Relationship Between Visual Complexity and Aesthetics of Webpages

Aliaksei Miniukovich
University of Trento
Trento, Italy
aliaksei.miniukovich@unitn.it

Maurizio Marchese
University of Trento
Trento, Italy
maurizio.marchese@unitn.it

ABSTRACT
Substantial HCI research investigated the relationship between webpage complexity and aesthetics, but without a definitive conclusion. Some research showed an inverse linear correlation, some other showed an inverted u-shaped curve, while the rest showed no relationship at all. Such a lack of clarity complicates hypothesis formulation and result interpretation for future research, and lowers the reliability and generalizability of potential advice for Web design practice. We re-collected complexity and aesthetics ratings for five datasets previously used in webpage aesthetics and complexity research. The results were mixed, but suggested an inverse linear relationship with a weaker u-shaped sub-component. A subsequent visual inspection of revealed several confounding factors that may have led to the mixed results, including some webpages looking broken or archaic. The second data collection showed that accounting for these factors generally eliminates the u-shaped tendency of the complexity-aesthetics relationship, at least, for a relatively homogeneous sample of English-speaking participants.

Author Keywords
Visual Aesthetics; Graphical User Interfaces; Web Design; User Study; Quantitative Analyses.

CSS Concepts
• **Human-centered computing~HCI theory, concepts and models** • *Human-centered computing~Empirical studies in HCI*

INTRODUCTION
The shape of aesthetics-complexity relationship determines answers to multiple questions related to Web design. For design practitioners, the questions may be whether minimalism is a good trend in web design or whether it makes webpages dull and un-engaging, what the optimal webpage complexity is, and whether such an optimum even exists. For researchers, the questions would relate to hypotheses formulation and result interpretation, e.g., whether an observation of u-shaped aesthetics-complexity relationship is expected or the study contains critical flaws and needs to be re-designed and re-run.

This paper explores the aesthetics-complexity relationship. A crowdsourcing-based study recruited a sample of English-speaking participants (primarily, from the US) and re-collected the scores of complexity and aesthetics for several datasets that past work relied on to observe a negative correlation, inverted u-shaped curve, or no relationship between complexity and aesthetics. The study at first largely replicated the results of past work, but also identified several potential confounders, which could have undermined the conclusions of past studies, presenting a serious problem, e.g., similar to [9] that showed an example of confounders undermining multiple HCI studies on Fitts's law. Accounting for one confounder – technical condition of a webpage – led us to conclude that the relationship between webpage aesthetics and complexity is likely linear, at least for the 1506 webpages that the study included, though this result may need to be replicated across different cultures (cf., [51]).

RELATED WORK
HCI research has actively studied visual aesthetics – a stimulus property that results in an immediate pleasant feeling of appreciation towards the appearance of the stimulus (cf., [44,29]) – and its effects on user experience (UX). Early research [16,43] demonstrated aesthetics to be highly relevant for everyday things, such as webpages and other GUIs, and not only for pieces of art, as might have been expected. Subsequent research positioned aesthetics relative to other UX constructs (such as, usability [45], overall system goodness [11] and trustworthiness [18]), extracted aesthetics sub-dimensions and developed multi-item measurement scales [17,30,28], linked aesthetics to design features (e.g., a text-to-image ratio [7] and predominant color [5]), and developed computational approaches to GUI aesthetics modeling and prediction [35,23,50,8]. However, despite these significant developments, our ability to state what exactly makes a webpage nice remains stymied by the lack of theoretical clarity on the mental processes underlying aesthetics impression, as alternative theories attribute aesthetics to different causes.

CHI 2020, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04...$15.00.
DOI: https://doi.org/10.1145/3313831.3376602

Psychological Theories of Aesthetics

Berlyne's [2] influential theory posited that two neurological systems determined aesthetics impression, one related to reward and the other to aversion. Both systems are activated by a class of stimulus properties, such as novelty, complexity and familiarity, but drive aesthetics impression in the opposite directions. The effect of both systems would add up to zero if their activation were the same, but the reward system activates earlier, quickly approaches its maximum effect and stays approximately the same with more stimulation, whereas the aversion system activates later and keeps on producing larger effects with more stimulation. Such activation patterns add up to an inverted U-shape relationship between the levels of stimulation by stimulus properties and aesthetics.

A large part of experimental evidence did not support the existence of an inverted U-shaped curve, and instead posited that stimuli properties, such as complexity, affected aesthetics linearly [20]. Two theories could explain such a linear relationship. The prototype theory of aesthetics [48,47] state that people prefer prototypical stimuli – stimuli that represent their category the best, e.g., most people choose *sparrow* over *ostrich* as a better representation of category *birds* – which leads to higher aesthetics with higher prototypicality. The processing fluency theory of aesthetics [34,33] encompasses prototypicality [49] as only one of several factors decreasing processing fluency – the effort to mentally process a stimulus – which in turn leads to higher aesthetics. The other factors include user familiarity with a stimulus and different aspects of stimulus complexity, such as the amount of information and symmetry.

The processing fluency theory was criticized as only accounting for the mild positive feeling of understanding a stimulus – ostensibly, a small part of overall aesthetics impression – while leaving out interest and curiosity that account for the rest of the impression [1,40]. The appraisal theory of aesthetics [39,40] attempts to correct this shortcoming. It asserts that stimulus properties do not directly result in an aesthetics impression, but are first appraised (e.g., as congruent with one's goals or self worth, understandable, or having potential for knowledge expansion), and the appraisals lead to an emotion related to stimulus aesthetics, which can be interest and curiosity, but also negative emotions, such as disgust or confusion. The appraisals depend strongly on inter-individual differences, including the level of expertise and familiarity with stimuli [41], which complicates hypothesizing the direction or shape of the effect of stimulus appearance on aesthetics impression, with possibly opposite directions and different shapes for different demographics. However, recent research found no effect of design expertise on webpage aesthetics [3], which potentially limits the generalizability of the appraisal theory to artistic stimuli, not everyday stimuli, such as GUIs.

Webpage Aesthetics and Visual Complexity

Attempts to directly link webpage aesthetics with certain features of design [13,12] often led to hard-to-interpret results with limited generalizability, with few exceptions, such as the general preference for more images and less text [7,27]. The difficulty in establishing such feature-aesthetics links likely stemmed from the large variety of design features that could be tweaked to produce numerous layout and element configurations. Only a tiny part of these configurations would be studied in any single study, making conclusions highly susceptible to spurious observations (e.g., better webpages using tiny fonts [13], despite larger fonts being far better for Web readability [36]) that can only be corrected by meticulous manual review (e.g., smaller fonts corresponded to the copyright notices that most good webpages featured [13]).

Instead of directly linking design features to webpage aesthetics, much of HCI research focused on an intermediary that would correlate with aesthetics in a predictable way and was easier to express in design features than aesthetics. Possible intermediaries included design colorfulness [35], diversity [25], prototypicality [46], and most often, visual complexity [46,42,32,21,35,26]. A recent analysis of visual complexity [22] suggested it encompassed several facets, including the amount and diversity of information, orderly organization of information, and sensory perceivability of detail. The analysis also listed several computational methods to describe these facets for GUIs, which could further facilitate the adoption of visual complexity as a proxy of GUI aesthetical impression.

Conflicting findings on the complexity-aesthetics relationship, however, hamper the actual adoption of visual complexity in aesthetics measurement. Reinecke et al. [35] observed an inverted U-shaped relationship, similar to the predictions of Berlyne's theory, but with strong linear tendencies: simpler webpages were only slightly less aesthetical than mid-complexity webpages, and complex webpages much less aesthetical than both simple and mid-complexity webpages. Güçlütürk et al. [10] found a similar U-shaped complexity-aesthetics relationship, and attributed it to the existence of two groups of users. One group liked lower complexity, whereas the other liked higher complexity, which would be compatible with the appraisal theory of aesthetics, emphasizing inter-personal differences. Tuch et al. [46] observed a linear relationship – negative correlation – between webpage aesthetics and complexity, and Miniukovich et al. [24,26] observed the same relationship for both webpages and mobile Apps, while also referencing the fluency theory. Finally, Boychuk & Bakaev [4] found no connection between complexity and aesthetics, but did observe a correlation between aesthetics and order – a concept closely related to visual complexity.

STUDY

We conducted a study to explore why past work could not agree on the shape and direction of the complexity-aesthetics relationship and whether these disagreements could be reconciled to arrive to a better-informed conclusion on the relationship. The study included two steps. The first step aimed at replicating past results, whereas the second step aimed at testing additional explanations for divergence in the past results. Both steps used the same datasets of webpages.

Datasets

Five previously-used in aesthetics and complexity research and one new datasets were combined to create two larger groups of datasets of webpage screenshots (Table 1): general-purpose (N = 749) and university websites (N = 757). All webpages were website homepages in English (a small number of non-English webpages were in the Uni dataset). We chose the past datasets to represent different perspectives on the aesthetics-complexity link: CHI_13 supported a U-shaped curve; IJHCS_12 and AVI_14 supported a negative correlation; ICWE_19 suggested no relationship. CHI_15 was not aimed at testing complexity-aesthetics link, but we included it to counterbalance the potentially outdated-looking webpages of IJHCS_12. The new CHI_20 was collected to counterbalance the prevalence of webpages of universities from rich, anglophone countries (408 out of 497, mostly the US) in ICWE_19, which resembled each other, were very well designed, and were relatively complex (e.g., the simplest webpage, figure 2 in [4], might still be considered complex). Such properties of ICWE_19 might have been a source of biases, e.g., if the range of complexity or aesthetics was narrow or if no genuinely simple webpages were sampled.

Group	Dataset	Size	Description
GEN	IJHCS_12 [46]	184	We obtained a part of the dataset [46] containing 119 corporate and 65 arts-rated webpages, with complexity ratings only available for the corporate webpages and aesthetics ratings for arts-related webpages. Complexity ratings were collected online; rating item "*I think this website is of high visual complexity*"; exposure duration not mentioned. Aesthetics ratings for the arts-related dataset were not collected as described in [46], but as a part of a pre-study, using classical and expressive aesthetics items [17]. Participants: not clear.
	CHI_13 [35]	350	We only used the full-color English webpages, leaving out non-English, monochrome and Webby-award webpages to minimize potential biases. Both complexity and aesthetics ratings were available for all webpages, collected online using two 9-point Likert-type scales (*not at all complex/very complex*); exposure duration .5 sec. Participants: general public, online users.
	AVI_14 [24]	140	Of the 140 webpages, 115 came from online repositories of beautifully designed examples, while the other 25 were chosen manually by the authors for being unappealing. Both complexity and aesthetics ratings were available, measured in-lab individually using two 1-5 semantic differential scales with anchors *simple/complex* and *ugly/beautiful*; exposure duration .05 sec. Participants: students and researchers.
	CHI_15 [23]	75	Of 300 webpages, we chose 75 homepages (the others were non-homepage webpages from the same websites). The websites were collected by crowdworkers (to counter experimenter biases) and came from a corporate, eCommerce, or news domain. Only aesthetics ratings were available, collected in-lab individually using a 1-7 semantic differential item (anchors *ugly/beautiful*), exposure duration .15 sec. Participants: students and university employees.
EDU	ICWE_19 [43]	497	The university webpages were manually chosen from a much larger sample. The authors avoided very famous university and strived to ensure a diversity of design layouts in their sample. Both complexity and aesthetics ratings were collected using 7-point Likert-type scales; exposure duration not mentioned. Participants: primarily students, IT specialists.
	CHI_20	260	We sampled university website URLs from an online repository[1] and excluded URLs with hostnames ending on *.edu*, *ac.uk*, *ac.nz*, *edu.au*, *.ca*, and *.ie*. The left URLs were manually visited to identify English-version homepages, if present. The homepages were automatically collected as full-page screenshots using a script, saved as PNG images; min height 1000px, width 1440px. The screenshots were then cropped to the top 1600 pixels and downscaled to half their original size.

Table 1. Six smaller datasets were combined in two groups (General-Purpose and University websites) to be used in the study[2]; five datasets were used in past research. Size shows the number of webpages in each dataset.

[1] https://univ.cc/world.php

[2] Full dataset available on https://doi.org/10.7910/DVN/XEYNYW

The first group dataset, GEN, contained the top-screen (height 768 or 800 pixels depending on a sub-dataset) screenshots of homepages of various-genre websites (width 1000 to 1280 pixels depending on a sub-dataset). The second dataset, EDU, could not be combined with GEN because it had non-cropped and often-lengthy webpages that could not be cropped to the top-screen (upper 800 pixels) without loosing much of their content and no longer looking like realistic, meaningful webpages (e.g., a screenshot would only contain a top horizontal menu and large rotating banner). Instead, the screenshots of EDU were cropped to the upper 1600 pixels and downscaled to half their size (Lanczos-3 kernel; max height 800px, width 720px), so they could be viewed on common-size monitors without scrolling. Because of this cropping and scaling related difference, user data were collected and analyzed separately for EDU and GEN (except demographics analyses).

Step 1: Complexity and Aesthetics

Step 1 aimed at replicating the results of past studies, by re-collecting complexity and aesthetics scores for the previously used datasets. We used the same experimental procedure for all datasets, which should have allowed us to test if the inter-dataset differences in observed aesthetics-complexity relationship were due to the differences in experimental procedures (measurement items, exposure duration, demographics). Evaluating all datasets together (though, within their dataset group) would also let us test if the datasets differed in complexity or aesthetics: if CHI_13 contained webpages of the whole simple-complex spectrum, and AVI_14 only contained mid-complex and complex webpages, one could conclude that AVI_14 did not suggest an inverted U-shape curve, because it only had examples representing the right half of the U-shape, which, in turn, manifested itself as a negative aesthetics-complexity correlation. Finally, we could use the re-collected scores to select and review low-complexity/low-aesthetics (low-VC/low-AE) and high-complexity/high-aesthetics (high-VC/high-AE) webpages for commonalities. Such commonalities could then be tested as potential confounding factors that could have morphed a negative correlation in a U-shaped curve or resulted in no correlation at all.

Participants

We recruited 390 English-speaking participants (159 female; M age = 33.4 years, SD = 1.41 years, range 18 to 73 years; $n_{color\ blind} = 3$) on two crowdsourcing platforms (MTurk.com and Microworkers.com), who were compensated 1.2 USD for a ~12 min experimental session. Participants primarily came form the US, but also other English-speaking countries, such as the UK, Canada and Australia. Approximately a half of participants had a university degree (BA – 170; MSc – 50; PhD – 4), while the rest attended a high school (with diploma – 146; no diploma -19); one person had no schooling. A majority of participants were employed (self-employed – 67, partially employed – 37, fully employed – 199), while the rest were students (n = 37), unemployed (n=28), unable to work (3), retired (n = 4), or worked as a homemaker (n=14) or in military (n=1). Participants indicated to be using the Internet 6.18 hours a day (SD = 4.02h).

Procedure

After being redirected from a crowdsourcing platform on our website, participants read a brief study summary and participation conditions. If they consented to participate, they then filled out a brief demographic questionnaire. A summary of screenshot-rating procedure was shown, and participants proceeded to rate 117 webpages (14 webpages were rated twice; the first three webpages were training webpages, not used in later analyses) one by one using keyboard keys “1” to “7” (with “Enter” to confirm rating), which should have reduced the rating effort relative to the use of a mouse or touchpad. The webpages were selected randomly from one of the two dataset groups (either GEN or EDU, Table 1). Each of 117 trials included a sequence of a gaze-fixation point shown for 1-1.5second, a webpage shown for 1 second, a black-white noise mask flashed for 50msec, and 1-7 semantic differential scale shown until a rating was confirmed. A participant rated either complexity (anchors *simple/complex*) or aesthetics (anchors *ugly/beautiful*), which reduced rating effort (no need to mentally switch from one quality to the other) and ensured complexity and aesthetics are not artificially correlated (due to the two qualities being shown together). Optional free-form feedback could be left at the end of the study.

Results

We used the twice-rated webpages for data quality control – if a participant systematically rated webpages inconsistently (e.g., as 1 the first time, but 7 the second time) we deemed their data to be untrustworthy and excluded them from further analyses. Pearson’s and interclass correlations were computed on the ratings of the twice-rated webpages to aid the data quality review. In total, the data of 113 participants were excluded from analyses for a variety of reasons besides inconsistency, including giving all webpages the same rating (e.g., all 114 ratings were ‘1’), taking unreasonably long to set a rating (e.g., regularly interrupting the test for > 30sec, which was problematic because a webpage was only shown for 1sec), or giving ratings at a very fast rate (e.g., with the median time to set a rating as low as .13sec). The relatively high exclusion rate is not surprising for crowdsourcing and could be as high as over a half of all data [15]. We then reviewed correlations between individual and averaged (across participants) scores, which were expected to be high for both complexity and aesthetics [46,24,26,23]. The mean individual-average correlations were relatively high for both aesthetics (mean r = .62) and complexity (mean r = .68). The scores of a small group of participants did not correlate with the average, and their data were excluded from further analyses (we used r < .3 as a cut-off; 14 participants’ complexity data and one participant’s aesthetics data) to further minimize the proportion of potentially untrustworthy data in the dataset. Individual complexity and aesthetics scores were aggregated across participants for subsequent analyses.

Correlations between the scores collected in the past studies and newly re-collected scores were relatively high (Table 2), except for ICWE_19, which suggested that our data-collection approach resembled the past approaches, with no major issues and deviations from the past studies in complexity and aesthetics measurement.

Dataset	Aesthetics, r (df)	Complexity, r (df)
IJHCS_12*	.68 (63)	.82 (117)
CHI_13	.80 (338)	.83 (346)
AVI_14	.75 (138)	.64 (138)
CHI_15	.69 (73)	NA
ICWE_19	.55 (488)	.27 (488)

* aesthetics correlation for arts-related webpages only (using the mean of classical and expressive aesthetics scores); complexity for corporate webpages

Table 2. Correlations between past and newly re-collected scores of the past datasets; all p < .001.

We then tested if the datasets differed in the ranges of complexity and aesthetics that they contained (Figure 2 and Figure 2). GEN and EDU datasets were tested separately because they were collected separately, which would have affected participants' use of measurement scale, and thus, group means. For GEN datasets, one-way ANOVAs with dataset ids as an independent factor were significant for both aesthetics ($F(3,745)=60.58$, $p < .001$) and complexity ($F(3,745) = 31.00$, $p < .001$). A post hoc Tukey test showed only AVI_14 to be significantly simpler than the other three GEN datasets (all $p_{adj} < .001$, Figure 1). Another Tukey test showed both AVI_14 and CHI_15 to be significantly nicer than IJHCS_12 and CHI_13 (all $p_{adj} < .001$); AVI_14 did not significantly differ from CHI_15, nor did IJHCS_12 differ from CHI_13 (Figure 2). For EDU datasets, t-tests[3] showed a significant difference in aesthetics of ICWE_19 and CHI_20 datasets ($t(491.12)=8.83$, $p < .001$), but no difference in complexity ($t(442.74)=-.47$, $p=.64$).

To test the existence, shape and strength of the relationship between complexity (both linear and quadratic terms) and aesthetics, two linear models with dataset ID as a moderator were fitted, Table 3. The models suggested that both negative linear and quadratic (inverted U-shape) components of aesthetics-complexity relationship were present, though the strength of the components and their statistical significance varied depending on the dataset. To further examine per-dataset aesthetics-complexity relationship, we fitted a series of regression models, Table 4. The relationship was present for all datasets, except CHI_20, though it was weaker than past research suggested. IJHCS_12 and CHI_15 showed a linear relationship, as in the respective past studies. CHI_13, AVI_14 and ICWE_19 showed a linear relationship with a U-shaped component, despite past-research ICWE_19 suggested no relationship, which past-research AVI_14 suggested a linear negative correlation. Figure 3 visualizes the relationship for several datasets.

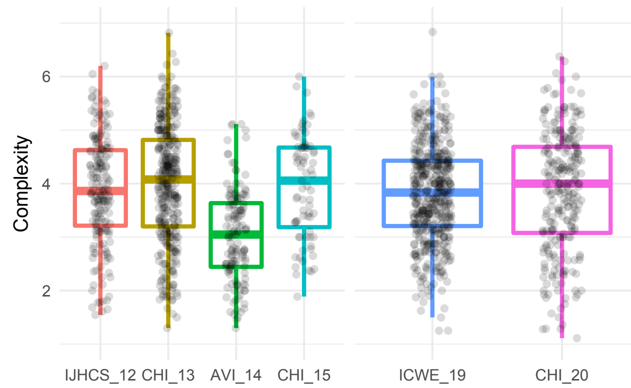

Figure 1. Boxplots show relative differences in complexity among datasets; overlaid dot plots show complexity (with random jitter) of individual webpages within datasets. GEN datasets were analyzed separately from EDU datasets because they were collected separately.

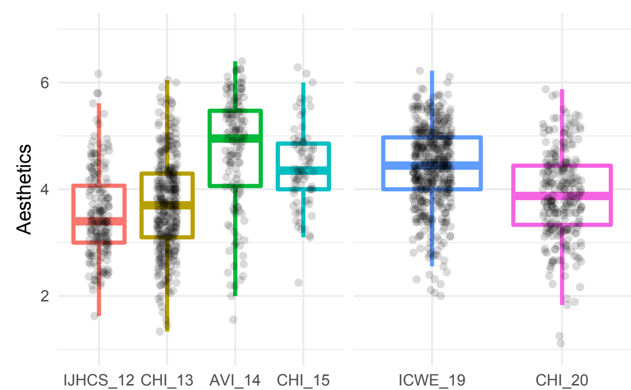

Figure 2. Boxplots show relative differences in aesthetics among datasets.

We further looked for factors that may have influenced the shape of the curve in the present and past research – either bending down the left side a linear relationship making it slightly U-shaped, or also bending the right side make the relationship non existent. We automatically selected for a visual inspection low-VC/low-AE and high-VC/high-AE webpages (Figure 3, bottom row plots, bottom-left and top-right corners delimited with orange dotted lines). The GEN dataset returned 64 low-VC/low-AE and only three high-VC/high-AE webpages, whereas EDU returned 30 low-VC/low-AE and 20 high-VC/high-AE webpages. A review of both GEN and EDU low-VC/low-AE webpages revealed three noticeable commonalities among them: many webpages looked outdated, designed by non-professional designers, or partially broken with possibly missing content, Figure 4. A review of EDU high-VC/high-AE did

[3] a t-test is equivalent to a one-way ANOVA with a two-level factor.

not reveal noticeable common factors that could have influenced aesthetics systematically, possibly except for a prevalence of large high-quality photographs.

			β (Std.Err.)	$β_{act}$
GEN	Intercept	Baseline (IJHCS_12)	-.36***(.06)	-.36
		CHI_13	.17*(.08)	-.19
		AVI_14	.68***(.13)	.31
		CHI_15	.88***(.12)	.52
	VC	Baseline (IJHCS_12)	-.16*(.07)	-.16
		CHI_13	.09(.08)	-.07
		AVI_14	-.57***(.16)	-.73
		CHI_15	-.19(.13)	-.35
	VC^2	Baseline (IJHCS_12)	.06(.07)	.06
		CHI_13	-.21*(.08)	-.15
		AVI_14	-.55***(.13)	-.49
		CHI_15	-.19(.16)	-.12
				Adj. R2=.25
EDU	Itcpt	Baseline (ICWE_19)	.21**(.04)	.21
		CHI_20	-.62**(.07)	-.41
	VC	Baseline (ICWE_19)	-.20**(.05)	-.20
		CHI_20	.17*(.07)	-.03
	VC^2	Baseline (ICWE_19)	-.13*(.05)	-.13
		CHI_20	.07(.07)	-.07
				Adj. R^2=.13

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 3. Two regression models (separately for GEN and EDU), with aesthetics as output, complexity (VC) and complexity 2nd order polynomial (VC^2) as predictors, and dataset ID as a moderator; β coefficients (incl. their significance) show intercept and slope changes relative to a baseline, and the actual intercepts and slopes (descriptive of aesthetics-complexity link per dataset) are estimated as $β_{act}$

Dataset	VC, β	VC^2, β	Std.Err.	R^2
IJHCS_12	-.18*	.07	.07	.04
CHI_13	-.11*	-.17**	.05	.04
AVI_14	-.18*	-.31***	.08	.13
CHI_15	-.45***	-.12	.01	.22
ICWE_19	-.21***	-.13**	.04	.06
CHI_20	-.03	-.08	.06	.01
GEN	-.26***	-.09*	.04	.08
EDU	-.14***	-.16***	.04	.04

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 4. A series of regression models, with aesthetics as a dependent variable, and complexity and complexity 2nd order polynomial as predictors; a model was fitted for each dataset, which simplifies interpreting β coefficients as a linear (VC) or quadratic (VC^2) component of aesthetics-complexity relation. Std.Err. is the same for both coefficients.

Step 2: Confounding Factors

To further disambiguate the complexity-aesthetics relationship, Step 2 tested the three factors – novelty, craftsmanship, and technical condition – that Step 1 established as possible confounders (i.e., variables affecting both complexity and aesthetics). This required evaluating the three factors for all webpages, which was done in a new data collection with users. In addition, Step 2 reviewed a past observation of the U-shaped pattern being caused by combining the scores of two distinct groups of users, one valuing simplicity (negative aesthetics-complexity correlation) and the other complexity (positive aesthetics-complexity correlation) [10].

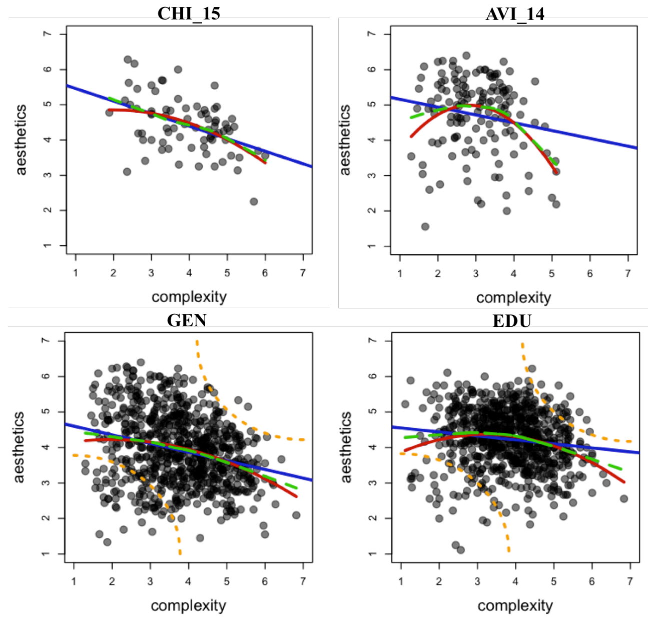

Figure 3. Relationship between complexity and aesthetics. Blue solid lines are regression lines; red curves are regression lines when complexity^2 is added in models; green dashed curves are Lowess curves. Orange dotted curves show plot areas from which webpages were selected for a visual inspection.

Participants

A total of 530 English-speaking crowdworkers (193 female; M age = 35.37 years, SD = 1.55, range 18-70 years; $n_{color\ blind}$ = 5) were recruited for the data collection. Over half of participants had a university degree (BA – 269, MSc – 66, PhD – 10), while the rest attended a high school (with diploma – 163; no diploma – 20); two participants did not attend a school. A majority of participants were employed (fully – 348; partially -39; self-employed – 63), with the rest being students (n=27), unemployed (n=24), retired (n=8) or unable to work (n=5), or worked as a homemaker (n=14) or served in the military (n=2). Participants said to be using the Internet 5.88 hours a day (SD = 3.81).

Procedure

Procedure was identical to the procedure of Step 1, except the duration of webpage exposure was increased to 1.5sec to accommodate for a possibly more demanding judgment of new webpage properties. The three factors were measured using 1-7 semantic differential scales, with anchors *novel/outdated* for novelty, design *amateur/professional* for craftsmanship, and *functional/broken* for technical condition.

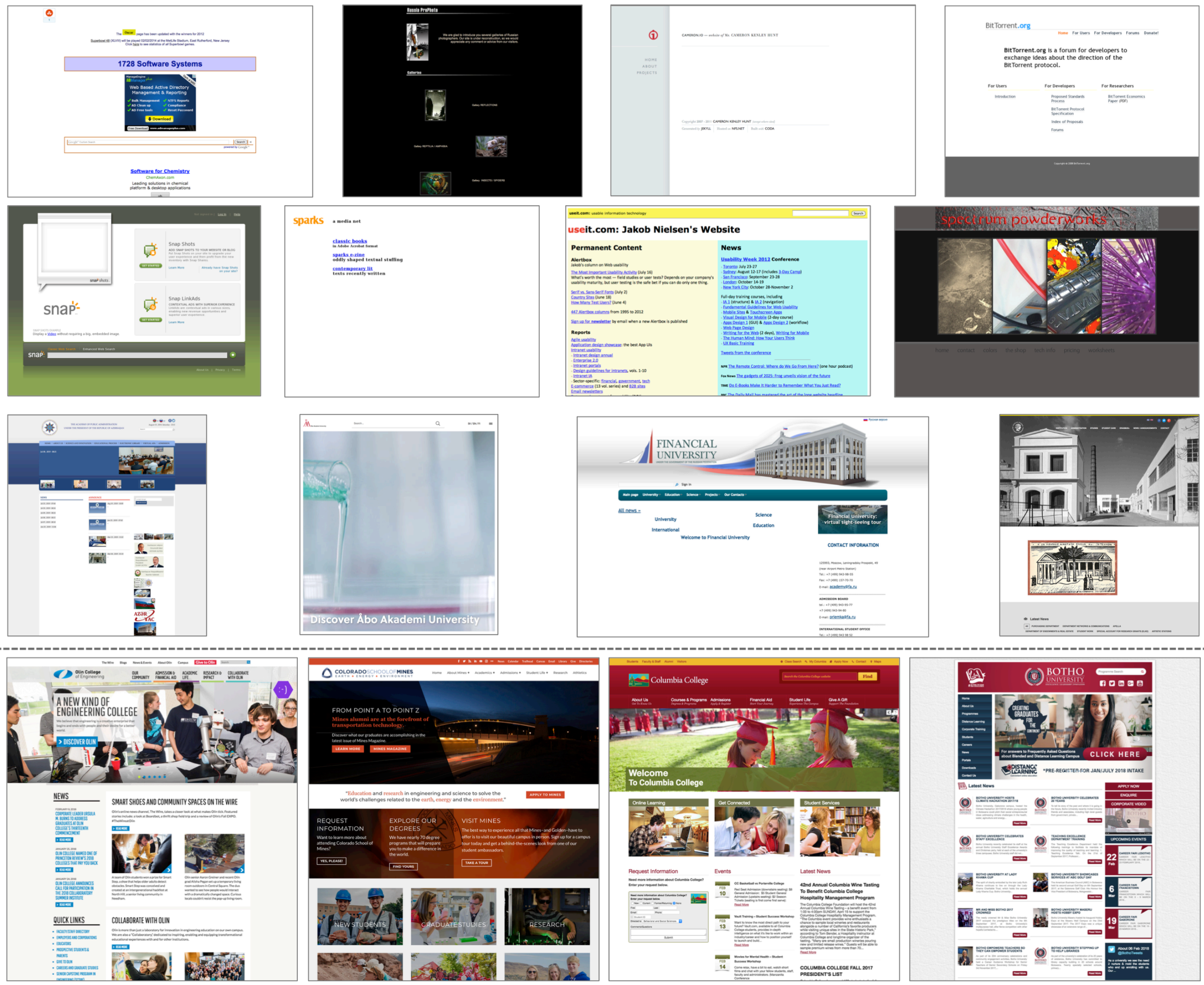

Figure 4. Selection of low-VC/low-AE (above the line) and high-VC/high-AE (below the line) webpages.

Results

Following the same principles of data quality control as in Step 1, we excluded the data of 137 participants form further analyses. The mean individual-average correlations were relatively high for novelty (mean r = .66) and craftsmanship (mean r = .63), and slightly lower for technical condition (mean r = .57).

The three new factors cross-correlated, but not excessively, Table 5. A subsequent per-dataset review of correlations between the factors and complexity/aesthetics (Table 6) showed that only technical condition could be a disambiguating confounder, since it tended to correlate negatively with both aesthetics and complexity, whereas novelty correlated positively with complexity and negatively with aesthetics (i.e., novelty likely explained the same as complexity variance in aesthetics), while craftsmanship did not consistently correlate with complexity.

	Tech. Condition	Novelty
Novelty	.52	
Craftsmanship	-.63	-.71

Table 5. Cross-correlations among the three potential confounders, all df = 1504, p < .001.

To factor out the variance of technical condition from aesthetics and estimate complexity-aesthetics relationship more precisely, technical condition was used as a control variable in two models regressing aesthetics on complexity linear and quadratic terms, Table 7. The presence of technical condition appeared to have strengthened the linear term and weakened the quadratic term (see Table 3 models without technical condition). To further disambiguate the aesthetics-complexity relationship, aesthetics was regressed on technical condition, and aesthetics residual variance – which is webpage aesthetics but without the effects of technical condition – was used as a dependent variable in a series of linear regressions, one per dataset, Table 8. The results mirrored the results of joined-dataset linear models (those in Table 7). The linear component of the complexity-aesthetics relationship became stronger for all datasets (relative to Table 4), and the u-shaped component became insignificant for all but AVI_14.

	Tech. Cond.		Novelty		Craftsmanship	
Dataset	Aesth	Compl	Aesth	Compl	Aesth	Compl
IJHCS_12	$-.39^{***}$	$-.27^{***}$	$-.71^{***}$	$.25^{**}$	$.60^{***}$	.12
CHI_13	$-.59^{***}$	$-.24^{***}$	$-.76^{***}$	$.15^{**}$	$.70^{***}$	$.11^{*}$
AVI_14	$-.72^{***}$	-.10	$-.85^{***}$	$.28^{**}$	$.76^{***}$	-.03
CHI_15	$-.51^{***}$	.07	$-.71^{***}$	$.51^{***}$	$.61^{***}$	-.22
ICWE_19	$-.47^{***}$	-.06	$-.64^{***}$	$.31^{***}$	$.59^{***}$	$-.12^{**}$
CHI_20	$-.40^{***}$	$-.38^{***}$	$-.69^{***}$	$.22^{***}$	$.66^{***}$	.02
GEN	$-.61^{***}$	$-.13^{***}$	$-.81^{***}$	$.31^{***}$	$.71^{***}$	-.01
EDU	$-.50^{***}$	$-.18^{***}$	$-.69^{***}$	$.27^{***}$	$.64^{***}$	-.07

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 6. Per-dataset correlations between the three potential confounders and, complexity and aesthetics; all scores are averaged across participants per webpage.

			β (Std.Err.)	β_{act}
GEN	Intercept	Baseline (IJHCS_12)	$1.26^{***}(.22)$	1.26
		CHI_13	.45(.25)	1.71
		AVI_14	$1.03^{***}(.28)$	2.29
		CHI_15	.58(.37)	1.84
	VC	Baseline (IJHCS_12)	$-.28^{***}(.06)$	-0.28
		CHI_13	.05(.07)	-0.23
		AVI_14	$-.34^{**}(.12)$	-0.62
		CHI_15	-.07(.11)	-0.35
	VC^2	Baseline (IJHCS_12)	$.12^{*}(.05)$	0.12
		CHI_13	$-.15^{*}(.06)$	-0.04
		AVI_14	$-.40^{***}(.10)$	-0.29
		CHI_15	-.15(.12)	-0.03
	TechCnd	Baseline (IJHCS_12)	$-.47^{***}(.06)$	-0.47
		CHI_13	-.14 (.07)	-0.72
		AVI_14	$-.24^{***}(.08)$	-0.72
		CHI_15	-.07(.13)	-0.55
			Adj. R^2=.31	
EDU	Itcpt	Baseline (ICWE_19)	$1.57^{***}(.13)$	1.57
		CHI_20	$-.51^{*}(.22)$	1.06
	VC	Baseline (ICWE_19)	$-.24^{***}(.04)$	-0.24
		CHI_20	.05(.06)	-0.19
	VC^2	Baseline (ICWE_19)	-.01(.04)	-0.01
		CHI_20	.00(.06)	-0.01
	TCnd	Baseline (ICWE_19)	$-.53^{***}(.05)$	-0.53
		CHI_20	.07(.07)	-0.45
			Adj. R^2=.31	

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 7. Two regression models of aesthetics (for GEN and EDU), with complexity (VC), complexity 2nd order polynomial (VC^2) and Technical Condition (TechCnd) as predictors, and dataset ID as a moderator.

We further looked at individual preference for complexity as an additional potential explanation of the U-shaped relationship. The relationship between an individual's aesthetics and average complexity was primarily non-significant, a negative correlation, or an inverted U-shape, Figure 5. Only in few cases the relationship was a positive correlation or non-inverted u-shaped curve (eight out of 128 participants). This observation did not suggest that two distinct user groups existed, as the small number of user who valued complexity (positive linear or non-inverted U-shaped link) was consistent with such result being a random aberration or such users being at an extreme end of a spectrum rather than in a distinct group.

Dataset	Compl, β	Compl^2, β (p)	Std.Err.	R^2
IJHCS_12	$-.31^{***}$	.13 (.07)	.07	.11
CHI_13	$-.32^{***}$	-.07 (.19)	.05	.07
AVI_14	$-.36^{***}$	$-.26^{**}$	.08	.09
CHI_15	$-.49^{***}$	-.03 (.79)	.10	.18
ICWE_19	$-.27^{***}$	-.02 (.70)	.04	.06
CHI_20	$-.20^{**}$	-.03 (.66)	.06	.04
GEN	$-.43^{***}$	-.01 (.82)	.03	.12
EDU	$-.26^{***}$	-.04 (.25)	.04	.05

** $p < .01$; *** $p < .001$.

Table 8. Per-dataset regression models, with residual aesthetics (after factoring out technical-condition scores) a dependent variance and complexity and complexity^2 as independent variables; R^2 refers to the explained extra variance in aesthetics, in addition to the technical-condition scores. Std.Err. applies for both coefficients.

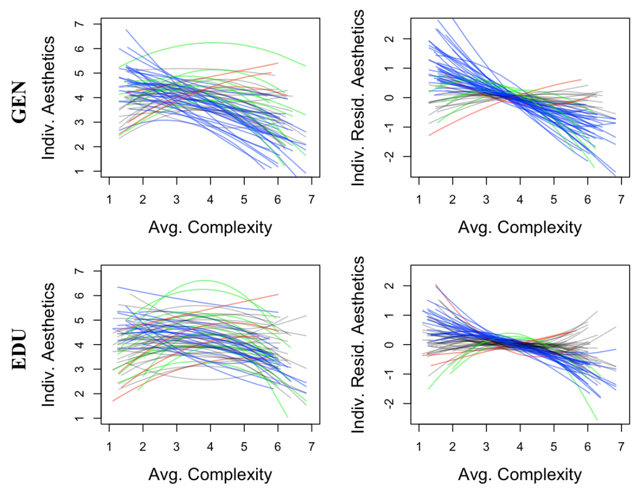

Figure 5. Regression lines show relationships between individuals' aesthetics and average complexity, before (left column) and after (right column) factoring out technical condition from individual aesthetics. Blue shows negative correlation; light gray – no relationship; green – inverted u-shapes; and red everything else.

Figure 5 shows a substantial amount of inter-personal difference in the shape of aesthetics-complexity curve, even after the effect of technical condition is factored out. We explored if the differences could be due to a systematic effect of demographic variables. Four variables – age, gender, hours of daily Web use, and education level – were entered one by one as moderators in linear mixed models, with non-aggregated aesthetics as output, aggregated complexity as fixed effects, and webpage ID and participant ID as random effects. The EDU and GEN data were combined to increase the number of observations per level of a demographic variable. The results show that none of age (Table 9), gender (Table 10), web use (Table 11) or education level (Table 12) had a direct effect on aesthetics, but all four variables had a small to moderate effect on the strength of complexity-aesthetics relationship.

	b coeff (Std.Err)	t val
Intercept	3.98(.20)	19.75
Age	.00(.01)	.44
VC	-13.99(1.54)	-9.06
VC*Age	.19(.04)	5.13
VC^2	-2.40(1.55)	-1.55
VC^2*Age	-.06(.04)	-1.61

Table 9. A linear mixed model shows the effect of each additional year of user age on complexity-aesthetics link.

	b coeff (Std.Err)	t val
Intercept (baseline)	4.11(.08)	52.59
Gender:Female	-.10(.12)	-0.84
VC	-5.94(.92)	-6.48
VC:Fem	-4.27(.84)	-5.07
VC^2	-5.45(.92)	-5.95
VC^2:Fem	2.81(.84)	3.36

Table 10. A mixed model shows the effect of gender on complexity-aesthetics relationship (Male serves as the baseline, intercept and slopes for Female are estimated relative to it).

	b coeff (Std.Err)	t val
Intercept	4.00(.12)	32.37
Web Use	.01(.02)	0.56
VC	-8.65(1.12)	-7.72
VC*Web Use	.22(.13)	1.70
VC^2	-4.19(1.12)	-3.74
VC^2*Web Use	-.05(.13)	-0.43

Table 11. A mixed model shows the effect of each additional hour of daily Web use on the complexity-aesthetics relationship.

	b coeff (Std.Err)	t val
Intercept (baseline)	4.08(.27)	15.36
EduL:HighSchool	-.19(.28)	-0.67
EduL:BA	.12(.28)	0.44
EduL:MSc	.19(.32)	0.60
VC	-4.98(2.10)	-2.38
VC:HighSchool	-4.60(2.04)	-2.25
VC:BA	-.44(2.05)	-0.21
VC:MSc	-2.41(2.33)	-1.03
VC^2	-6.12(2.02)	-3.03
VC^2:HighSchool	2.49(1.96)	1.27
VC^2:BA	1.08(1.98)	0.55
VC^2:MSc	1.23(2.25)	0.55

Table 12. A mixed model shows the effect of education level on complexity-aesthetics relationship. (High School without Diploma served as the baseline.)

DISCUSSION

This study aimed at determining the shape and direction of aesthetics-complexity relationship for webpages, which past studies did not agree on. Step 1 re-measured aesthetics and complexity for 5 past datasets (Table 4). The results of past studies were largely replicated, with both a negative correlation and inverted u-shaped curve supported by different dataset, which suggested that the difference in past study results did not occur due a difference in their experimental set-up – we re-measured aesthetics and complexity using the same procedure for all datasets. Not replicated were the results for AVI_14 (both a linear and u-shaped terms significant instead of just a linear term) and ICWE_19 (a significant relationship instead of no relationship). We might speculate AVI_14 collected both aesthetics and complexity with the same participants, which may have artificially strengthened the relationship between the two constructs; ICWE_19 collected complexity together with order using the same participants, which may have artificially weakened the aesthetics-complexity connection due to participant confusion over whether order is different or a part of complexity, as might be [22]. Alternatively, ICWE_19 participants were largely students, unlike our participants, and their webpage evaluations may have been strongly affected by students' expectations of what a typical university website should look like (cf., user expectations [37,38] and prototypicality [46] effects).

Any study relying on Berlyne's theory would need to sample stimuli from the entire simple-complex continuum to observe an inverted U-shaped curve; otherwise, it would find a positive (largely simple stimuli, left side of Berlyne's hypothesized u-shape) or negative (largely complex stimuli, right side of the u-shape) correlation. Given Berlyne's theory is correct, such sampling could have explained the negative correlation in IJHCS_12 and AVI_14. However, IJHCS_12 stimuli covered the entire complexity spectrum and were no different from the rest of webpages (Figure 1), while AVI_14 stimuli were simpler than the rest – the opposite of what would be expected for the sampling explanation to be correct.

Other explanations of the past mixed results might include the presence of confounders in the past datasets, with sampling biases [6] present during dataset building. For example, AVI_14 mostly contained exceptionally well-designed webpages from online repositories, with several unappealing webpages manually found and injected in the dataset to counterbalance the well-designed webpages. Such an unsystematic sampling approach may have introduced biases, and a better approach may include multiple crowdworkers sampling webpages (as in [23]).

Step 2 explored three possible confounders: technical condition, novelty and craftsmanship of webpage design. We identified them by reviewing the webpages of low-complexity/low-aesthetics and high-complexity/high aesthetics for prominent shared features. Future research may rely on this technique to look for tentative explanations of possibly anomalous results. The analysis of identified shared features (Table 6) showed that technical condition was likely a confounder. After accounting for technical condition, the aesthetics-complexity link became largely linear (Table 8), which supported the processing fluency theory of aesthetics and did not support Berlyne's theory.

Berlyne's prediction of aesthetics-complexity connection being a u-shape is not necessarily incorrect (though, a large amount of evidence does say it is [20]). Stimuli may need to be much simpler than the webpages we studied to observe the boredom and disinterest effects that would make the aesthetics-complexity relationship a u-shape. But such stimuli would hardly be valid webpages. Berlyne's theory of aesthetics appears to be, at least, impractical for HCI.

Our results (Figure 5) did not support the past work conclusion [10] of Berlyne's predicted u-shape being a sum of preferences of two distinct groups of users, one seeing complexity as beautiful (positive correlation) and the other seeing simplicity as beautiful (negative complexity-aesthesis correlation). Very few participants saw complexity as beautiful in our study. In addition, two opposite-direction correlations would likely add up to no correlation, or, if one group is larger than the other, to a correlation, but not to a u-shaped curve.

Figure 5 shows substantial variation among individuals in the effect of complexity on their aesthetics scores. The analyses of demographic variables (Table 9, Table 10,Table 11 and Table 12) suggested that some of that variance might be systematic, which corroborates past work [51]. However, the effects of demographic variables were modest, e.g., each extra year of age strengthened the linear component of aesthetics-complexity relationship by 1.4%, whereas each extra hour of daily Web use weakened the same component by 2.5%. Only gender appeared to have a more substantial effect, with the linear component of aesthetics-complexity relationship being 72% stronger and quadratic 52% weaker for women relative to men. These demographic-related results are promising, but tentative, and future research will need to validate them in dedicated studies with many more participants than this study recruited. We would note that demographic factors could not explain between-dataset differences (e.g., Table 4), as the same users rated webpages across different datasets.

Correlational studies do carry the risks of biased stimuli sampling, as we possibly observed at Step 2. However, they also avoid the risks of unrealistic manipulation biases of controlled studies. If complexity of a webpage is manipulated, it, e.g., should not make the webpage look unrealistic or broken (e.g., figure 1 in [14]) or involve unrealistic, invisible font sizes (e.g., figure 2 in [19]). Otherwise, research conclusions are non-generalizable or non-informative for design practice. Future research should control for potential confounders, such as webpage technical condition or novelty, and ensure stimuli validity, e.g., by pre-testing them on being realistic, as in [31].

Our results suggested visual complexity to be only a modest predictor of webpage aesthetics, explaining 4-18% of aesthetics variance depending on the dataset (Table 8). Future research may need to look beyond complexity to explain the bulk of aesthetics variance, e.g., focusing on prototypicality [46] or craftsmanship [29].

The observed weakness of aesthetics-complexity relationship and presence of systematic demographic-variable effects suggested that neither Berlyne's model nor processing-fluency theory explain aesthetics preferences satisfactorily, as both of them attribute a large role to complexity as a predictor and do not rely on demographic factors. However, the processing fluency theory explains our data better than Berlyne's model, and we might suggest design practice relies on it and strives to minimize visual complexity in design, instead of increasing it to medium levels (as Berlyne's model suggests). Complexity minimization should not be done at the expense of other design aspects, such as a webpage looking like an actual, well-functioning webpage, e.g., Figure 6.

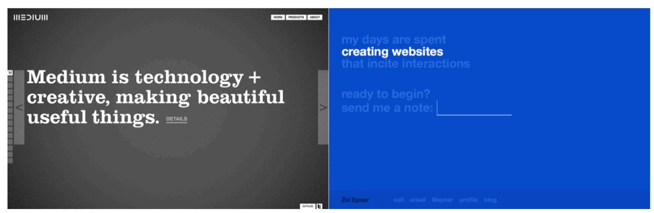

Figure 6. Two examples of webpages that were rated as low complexity, but also as not well functioning, resulting in a relatively low aesthetics score for both.

LIMITATIONS

The study may need to be replicated with confounders measured using multi-item validated scales (e.g., craftsmanship [29]), as well as aesthetics [17]. The study did not test for cross-cultural differences, instead opting for relative participant homogeneity by recruiting participants from English-speaking countries, primarily the US. Such homogeneity reduced conclusion generalizability, as past work showed country-of-origin to determine the shape of aesthetics-complexity relationship [51]. The conclusions are also limited to visual complexity, as participants only viewed, but not interacted with webpages, which future work should also address.

CONCLUSION

A study explored the complexity-aesthetics relationship for webpages using several past datasets. After accounting for a potential confounder – webpages looking broken – we concluded that the relationship appeared to be linear, a negative correlation.

ACKNOWLEDGEMENTS

We thank Alexandre Tuch and Maxim Bakaev for kindly sharing with us the IJHCS_12 and ICWE_19 datasets.

REFERENCES

[1] Armstrong, T. and Detweiler-Bedel, B. Beauty as an emotion: the exhilarating prospect of mastering a challenging world. *Review of general psychology*, 12, 4 (2008), 305-329.

[2] Berlyne, D. E. *Aesthetics and psychobiology*. Appleton-Century-Crofts, New York, 1971.

[3] Bölte, J., Hösker, T. M., Hirschfeld, G., and Thielsch, M. T. Electrophysiological correlates of aesthetic processing of webpages: a comparison of experts and laypersons. *PeerJ*, 5, e3440 (2017).

[4] Boychuk, E. and Bakaev, M. Entropy and Compression Based Analysis of Web User Interfaces. In *International Conference on Web Engineering* (2019), Springer, 253-261.

[5] Cyr, D., Head, M., and Larios, H. Colour appeal in website design within and across cultures: A multi-method evaluation. *International Journal of Human-Computer Studies*, 68, 1 (2010), 1-21.

[6] Delgado-Rodriguez, M. and Llorca, J. Bias. *Journal of Epidemiology & Community Health*, 58, 8 (2004), 635-641.

[7] Douneva, M., Jaron, R., and Thielsch, M. T. Effects of different website designs on first impressions, aesthetic judgements and memory performance after short presentation. *Interacting with Computers*, 28, 4 (2016), 552-567.

[8] Dou, Q., Zheng, X. S., Sun, T., and Heng, P. A. Webthetics: quantifying webpage aesthetics with deep learning. *International Journal of Human-Computer Studies*, 124 (2019), 56-66.

[9] Gori, J., Rioul, O., Guiard, Y., and Beaudouin-Lafon, M. The Perils of Confounding Factors: How Fitts' Law Experiments can Lead to False Conclusions. In *the 2018 CHI Conference on Human Factors in Computing Systems* (2018), ACM.

[10] Güçlütürk, Y., Jacobs, R. H., and Lier, R. V. Liking versus complexity: decomposing the inverted U-curve. *Frontiers in Human Neuroscience*, 10, 112 (2016).

[11] Hassenzahl, M. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19, 4 (2004), 319-349.

[12] Ivory, M. Y. and Hearst, M. A. Improving web site design. *Internet Computing*, 6, 2 (2002), 56-63.

[13] Ivory, M. Y. and Hearst, M. A. Statistical profiles of highly-rated web sites. In *SIGCHI conference on Human factors in computing systems* (2002), ACM, 367-374.

[14] King, A. J., Lazard, A. J., and White, S. R. The influence of visual complexity on initial user impressions: testing the persuasive model of web design. *Behaviour & Information Technology* (2019), 1-14.

[15] Kittur, A., Chi, E. H., and Suh, B. Crowdsourcing user studies with Mechanical Turk. In *the SIGCHI conference on human factors in computing systems* (2008), ACM, 453-456.

[16] Kurosu, M. and Kashimura, K. Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In *Human factors in computing systems* (1995), ACM, 292-293.

[17] Lavie, T., and Tractinsky, N. Assessing dimensions of perceived visual aesthetics of web sites. *Int. journal of human-computer studies*, 60, 3 (2004), 269-298.

[18] Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., and Noonan, P. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *Transactions on computer-human interaction*, 18, 1 (2011), 1-30.

[19] Maity, R. and Bhattacharya, S. Is My Interface Beautiful?—A Computational Model-Based Approach. *IEEE Transactions on Computational Social Systems*, 6, 1 (2019), 149-161.

[20] Martindale, C., Moore, K., and Borkum, J. Aesthetic preference: Anomalous findings for Berlyne's psychobiological theory. *The American Journal of Psychology* (1990), 53-80.

[21] Michailidou, E., Harper, S., and Bechhofer, S. Visual complexity and aesthetic perception of web pages. In *the 26th annual ACM international conference on Design of communication* (Lisbon 2008), ACM, 215-224.

[22] Miniukovich, A., Sulpizio, S., & De Angeli, A. Visual complexity of graphical user interfaces. In *the 2018 International Conference on Advanced Visual Interfaces* (2018), ACM, 1-9.

[23] Miniukovich, A. and De Angeli, A. Computation of Interface Aesthetics. In *33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul 2015), ACM, 1163-1172.

[24] Miniukovich, A. and De Angeli, A. Quantification of Interface Visual Complexity. In *the 2014 International Working Conference on Advanced Visual Interfaces* (Como 2014a), ACM, 153-160.

[25] Miniukovich, A. and De Angeli, A. Visual diversity and user interface quality. In *the 2015 British HCI Conference* (2015b), ACM, 101-109.

[26] Miniukovich, A. and De Angeli, A. Visual Impression of Mobile App Interfaces. In *the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (2014), ACM, 31-40.

[27] Miniukovich, A. and De Angeli, A. Webpage Aesthetics: One Size Doesn't Fit All. In *9th Nordic*

Conference on Human-Computer Interaction (2016), ACM.

[28] Moshagen, M. and Thielsch, M. A short version of the visual aesthetics of websites inventory. *Behaviour & Information Technology*, 32, 12 (2013), 1305-1311.

[29] Moshagen, M. and Thielsch, M. T. Facets of visual aesthetics. *International Journal of Human-Computer Studies*, 68, 10 (2010), 689-709.

[30] Moshagen, Morten and Thiesch, Meinald T. Facets of visual aesthetics. *International jouranl of human-computer studies*, 68, 10 (2010), 689-709.

[31] Post, R., Nguyen, T., and Hekkert, P. Unity in Variety in website aesthetics: A systematic inquiry. *International Journal of Human-Computer Studies*, 103 (2017), 47-62.

[32] Purchase, H. C., Hamer, J., Jameson, A., and Ryan, O. Incesitigating objective measures of web page aesthetics and usability. In *12th Australasian user interface conference (AUIC 2011)* (Perth 2011), Australian computer society, Inc., 19-28.

[33] Reber, R. Processing fluency, aesthetic pleasure, and culturally shared taste. In *Aesthetic Science: Connecting Minds, Brains, and Experience*. Oxford University Press, Oxford, 2012.

[34] Reber, R., Schwarz, N., and Winkielman, P. Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Personality and social psychology review*, 8, 4 (2004), 364-382.

[35] Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., and Liu, J., Gajos, K. Z. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *CHI* (Paris 2013), ACM, 2049-2058.

[36] Rello, L., Pielot, M., and Marcos, M. C. Make it big! The effect of font size and line spacing on online readability. In *CHI'16* (2016), ACM, 3637-3648.

[37] Roth, S. P., Schmutz, P., Pauwels, S. L., Bargas-Avila, J. A., and Opwis, K. Mental models for web objects: Where do users expect to find the most frequent objects in online shops, news portals, and company web pages? *Interacting with Computers*, 22, 2 (2010), 140-152.

[38] Roth, S. P., Tuch, A. N., Mekler, E. D., Bargas-Avila, J. A., and Opwis, K. Location matters, especially for non-salient features–An eye-tracking study on the effects of web object placement on different types of websites. *International journal of human-computer studies*, 71, 3 (2013), 228-235.

[39] Silvia, P. J. Cognitive appraisals and interest in visual art: Exploring an appraisal theory of aesthetic emotions. *Empirical studies of the arts*, 23, 2 (2005), 119-133.

[40] Silvia, P. J. Human emotions and aesthetic experience: An overview of empirical aesthetics. In Shimamura, A. P. and Palmer, S. E., eds., *Aesthetic science: Connecting minds, brains, and experience*. Oxford University Press, 2012.

[41] Silvia, P. J. and Barona, C. M. Do people prefer curved objects? Angularity, expertise, and aesthetic preference. *Empirical studies of the arts*, 27, 1 (2009), 25-42.

[42] Tinio, P. P. and Leder, H. Just how stable are stable aesthetic features? Symmetry, complexity, and the jaws of massive familiarization. *Acta Psychologica*, 130, 3 (2009), 241-250.

[43] Tractinsky, N. Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In *the ACM SIGCHI Conference on Human factors in computing systems* (1997), ACM, 115-122.

[44] Tractinsky, N. Visual Aesthetics. In Soegaard, Mads and Dam, Rikke Friis, eds., *The Encyclopedia of Human-Computer Interaction, 2nd Ed.* The Interaction-Design.org Foundation, Aarhus, 2013.

[45] Tractinsky, N., Katz, A. S., and Ikar, D. What is beautiful is usable. *Interacting with computers*, 13, 2 (2000), 127-145.

[46] Tuch, A. N., Presslaber, E. E., Stocklin, M., Opwis, K., and Bargas-Avila, J. A. The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. *International Journal of Human-Computer Studies*, 70 (2012), 794-811.

[47] Whitfield, T. W. A. Predicting preference for familiar, everyday objects: An experimental confrontation between two theories of aesthetic behaviour. *Journal of environmental psychology*, 3, 3 (1983), 221-237.

[48] Whitfield, T. W. and Slatter, P. E. The effects of categorization and prototypicality on aesthetic choice in a furniture selection task. *British Journal of Psychology*, 70, 1 (1979), 65-75.

[49] Winkielman, P., Halberstadt, J., Fazendeiro, T., and Catty, S. Prototypes are attractive because they are easy on the mind. *Psychological science*, 17, 9 (2006), 799-806.

[50] Wu, O., Chen, Y., Li, B., and Hu, W. Evaluating the visual quality of web pages using a computational aesthetic approach. In *the fourth ACM international conference on Web search and data mining* (2011), ACM, 337-346

[51] Reinecke, K., & Gajos, K. Z. (2014, April). Quantifying visual preferences around the world. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 11-20). ACM.