literature to impact people's visual preferences [12, 25, 38] or could be potential confounds for performance, including hours spent on a computer and input device (see Table 1 for an overview).

Table 1. Demographic distribution across preference levels. Only age and English proficiency levels had a significant effect on performance.

	low	medium	high
n	39	93	33
Age mean (SD)	30.1 (11.7)	28.8 (13.0)	29.3 (12.0)
Gender			
Male	23 (59%)	33 (40%)	13 (48%)
Female	13 (33%)	49 (59%)	14 (52%)
Non-binary	3 (8%)	1 (1%)	0
English Proficiency median (SD)	Native (1.1)	Native (1.2)	Native (1.4)
Americans	15 (38%)	38 (41%)	14 (42%)
Hours on Computer mean (SD)	6.2 (3.9)	6.5 (4.0)	6.9 (3.9)
Input Device			
Mouse	25 (64%)	43 (46%)	18 (55%)
Trackpad	10 (26%)	42 (45%)	11 (33%)
Finger	3 (8%)	7 (8%)	4 (12%)
Trackball Mouse	1 (3%)	1 (1%)	0

Participants

Participants were volunteers recruited through the experiment platform LabintheWild [26]. Altogether, 218 people completed the study. We removed 25 participants who indicated they had previously completed the study, experienced technical difficulties, or cheated in any way. We also removed three people who chose the same rating for every website (so-called "straight-lining"), as well as 25 people who did not have a clear preferred complexity level. The final number of participants included in the following analyses was 165.

Of the 165 participants, 40 different countries were represented with a plurality of participants coming from the United States ($n = 67, 40.6\%$). Participants were between 12 and 75 years old ($M = 29.2, SD = 12.5$), and the gender distribution was 43% men, 46% women, 2% non-binary, and 9% chose not to disclose. 87% ($n = 144$) of participants reported proficiency in English or higher, as measured on a 5-option scale labeled with limited knowledge, conversational, proficient, fluent, and native. This scale was adapted from a 5-point Likert scale to measure language fluency used in [5, 40]. Participants spent 6.5 hours on the computer per day on average ($SD = 3.9$), with a majority of them (52%) using a mouse as an input device, followed by a track pad (38%).

RESULTS

Statistical Analysis

Search Response Time. To evaluate the impact of website complexity and preferences on search response time, we fit two linear mixed-effects models with time and log-transformed time as dependent variables. A Shapiro-Wilk test indicated that the residuals of the log-transformed time model ($W = .987, p < .01$), were more normal than that of the raw time model ($W = .948, p < .001$). We therefore report on the linear mixed-effect model with log-time as the dependent variable and participant and website ID as random effects.

Fixed effects were website complexity, preferred complexity level, as well as the interaction between the two. In addition, an initial model included daily hours spent on a computer, input device, gender, and nationality, none of which had a significant effect on the model. These variables were therefore excluded from our final model. Age and English proficiency were included as control variables since age is known to affect visual search and motor ability [12, 38] and English proficiency had a borderline significant effect on search response time.

Table 2 shows the results of the final mixed-model. We additionally conducted *post hoc* tests to evaluate the effects of complexity and preference on search response time. All *post hoc* t-tests were corrected with Holm's sequential Bonferroni procedure.

Errors. A Chi-Squared goodness-of-fit test was run on errors for all three levels of preference, which indicated non-detectable deviations from a Poisson distribution for all levels except preference for medium complexity ($\chi^2(1,165) = 21.78, p < .001$). Even with the deviation, this distribution fit the data better than any other distributions evaluated (negative binomial, logarthmic, linear, quasipoisson, and exponential). Errors were modeled as a function of complexity, performance, and their interaction. The control variables tested included age, English proficiency, computer literacy, nationality, gender, and input device, with participant ID as a random effect. The demographic variables were removed from the model because they did not have a significant effect.

Information Recall. Before analyzing the effect of website complexity and preference on information recall, we first removed questions answered with "I don't know" from the dataset, resulting in 738 answered questions for 121 participants (out of 165 total).

We then ran a linear mixed-effects model with fixed effects of website complexity, preferred complexity level, their interaction, and the demographic variables: age, gender, nationality, average hours spent on a computer, input device, English proficiency and time to complete the trial task. Participant ID was modeled as a random effect. The demographic variables and time were removed from the model because an ANOVA on the mixed-model revealed that they did not have a significant effect. The result of the Shapiro-Wilk test on the residuals of the model with all variables was not statistically significant ($W = 0.991$, $p = n.s.$), indicating compliance with the normality assumption. After removing these variables the residuals of the model indicated a deviation from normality ($W = 0.986$, $p < .01$). We repeated the above process with a variety of other models, including Poisson, quasi-poisson, logarithmic, and negative binomial. The residuals of the linear mixed model deviated the least from normality, so we use it to analyze our results.

We ran *post hoc* tests to evaluate the effect of website complexity and preferred complexity level on correct answers. All *post hoc* t-tests were corrected with Holm's sequential Bonferroni procedure for multiple comparisons.

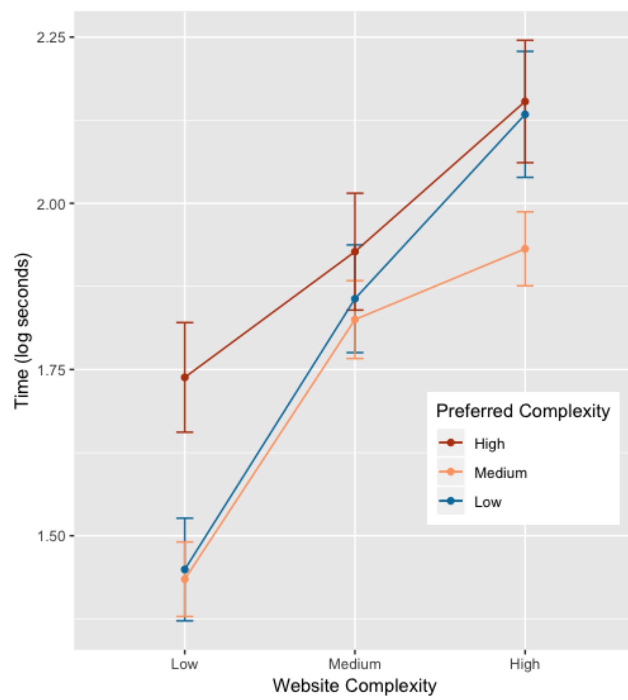
Hypothesis Testing

H.1: Visual Complexity Negatively Affects Search Efficiency

Our analysis revealed that website complexity has a significant negative effect on search efficiency, confirming our first hypothesis (see Table 2). A linear mixed-model ANOVA confirmed the significant effect of complexity on search response time ($F(2,8.99) = 4.44, p < .05$). Participants were fastest on simple websites ($M = 4.2$s, $SD = 4.7$s) and slowest on highly complex websites ($M = 7.3$s, $SD = 7.0$s), regardless of their preferred complexity level.[1] Figure 2 shows this negative correlation between search response time and website complexity for all three preference groups. All participants were statistically significantly faster with low vs. high complexity websites, and those who preferred medium and low complexity were also statistically significantly faster with low vs. medium complexity websites (see Table 3).

While participants' time and errors are correlated (i.e., participants who make many errors will naturally need more time to find the target), several participants who took longer-than-average did not make any errors. Additionally, we found no relationship between complexity and error rate from our ANOVA on the mixed-effects Poisson regression ($\chi^2(2, N = 165) = 0.01, n.s.$).

Figure 2. Relationship between search response time, website complexity, and participants' preferred website complexity. Error bars show the standard error. Across all preference groups, website complexity significantly impacts time spent in visual search tasks. Note that the log-transformed time does not correlate with absolute values, but rather shows the relative difference between groups.



[1] Means and standard deviations were calculated from the non-transformed time data.

Table 2. Results of the linear mixed effect model for log-time. Higher website complexity positively correlates with more time spent on the visual search task, as does age. English proficiency correlated with reduced time spent on tasks, as does the interaction of website complexity and preference. Note that "complexity" refers to website complexity, and "preference" refers to preferred complexity level.

	β	std. error	Z	p
(Intercept)	1.581	0.121	13.07	<.001
Complexity	0.406	0.137	2.96	<.01
Preference	0.024	0.073	0.329	n.s.
Age	0.014	0.003	4.667	<.001
English Proficiency	-0.415	0.106	-3.915	<.001
Complexity:Preference	-0.154	0.071	-2.169	<.05

H.2: Preferences Interact with Complexity to Influence Search Efficiency for Low and High Complexity

Our regression analysis revealed a significant interaction effect between website complexity and preferred website complexity on speed (Table 2). As shown in Figure 2, people who prefer low complexity websites are more negatively impacted by highly complex websites than those who prefer highly complex websites. This confirms our second hypothesis, albeit only for participants who prefer low and high complexity.

Figure 2 also provides more insights into the magnitude of the difference between participant groups with different preferred website complexity. For example, while participants who preferred low and medium websites perform similarly with low complexity websites ($t(249) = 0.15, n.s.$), participants with a preference for highly complex websites were significantly slower than those who preferred low complexity ($t(129) = 2.54, p < .05, d = 0.45$). Our analyses also showed that all preference groups performed similarly with medium and high complexity websites (See Table 3).

While the interaction of complexity and preference correlates with speed ($\beta = -0.15, p < .05$), an ANOVA of the mixed-model revealed that there was not a significant effect ($F(4, 727.07) = 1.53, n.s.$). As this deviates from the model and the graph in Figure 2, we created a follow-up linear model isolating only those who prefer low and high visual complexity and their time when searching on low and highly complex websites. In this model, the ANOVA revealed that there was a significant effect of the interaction of complexity and performance on speed ($\beta = -0.17, p < .05$, $F(1, 184.82) = 5.72, p < .05$). This partially confirms our second hypothesis, showing that there is a significant interaction of preference and complexity, albeit only for people with a preference for low or high visual complexity. Those who prefer low complexity websites more than double their time spent on high (8.7s) versus low (4.2s) complexity websites, while people who prefer high complexity websites only spent 70% more time on high (8.9s) versus low (5.2s) complexity websites.The effect size also reflects this, as people who prefer low complexity have a large effect ($d = .95$) of using high versus low complexity websites, compared to those who prefer high complexity ($d = .67$).

We found a relationship between preference, complexity, and error rate from our ANOVA on the mixed-effects Poisson regression ($\chi^2(4, N = 165) = 11.45, p < .05$). However, *post hoc* tests revealed no statistically significant relationships across complexity levels or preference groups.

Table 3. Results of pairwise t-tests evaluating significant differences in speed across levels of complexity and preference. All t-tests were corrected with Holm's sequential Bonferroni procedure. H, M, and L are used for high, medium, and low complexity.

Preference	Complexity	t-test	p value	Cohen's d
H	H vs L	$t(101) = 3.36$	<.01	0.67
H	H vs M	$t(107) = 1.74$	n.s.	
H	M vs L	$t(118) = 1.56$	n.s.	
M	M vs L	$t(350) = 4.82$	<.001	0.51
M	H vs L	$t(327) = 6.25$	<.001	0.51
M	H vs M	$t(325) = 1.31$	n.s.	
L	M vs L	$t(148) = 3.64$	<.01	0.59
L	H vs L	$t(138) = 5.65$	<.001	0.96
L	H vs M	$t(140) = 2.24$	n.s.	
H vs L	L	$t(129) = 2.54$	<.05	0.45
H vs L	H	$t(110) = 0.14$	n.s.	
M vs L	M	$t(249) = 0.30$	n.s.	
H vs M	M	$t(236) = 0.92$	n.s.	
H vs L	M	$t(137) = 0.59$	n.s.	

H.3: Visual Complexity Negatively Affects Information Recall

We found a significant negative effect of website complexity on information recall, confirming our third hypothesis ($F(2, 299.53) = 11.21, p < .001$), see Table 4 for the results of the linear model). Participants generally answered the most questions correctly with low complexity websites ($M = 1.3, SD = 0.8$) and the least with high complexity websites ($M = 0.8, SD = 0.8$).

People who preferred medium complexity remembered more from low than medium complexity websites ($t(220) = 3.80, p < .01, d = 0.40$)) and more from low than high complexity websites ($t(220) = 4.22, p < .001, d = 0.45$). All other within group comparisons were statistically insignificant, including those who prefer high complexity (note that the graph in Fig. 3 shows standard error, not confidence intervals). Grouping across all participants, people answered more questions correctly from low complexity websites than either high ($t(368) = 4.19, p < .001, d = 0.27$) or medium ($t(368) = 3.30, p < .01, d = 0.35$), as illustrated in Figure 3. There was no statistically significant difference between correctly answered questions on high and medium complexity sites ($t(368) = 0.70, n.s.$).

Table 4. Results of the linear mixed effect model for information recall. High complexity negatively correlates with more correct answers.

	β	se	zval	pval
(Intercept)	1.163	0.067	17.358	<.001
Complexity	-0.317	0.068	-4.662	<.001
Preference	0.236	0.131	1.802	.072
Complexity:Preference	-0.319	0.146	-2.185	<.05

H.4: Preferences for Low and High Complexity Influence Information Recall with Medium Complexity Websites

We found a significant interaction effect between complexity and preference on information recall ($F(4, 297.49) = 2.65, p < .05$). In *post hoc* tests, we found that information recall was only statistically significantly differentiated in one scenario: people who preferred high visual complexity remembered more from medium complexity websites than people who preferred low complexity ($t(128.6) = 2.86, p < .05, d = 0.45$), but more not statistically significantly more than those who prefer medium complexity ($t(111.5) = 2.13, n.s.$). Once again, our hypothesis on the effect of the interaction of preference and complexity is confirmed only for those who prefer simple and highly complex websites.

DISCUSSION

In this work, we set out to investigate the impact of visual complexity on search efficiency and information recall in websites, but with a new perspective: given that prior work has shown that people prefer different levels of website complexity, we aimed to evaluate whether their preferences are indicative of the type of website they can work with most efficiently.

Our results show that all participants, independent of their preference for low, medium, or high complexity websites, found and recalled information best with low complexity websites, and worst with high complexity websites. This confirms and extends prior work [36] by showing that the benefit of simple websites for visual search tasks remains true for a diverse set of participants with different visual preferences.

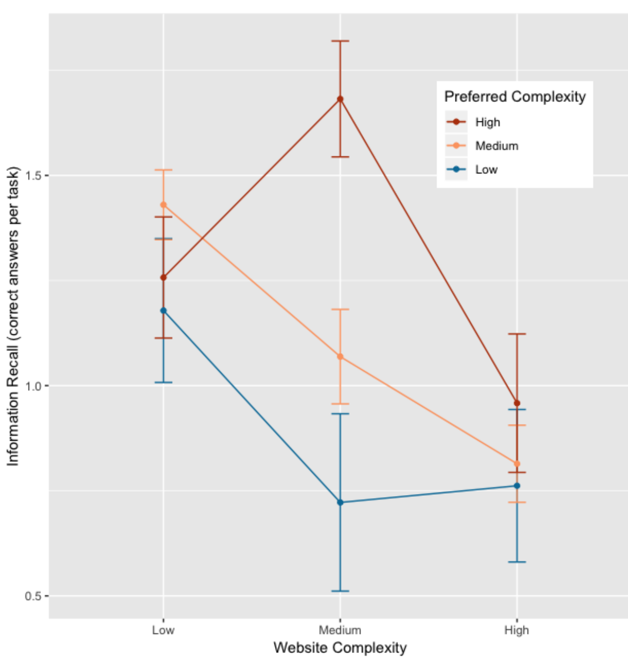


Figure 3. Relationship between information recall (correct answers), website complexity, and participants' preferred website complexity. Error bars show the standard error. Website complexity significantly impacts information recall regardless of preferences. Participants who prefer high visual complexity remembered significantly more from medium complexity websites than those who prefer low visual complexity.

However, our analyses showed that visual preferences do affect performance with websites of varying levels of complexity. In particular, we found that the negative effect of website complexity on search efficiency was amplified by a preference for simple websites: People who prefer simple websites are more negatively affected by highly complex websites than those who prefer high visual complexity.

Our results also showed an interaction effect between website complexity and preferences on information recall: People who prefer high complexity recalled significantly more information than people who prefer low complexity when interacting with medium complexity websites. While this confirms our hypotheses, the results are nuanced: our results only indicate a significant interaction effect between website complexity and preference for participants who prefer simple or highly complex websites.

There are several possible explanations for why people who prefer highly complex websites performed slower with low complexity websites than those who prefer a low level of complexity. It may be that preference groups differ in their cognitive styles, which have been shown to impact what content people pay attention to and how they prefer websites to look [9]. People who prefer low complexity websites may be more impulsive and visual, whereas people who prefer high complexity websites may tend to be verbal and holistic thinkers, focusing on the text and relationships of website elements [9, 20]. This could mean that people with a preference for highly complex websites approach websites more deliberately, and hence, take more time.

Additionally, it is likely that people tend to spend more time in the real world on websites with a level of complexity that they like, as visual appeal is linked to higher trust and intention to use a website [14]. Additionally, people from specific countries are likely to spend more time interacting with websites that are localized to their specific country or language [21, 25]. This could mean that they are especially familiar with these kinds of websites and can process them efficiently under time pressure because "practice makes perfect" [16, 28, 39]. There could be a cultural influence to both web design preferences (such as shown in [21, 25]) and search styles. For instance, there is evidence that people from East Asia may be more holistic thinkers and prefer high complexity websites [20, 25].

Finally, to contextualize our findings with prior works which have found appealing stimuli to be related to both more time spent [31], less time spent [18, 30, 32], and no relationship at all [8, 34], this work is the first to evaluate the effects of various preferences and complexity levels on time spent and information recall. The varied results in prior works may be due to different visual preferences of participants across studies, which were unaccounted for.

It is also possible that visual appeal and preferences are not traits that permeate all interactions on the web. They may be states, meaning that they are temporary and contextually influenced. Visual appeal is often measured by showing screenshots of websites for 50-500ms in order to limit the effect of the website content itself on ratings of appeal [15, 27, 37].

However, this isolates judgements of appeal from the natural use of such a website, which may in and of itself influence judgements of appeal. Prior work has not identified whether visual appeal varies depending on user goals, familiarity, and off-screen environment. It would be interesting future work to investigate how judgements of visual appeal vary across different contexts, with relevant impacts for designers.

Design Implications

There are several implications of our work for website design. First, our results show that website complexity negatively impacts performance, even for a relatively diverse sample, which suggests that when efficiency is important, designers should strive to reduce the visual complexity of their website to a minimum. Since perceived visual complexity increases with the number of images, text, and colors [27], reducing such elements to the most necessary components would improve search efficiency and information recall. This is also consistent with past usability guidelines [19] and findings on the negative impact of website complexity on search efficiency [36]. However, it is not always desirable to lower the complexity of otherwise complex websites. There are specific demographic groups that prefer highly complex websites over simple ones [25]. Since visual preferences have been shown to correlate with trustworthiness and perceived usability [14], re-designing complex websites to be simpler may be disadvantageous for these user groups.

Given this, there is an obvious tradeoff to consider: design websites that are most similar to users' visual preferences, which has been shown to correlate with trustworthiness, perceived usability, and intentions to use a website [14], or design websites that maximize search efficiency as part of usability of the site. To help in deciding between such a tradeoff, we propose the creation of an adaptive user interface which could potentially (1) infer user preferences and (2) determine whether perceived usability or efficiency of use is more important to users, and in what contexts. To infer user preferences, one could use available demographic information and assume preferences that are most common to those demographics, e.g., by referring to the results of Reinecke et al. [25]. Alternatively, a similar approach as Hauser et al. [9] could rely on inferring cognitive style from clickstream data, and adjusting complexity accordingly. Once user preferences are determined, websites could A/B test whether people prefer websites that they find appealing (with better perceived usability [14]) or low complexity websites (with more efficiency of use).

Most importantly, our results suggest that a user's visual complexity preferences need to be taken into account when assessing search time and information recall. This is important for researchers evaluating performance on user interfaces, as visual preferences may inadvertently have a differential effect on users' performance. It is also important for designers of online educational testing software and marketers conducting A/B tests with an interest in time spent. Our results suggest that whenever time is essential, researchers and designers should assess visual preferences or choose a medium complexity website layout, as this is where all people converge on time-based performance.

LIMITATIONS AND FUTURE WORK

This study specifically focused on completing a task as quickly as possible, which limits generalization about the results to speed-based tasks on the web. However, as people generally want to maximize their productivity, there may be cautious evidence towards this behavior more generally. It would be interesting to see if similar results are replicated across a variety of other situations that imitate behavior on the web (such as in creative, open-ended tasks). Studying response time under a variety of circumstances might shed light on why there are such varied reports of aesthetics' impact on usability.

Our study design was unable to answer what it is about these preference groups that make them interact with websites differently. It is possible that people's preference for low or high complexity websites determines their search strategies. Prior work has shown a relationship between personality traits and different search strategies [22]. If people with varying preferences also similarly vary in personality traits, this could affect their performance when working with websites that do not conform to their preferences. Future studies could combine tasks like ours with questionnaires of cognitive style or cultural predisposition to find out what drives these differences. Eye tracking studies can further shed light on this by evaluating if preferences correspond with eye movement or prolonged attention on different components of websites.

In our study, we aimed to use ecologically valid website stimuli rather than manually designed websites, in which we could have systematically controlled for other design variables while increasing complexity. However, given that colors are known to draw attention (and can lead to involuntary eye saccades [11]) and website elements of different shapes could make targets more or less easy to find [11]), our results may have been influenced by differences in specific website elements, rather than differences in the website complexity. To the extent possible, we tried to balance the use of colors across website complexity levels in order to rule out this possibility. However, a follow-up study could disentangle these potentially confounding factors.

CONCLUSION

This paper presents the first study to evaluate the effect of preference for visual complexity levels on performance on websites. We hypothesized that participants would perform best on search efficiency and information recall with low complexity websites, and that preferred level of complexity would impact how participants performed across websites of different complexity. Our results show that all participants performed most efficiently with low complexity websites, and the least efficiently with high complexity websites, independent of their visual preferences. We found that the negative effect of website complexity on search efficiency was amplified by preference for simple websites. Additionally, we found that people who prefer high complexity websites accurately recall more information than those who prefer low complexity at medium complexity levels. We encourage that researchers and designers measuring performance take into account how diverse visual preferences for website complexity lead to differentiated performance.

ACKNOWLEDGEMENTS
We thank the undergraduate researchers, Dean Barlan and Blue Jo for their work in developing the initial study used in this research. We also thank the LabintheWild participants and the anonymous reviewers for their valuable comments. This work was supported by NSF award 1651487.

DATASET AND MATERIALS
We make available our website stimuli, the dataset, and R-code used for analysis with the supplemental materials.

REFERENCES
1. Daniel E Berlyne. 1974. *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation.* Hemisphere.
2. George David Birkhoff. 1933. *Aesthetic Measure*. Cambridge, Mass.
3. Michelle A Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2306–2315.
4. Gordon C Bruner and Anand Kumar. 2000. Web commercials and advertising hierarchy-of-effects. *Journal of Advertising Research* 40, 1-2 (2000), 35–42.
5. Tam K Dao, Donghyuck Lee, and Huang L Chang. 2007. Acculturation level, perceived English fluency, perceived social support level, and depression among Taiwanese international students. *College Student Journal* 41, 2 (2007).
6. Maria Douneva, Rafael Jaron, and Meinald T. Thielsch. 2016. Effects of Different Website Designs on First Impressions, Aesthetic Judgements and Memory Performance after Short Presentation. *Interacting with Computers* 28, 4 (2016), 552–567.
7. Gary L Geissler, George M Zinkhan, and Richard T Watson. 2006. The influence of home page complexity on consumer attention, attitudes, and purchase intent. *Journal of Advertising* 35, 2 (2006), 69–80.
8. Jan Hartmann, Alistair Sutcliffe, and Antonella De Angeli. 2007. Investigating attractiveness in web user interfaces. In *Proceedings of the 2007 SIGCHI conference on Human Factors in Computing Systems (CHI '07)*. ACM, 387–396.
9. John R Hauser, Glen L Urban, Guilherme Liberali, and Michael Braun. 2009. Website morphing. *Marketing Science* 28, 2 (2009), 202–223.
10. Angela V Hausman and Jeffrey Sam Siekpe. 2009. The effect of web interface features on consumer online purchase intentions. *Journal of Business Research* 62, 1 (2009), 5–13.
11. John M Henderson, Myriam Chanceaux, and Tim J Smith. 2009. The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision* 9, 1 (2009), 32–32.
12. Simeon Keates and Shari Trewin. 2005. Effect of age and Parkinson's disease on cursor positioning using a mouse. In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility (Assets '05)*. 68–75.
13. Gitte Lindgaard. 2007. Aesthetics, visual appeal, usability and user satisfaction: what do the user's eyes tell the user's brain? *Australian Journal of Emerging Technologies & Society* 5, 1 (2007).
14. Gitte Lindgaard, Cathy Dudek, Devjani Sen, Livia Sumegi, and Patrick Noonan. 2011. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 1 (2011), 1–10.
15. Gitte Lindgaard, Gary Fernandes, Cathy Dudek, and Judith M. Brown. 2006. Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology* 25, 2 (2006), 115–126.
16. AH Maslow. 1937. The influence of familiarization on preference. *Journal of Experimental Psychology* 21, 2 (1937), 162.
17. Eleni Michailidou, Simon Harper, and Sean Bechhofer. 2008. Visual complexity and aesthetic perception of web pages. In *Proceedings of the 26th Annual ACM International Conference on Design of Communication (SIGDOC '08)*. 215–224.
18. Morten Moshagen, Jochen Musch, and Anja S Göritz. 2009. A blessing, not a curse: Experimental evidence for beneficial effects of visual aesthetics on performance. *Ergonomics* 52, 10 (2009), 1311–1320.
19. Jakob Nielsen. 1995. 10 usability heuristics for user interface design. *Nielsen Norman Group* 1, 1 (1995).
20. Richard E Nisbett, Kaiping Peng, Incheol Choi, and Ara Norenzayan. 2001. Culture and systems of thought: holistic versus analytic cognition. *Psychological Review* 108, 2 (2001), 291.
21. Manuel Nordhoff, Tal August, Nigini Abilio Oliveira, and Katharina Reinecke. 2018. A case for design localization: Diversity of website aesthetics in 44 countries. In *Proceedings of the 2018 SIGCHI Conference on Human Factors in Computing Systems (CHI '18)*. 337.
22. Alvitta Ottley, Huahai Yang, and Remco Chang. 2015. Personality As a Predictor of User Strategy: How Locus of Control Affects Search Strategies on Tree Visualizations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3251–3254. DOI: http://dx.doi.org/10.1145/2702123.2702590

23. Muzeyyen Pandir and John Knight. 2006. Homepage aesthetics: The search for preference factors and the challenges of subjectivity. *Interacting with Computers* 18, 6 (2006), 1351–1370.

24. Rolf Reber, Norbert Schwarz, and Piotr Winkielman. 2004. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review* 8, 4 (2004), 364–382.

25. Katharina Reinecke and Krzysztof Z. Gajos. 2014. Quantifying visual preferences around the world. In *Proceedings of the 2014 SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. 11–20.

26. Katharina Reinecke and Krzysztof Z Gajos. 2015. LabintheWild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1364–1378.

27. Katharina Reinecke, Tom Yeh, Luke Miratrix, Rahmatri Mardiko, Yuechen Zhao, Jenny Liu, and Krzysztof Z. Gajos. 2013. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the 2013 SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*.

28. Irene Reppa and Siné McDougall. 2015. When the going gets tough the beautiful get going: aesthetic appeal facilitates task performance. *Psychonomic Bulletin & Review* 22, 5 (2015), 1243–1254.

29. Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. 2007. Measuring visual clutter. *Journal of Vision* 7, 2 (2007), 17–17.

30. Carolyn Salimun, Helen C Purchase, David R Simmons, and Stephen Brewster. 2010. The effect of aesthetically pleasing composition on visual search performance. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction (NORDICHI '10)*. 422–431.

31. Juergen Sauer and Andreas Sonderegger. 2011. The influence of product aesthetics and user state in usability testing. *Behaviour & Information Technology* 30, 6 (2011), 787–796.

32. Andreas Sonderegger and Juergen Sauer. 2010. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics* 41, 3 (2010), 403–410.

33. Julie S Stevenson, Gordon C Bruner, and Anand Kumar. 2000. Webpage background and viewer attitudes. *Journal of Advertising Research* 40, 1-2 (2000), 29–34.

34. Manfred Thüring and Sascha Mahlke. 2007. Usability, aesthetics and emotions in human–technology interaction. *International Journal of Psychology* 42, 4 (2007), 253–264.

35. Alexandre Tuch, S Roth, Kasper HornbÃęk, Klaus Opwis, and Javier Bargas-Avila. 2012b. Toward understanding the relation between usability, aesthetics, and affect in HCI. *Computers in Human Behavior* [manuscript under revision] (04 2012).

36. Alexandre N Tuch, Javier A Bargas-Avila, Klaus Opwis, and Frank H Wilhelm. 2009. Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-Computer Studies* 67, 9 (2009), 703–715.

37. Alexandre N Tuch, Eva E Presslaber, Markus StöCklin, Klaus Opwis, and Javier A Bargas-Avila. 2012a. The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. *International Journal of Human-Computer Studies* 70, 11 (2012), 794–811.

38. Neff Walker, David A. Philbin, and Arthur D. Fisk. 1997. Age-related differences in movement control: Adjusting submovement structure to optimize performance. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* 52, 1 (1997).

39. Huaitang Wang, Takahiko Masuda, Kenichi Ito, and Marghalara Rashid. 2012. How much information? East Asian and North American cultural products and information search performance. *Personality and Social Psychology Bulletin* 38, 12 (2012), 1539–1551.

40. Christine Yeh and Mayuko Inose. 2003. International students' reported English fluency, social support satisfaction, and social connectedness as predictors of acculturative stress. *Counselling Psychology Quarterly* 16 (03 2003), 15–28.

41. Xianjun Sam Zheng, Ishani Chakraborty, James Jeng-Weei Lin, and Robert Rauschenberger. 2009. Correlating low-level image statistics with users-rapid aesthetic and affective judgments of web pages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, 1–10.