

Investigating Attractiveness in Web User Interfaces

Jan Hartmann, Alistair Sutcliffe & Antonella De Angeli

Centre for HCI Design, School of Informatics, University of Manchester

P.O. Box 88, Manchester M60 1QD, UK

jan.hartmann@postgrad.manchester.ac.uk; {a.g.sutcliffe} {antonella.de-angeli}@manchester.ac.uk

ABSTRACT

A theoretical framework for assessing the attractiveness of websites based on Adaptive Decision Making theory is introduced. The framework was developed into a questionnaire and used to evaluate three websites which shared the same brand and topic but differed in aesthetic design. The DSchool site was favoured overall and was best for aesthetics and usability. The subjective ratings of the sites were in conflict with the subject-reported comments on usability problems. Subjects were given two scenarios for their preference. They changed their preference from the DSchool to the HCI Group's site for the more serious (PhD study) scenario; however, design background students remained loyal to the DSchool. The implications of framing and halo effects on users' judgement of aesthetics are discussed.

Author Keywords

Aesthetics, usability, attractiveness, websites, judgement biases.

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – *Theory and methods*.

INTRODUCTION

Aesthetic factors beyond traditional usability are increasingly recognised as contributing to the overall success of a product or system [5, 20]. A number of studies have shown correlations between the perceived aesthetic quality of a system's user interface and overall user satisfaction [13, 29], leading to claims that aesthetic design can be a more important influence on users' preference than traditional usability [20]. In a comparison of six different designs of a process control application, with questionnaire inventories for experience, hedonic and appeal qualities, Hassenzahl et al. [8] concluded that both experience and hedonic qualities

contributed approximately equally to the overall judgement of appeal. In experiments using simplified designs of automated teller machines, aesthetics was correlated with perceived usability, and the designs rated to be more aesthetically pleasing were also rated as more usable [29]. However, this study by Tractinsky used very simple variations of an ATM layout and limited manipulations of usability (poor response time). In two experiments comparing users' judgement between two websites which had identical information content but radically different user interface design (one aesthetically metaphor-based, one traditionally menu-based), we found reliable differences between the designs on the expressive aesthetic dimension but not on the classic aesthetics dimension [27, 3]. In these studies we also discovered a framing effect in which the user judgement of aesthetics and overall preference varied according to the question. Users rated the aesthetic metaphor-based site to be superior and preferred it overall even though it had worse usability. When given a preference question framed for serious use, their preferences reversed and the menu site was favoured. It appears therefore that the relationship between users' perception of aesthetic design quality and their overall judgement of website quality is more complex than has been suggested in previous studies.

In this paper we place user perception of aesthetic design in a theoretical framework of decision making and test the hypotheses that aesthetic judgement is liable to framing effects of task as well as the users' background. In the next section we review related research on perception of aesthetics and engagement, while subsequent sections describe the methods followed in the experimental evaluation of three related websites. The results are reported, and the paper concludes with a brief discussion.

RELATED WORK

Several authors have drawn attention to aspects of user interfaces that are variously described as user engagement, experience or emotion in design [4, 15, 20]. McCarthy and Wright [15] view interactive technology as an experience, introducing a framework that describes compositional, sensual, emotional, and spatio-temporal threads of experience as "ways of talking about technology"; while Norman [20] claims that aesthetic design can outweigh usability in the users' overall experience and argues for the emotional impact of good design. Hallnäs and Redström [6] see aesthetics as the logic of expressionals and a foundation for designing for

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2007, April 28–May 3, 2007, San Jose, California, USA.
Copyright 2007 ACM 978-1-59593-593-9/07/0004...\$5.00.

presence, defining an expressional as “a thing that is designed to be the bearer of a certain expression”. Aesthetics as the logic of expressionals is then “concerned with how material builds expressive things”.

Based on Shusterman’s concept of Pragmatist Aesthetics [24], Petersen et al. [23] suggest a framework to distinguish between the aesthetics of use and appearance. They argue that interaction and playful design is an important influence on aesthetic perception, while Djajadiningrat et al. [2] discuss the role of affordances in aesthetic design and Norman’s propositions for affective design with feedback dimensions, and note that the socio-cultural context of aesthetic perception needs to be taken into account.

Lavie and Tractinsky [12] developed a questionnaire-based measurement instrument of perceived website aesthetics and defined their two dimensions: *classical* aesthetics emphasises orderly and clear design and is related to design rules advocated by usability experts, e.g. principles such as consistent and structured layout, symmetry, clean and clear design; whereas *expressive* aesthetics is “manifested by the designer’s creativity and originality and by the ability to break design conventions” with questions focused on users’ perception of user interface qualities, such as “beautiful”, “challenging” and “fascinating”. However, these questionnaires only elicited users’ high-level perceptions of design quality, essentially the look and feel, rather than requested judgement about components of the design.

Kim et al. [11] relate design factors and aesthetic responses to emotions that users frequently feel when viewing relationships to aid design of emotionally evocative home pages. Park et al. [21] analysed critical factors that determine the degree to which users feel the impressions intended by the designers (aesthetic fidelity), and found that variability of user perception was closely related to the aesthetic fidelity of web pages.

In previous research we proposed a set of evaluation heuristics which did focus more clearly on judgement of design features to analyse a more widely based construct of “attractiveness” [26]. This feature-based evaluation of attractiveness introduces heuristics for attractive design of interactive systems that can also be operationalised as design-guidelines. Assessing different interaction styles for website user interfaces, we demonstrated a trade-off between engaging and usable design [27, 3], and the impact of interaction style on the perception of content. This paper extends current work by reporting an exploratory investigation into the influence of personal and contextual factors on the relative importance of various qualities that contribute to the overall attractiveness of a system.

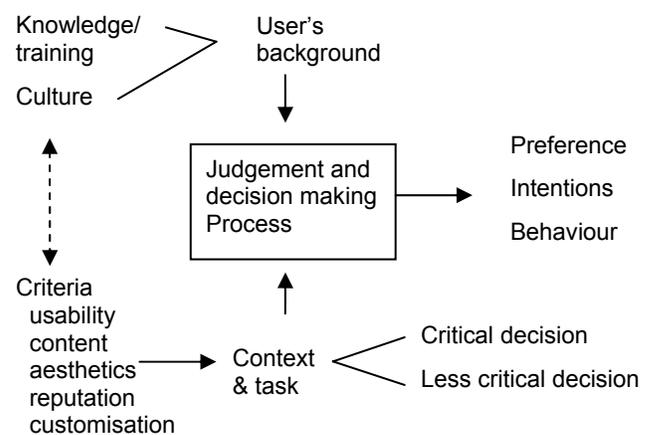


Figure 1: Theoretical Framework for Judgement.

FRAMEWORK FOR ATTRACTIVENESS

Based on previous evaluations of websites [10, 26, 27], we propose a theoretical framework for judgement of “attractiveness” (“pleasing or appealing to the senses, arousing interest” Oxford English Dictionary) based on Adaptive Decision Making theory [22], illustrated in Figure 1. ADM theory asserts that people’s decision making is adaptive and contingent upon the task, context, and their background-experience.

Our elaboration of ADM hypothesises that users’ judgement will depend on their background, in particular culture and training, the nature of the task, the importance or criticality of the decision, and interactions between decision-making criteria (e.g. design qualities such as content, functionality, usability) that are conditioned by the task context. The outcomes of users’ judgement are preferences between designs, intention to use, and actual use (behaviour).

Five judgement criteria are proposed: *usability* following traditional definitions (e.g. [9]), including ease of learning, efficiency of use, memorability, low error frequency, and subjective satisfaction. Appropriate and interesting *content* is widely cited to be a key factor in successful website design [14, 17, 19] with *services* to describe the functions that aggregate into utility. *Aesthetics* reflects the format in which the content and services are presented as well as the design look and feel and overall experience with a system [6, 15, 20]. *Reputation/identity* relate to the identity of the website owner and the brand-product identity which can be projected by a consistent visual style, logos, and product presentation [16]. Finally, *customisability* describes the ability for the user to adapt the system to his or her needs, which can encourage users to take ownership over a system and has been found to influence perceived usability and aesthetics [1].

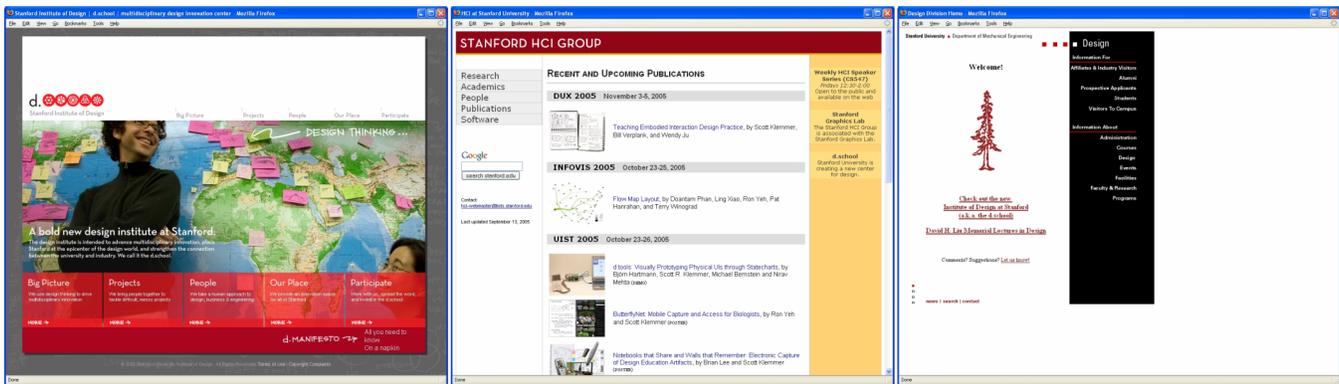


Figure 2: The home pages of the DSchool, HCI Group, and Design Division in October 2005.

In this paper we test three hypotheses that cover part of the theory's scope:

H1. User preference will be determined by interactions between decision criteria and subject background, specifically design-training and aesthetics, culture and identity

H2. User intentions will be determined by interactions between decision criteria and the task context; specifically, serious use will favour usability and content, less serious use will favour aesthetics.

H3. User judgement will be determined by interactions among decision criteria; specifically, positive aesthetics will over-rule poor usability.

METHODS

Three live departmental websites of Stanford University were compared: the Stanford Design Division¹, the Stanford HCI Group², and the Hasso Plattner Institute of Design at Stanford³ (DSchool). The sites were chosen for their expected variation in usability, content and aesthetic design. Although all three sites belonged to Stanford University and hence shared the same reputation, the DSchool was a new entity within the organisation. Figure 2 shows the home pages of the three departmental websites in September / October 2005, when the study was carried out.

The three sites represent departments generally concerned with the design of technical systems, and all promote an interdisciplinary approach. The departmental websites differ significantly in the amount of content provided, as well as in their visual presentation. Table 1 summarises an informal assessment of the three websites regarding content, aesthetics, usability (especially navigation), reputation, and customisability.

The complete website of the DSchool consists of 26 pages; in contrast, the websites of the HCI Group and Design Division

each contain over 100 pages. The HCI Group's website followed a traditional tabular layout and a text-dominant style with small pictures complementing the text sections. The Design Division's website implemented a simple style using only three colours and dividing the website into navigation and content components. Apart from the logo on the home page, it was exclusively text based. The DSchool had a more interactive style, dynamically reacting to mouse movements (underlining links, pop-up navigation menus), and a colourful and image-intensive design. All the websites could be navigated via a simple menu structure, so the complexity of navigation was similar. As all departments were from the same university, no significant inter-site effect of brand was expected. None of the websites could be customised, so this aspect of the framework is not addressed further in this paper.

Participants

The subject population (mean age 25.8, range 20-40, 24 male and 19 female) was composed of undergraduate, postgraduate, and HND students of the University of Manchester, Manchester Metropolitan University, and the Manchester College of Arts and Technology. The participants were categorised in three groups based on their background: 14 students from a Western European / North

	DSchool	HCI Group	Design Divn
Content	26 pages in total; two levels	> 100 pages in first two levels	> 100 pages in first two levels
Aesthetics	Heavy use of imagery, colourful, hover-links	Traditional tabular style	Plain, simple, text-based
Usability / Navigation	5-section menu with 2-5 items / section	5-item menu	13-item menu, organised in 2 sections
Reputation	Same university		
Customisability	Not customisable at all		

Table 1: Informal assessment of website qualities.

¹ <http://design.stanford.edu>

² <http://hci.stanford.edu>

³ <http://dschool.stanford.edu>

American origin and with a technical course-background (e.g. computer science, engineering degrees), 15 Western students of design-related courses (e.g. multimedia design, filmography and design, freelance web-designers) and 14 Chinese students with technical course-background. Chinese design students were not available hence a partially factorial design was adopted, i.e. no design x culture comparison could be made although a culture comparison among technical students was possible. All subjects had some technical design component in their studies, hence they all possessed relevant knowledge and a similar understanding for assessing the research areas in the three departments. All subjects were rewarded with £10.

Procedure

Data was collected in individual interviews lasting about one hour. Each participant was briefed about the intent of the study and completed a demographic questionnaire, and was then asked to complete the following tasks for each website. The order of the websites was randomised.

- (1) Browse the website to familiarise themselves with the department. While browsing, subjects were asked to report their impressions of the website, including thoughts on content, presentation and usability issues.
- (2) At the end of the browsing task, subjects were asked to list three keywords that described the department and to list three key people in the department. Subjects were allowed to refer to the respective website at all times during the study.
- (3) The website was assessed with a questionnaire based on the framework, with five questions each on the perceived quality of the content, usability, aesthetics, reputation/identity, and customisability. The questions for each component and an overall preference question used a five-point Likert scale. Finally, subjects were asked to name the top three to five best and worst features of the respective websites.
- (4) Subjects were asked to briefly revisit all three websites, and then compare them in two scenarios. In both scenarios they were asked to imagine being a student who will be going to study at Stanford University and to rank the three departments in order of preference. In one scenario the subjects were asked to imagine being a Bachelor student who is going to Stanford University for a one-month summer internship (Figure 3), whereas in the other the subject was asked to imagine being a Masters student who is going to Stanford for a PhD research studentship. The order of the scenarios was randomised between subjects.
- (5) At the end of the study, the subjects were asked to rank the five quality criteria, aesthetics, content, customisability, reputation/identity, and usability, in order of relative importance for their overall judgement of a website and to note further comments on the designs.

TASK

Imagine you are a Bachelor student who is interested in the design of technical systems. You apply for a funded **one-month summer internship** at Stanford University and are accepted.

To complete the process you are asked to **prioritise the three** departments of the university, which are **all** pursuing research in your area of interest. For more information you are kindly asked to refer to the respective websites of the three departments, the HCI Group (<http://hci.stanford.edu>), the Design Division (<http://design.stanford.edu>), and the Institute of Design (<http://dschool.stanford.edu>).

Please rank the departments in order of preference! What is the basis for your decision?

Figure 3: Description of “summer-internship” scenario.

RESULTS

Quality Assessment and Performance

The questionnaire scales used for measuring the perceived quality of the websites’ content, usability, and aesthetics were all reliable (each scale’s Cronbach’s Alpha > .88). The reliability for perceived reputation/identity was weak, so the question on perceived brand strength that did not correlate with the other items of this scale was removed. The new four-item scale used in all subsequent analyses showed improved reliability (Cronbach’s Alpha = .80). See Table 2 for the correlation matrix of the final evaluation measures.

For each of the framework components, inter-site and inter-group differences were analysed by a 3 x 3 mixed model ANOVA with website (3) as the within-subjects factor and participant background (3) as the between-subjects factor. Simple contrasts were run to test the effect of training (2, Western technical vs. Western designer) and culture (2, technical Western versus technical Chinese).

Qualitative data, consisting of the subjects’ comments on content, presentation, and usability issues, complemented the quantitative analysis of differences in perceived quality for each framework component. The comments were recorded as the subjects were browsing the website. There were no performance differences in task completion times. All subjects found three people who they considered to be important in each site and the choices were consistent within the HCI Group and Design Division sites. For the DSchool the choice was less consistent, as some participants experienced some difficulty in selecting people, because information was only displayed when hovering over or after clicking on a person’s photograph.

	Content	Usability	Aesthetics	Reputation
Content		.431 *	.168	.515 *
Usability	.431 *		.429 *	.503 *
Aesthetics	.168	.429 *		.572 *
Reputation	.515 *	.503 *	.572 *	

Table 2: Correlations of evaluation measures (* = $p < .01$).

	DSchool	HCI Group	Des Divn
Total Problems	19	19	25
Mean problems per subject	.44	.44	.58
Mean Rating of Usability	4.07	3.64	3.30

Table 3: Number of usability problems and mean ratings of usability per website.

Usability

The perceived usability of the three websites differed significantly between sites ($F_{(2, 80)} = 7.12, p < .01$) with no significant interactions. On mean rating scores the DSchool's was most usable (mean = 4.07, SD = .97), followed by the HCI Group (mean = 3.64, SD = 1.02) and Design Division (mean = 3.30, SD = 1.16). There was no significant effect for the subjects' backgrounds, i.e. training or culture.

Table 3 shows the number of usability problems reported per website. Overall an average of .48 usability issues was reported per participant per website. The most common problems for the HCI Group were excessive scrolling (reported by 7 participants), acronyms not explained (7) and inconsistent positioning of the navigation bar (4). The Design Division also encountered scrolling complaints (4), as well as broken links / navigation problems (8) and the search facility not working (4), while the DSchool problems were small font sizes (6) clicking on images inducing unpredictable actions (5), and general difficulty of navigation (4).

Even though an equal number of similarly severe usability problems were reported for the DSchool and HCI Group and the assessments of the navigational complexity of the website (Table 1) were very similar, the DSchool was rated significantly better for usability ($p < .05$). This indicates a mismatch between the subjects' actual usability experience and their overall impression.

Content

The perceived quality of content differed between the websites ($F_{(2, 80)} = 5.70, p < .01$) with no significant interactions. Taking the rank order of mean scores, the HCI Group was best (mean = 4.02, SD = .84) followed by the Design Division (mean = 3.68, SD = .79), then the DSchool (mean = 3.43, SD = .96). There was no significant effect for the subjects' backgrounds.

The subjects' comments on content were classified as positive, neutral, or adverse (i.e. critical). The DSchool's content attracted mainly negative comments. The most frequent complaint about the DSchool was insufficient detail / lack of information (reported by 20 participants), a generally "advertising-like" style (5), and specifically the slogans on each page (3), although several other subjects liked the slogans (4). The HCI Group received more favourable than unfavourable comments overall. The most frequent adverse criticism concerned the lack of an

introduction or overview about the department (18). The Design Division's content comments were either neutral or favourable.

Aesthetics

Aesthetics ratings were significantly different between websites ($F_{(2, 76)} = 65.56, p < .001$), with a significant interaction between participant background and sites ($F_{(4,76)} = 3.09, p < .05$) (Figure 4).

On average ratings, the DSchool's website was best (mean = 4.23, SD = .71), followed by the HCI Group (mean = 2.60, SD = 1.27), with the Design Division in third place (mean = 2.13, SD = .97). There was no significant difference in the ratings by cultural background, but significant difference by training (contrast est. = -.686, $p < .01$). While participants with a technical background rated the HCI Group in second place (mean = 2.78, SD = 1.34), designers rated it the lowest of all three websites (mean = 1.59, SD = .76).

The subjects' comments followed the same pattern as their ratings with the quality order DSchool, HCI Group and the Design Division in third place, and a similar difference between participants with technical and design backgrounds.

The DSchool received more comments than the other sites. They were nearly all favourable, and all groups commented that the DSchool's website was well designed aesthetically, with good use of imagery (19 comments), the colour-scheme (11), interactive links (11), and the hand-written fonts (5).

The Design Division's webpage received mostly negative comments. It was especially criticised for the lack of images (13), bad use of space (12), and its colour-scheme (8).

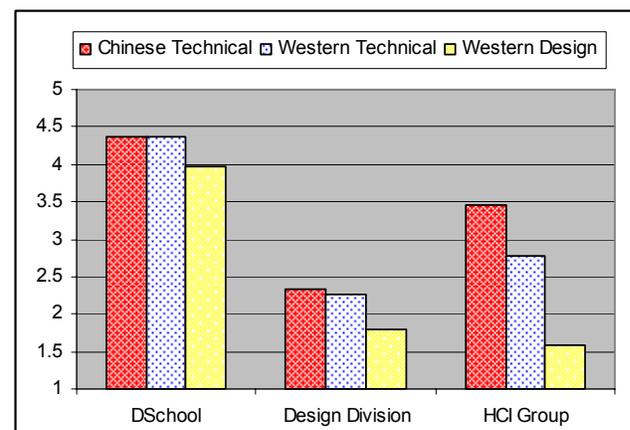


Figure 4: Aesthetics ratings per website per participant group.

The HCI Group's comments are listed in Table 4 by participants' training, as the subjects' training backgrounds had a strong effect, $p < .01$ Chi Square on distribution of positive / negative comments by site. The design subjects were more forthcoming with comments, with an average 3.0

comments for each website; while technical subjects were more restrained with 1.69 comments each on each website. Furthermore, the design students' comments were more detailed and emotive, e.g. they commented not only on the presence of images, but also on their quality, expressing strongly worded opinions.

Design-Students (total comments = 40)	Students with technical course-background (15)
<i>Negative comment on design in general - from "awful" to "not really the most attractive" (15)</i>	Pictures in people section nice (4)
Technical design (6)	<i>Not very attractive, but functional (3)</i>
<i>Quality of images bad (6)</i>	Colour scheme good (2)
Good to have pictures along-side people and projects (5)	
<i>Colour scheme bad (3)</i>	
<i>5 other comments (5)</i>	6 other comments (6)
Aesthetic Rating	
Mean = 1.61, SD = 0.75	Mean = 2.78, SD = 1.34

Table 4: Comments by Western participants on aesthetics by training (technical vs. design) for the HCI Group's website and mean ratings. Negative comments are in italics.

Reputation / Identity

Perception of reputation / identity differed significantly between websites ($F_{(2,76)} = 6.31, p < .01$), with a significant interaction between site and participant background ($F_{(4,76)} = 3.24, p < .05$).

The DSchool (mean = 4.07, SD = .85) was perceived as best, followed by the HCI Group (mean = 3.75, SD = .88), and Design Division (mean = 3.46, SD = 1.03). Subject training significantly affected judgement (contrast est. = -.601, $p < .01$).

All groups rated the DSchool positively, but differed in their judgements of the HCI School and Design Division. Technical background students showed no significant difference between the three websites; in contrast, the designers' judgement of HCI and Design Division's reputation was poor and this was correlated with their perception of aesthetics for the HCI Group ($r = .72, p < .01$ Spearman) and Design Division ($r = .70, p < .01$).

We hypothesise that the design students identify with and treat the DSchool as a sub-brand, and this may be linked to their judgement of the DSchool's aesthetic presentation; in contrast, technical students seem to rate the Stanford brand equally across all three websites.

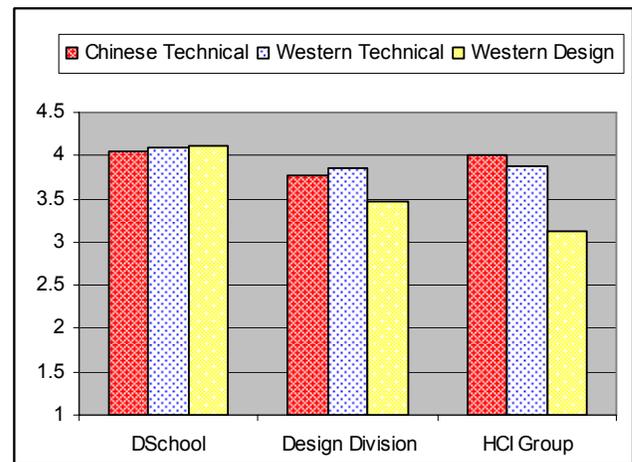


Figure 5: Average ratings for reputation / identity by website and participant group.

Overall Preference

The overall preference for a website varied significantly between sites ($F_{(2,80)} = 10.06, p < .001$) with an interaction between participant background and site ($F_{(4,80)} = 4.34, p < .01$).

The DSchool was rated most favourably (mean = 3.63, SD = .85) followed by the HCI Group (mean = 3.02, SD = 1.06), and the Design Division (mean = 2.79, SD = 1.01) (Figure 6). Subjects' training did have a significant effect on overall preference (contrast estimate = -.654, $p < .01$). At the site level the technical subjects preferred the Design Division more strongly ($t_{(27)} = 3.216, p < .01$).

While there were no differences by cultural background in the ratings for the individual qualities, for overall preference there was a difference by culture for the DSchool ($t_{(26)} = 3.045, p < .01$). The Western students preferred the DSchool, while the Chinese participants did not rate the DSchool better than the HCI Group.

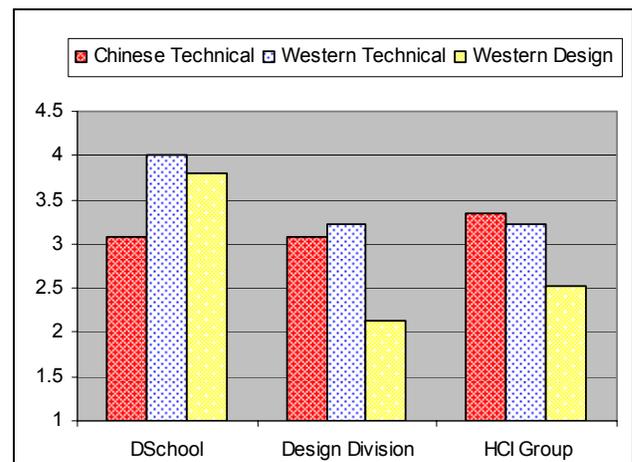


Figure 6: Average ratings for overall preference by website and participant group.

	Usability	Content	Aesthetics	Reputation	Overall
DSchool	+++	+	+++	+++	++
HCI Group	++	+++	=	++	+
Design Division	+	++	-	+	=

Figure 7: Summary ratings for the three sites; shading denotes where subjects' comments conflicted with their ratings (+ positive, = neutral, - negative).

To summarise, the DSchool is perceived as the most aesthetic and usable, and is preferred overall (Figure 7). This perception seems to override the subjects' criticism for poor usability and content. The HCI Group was perceived to be weaker on aesthetics and usability, even though the objective evidence of usability problems did not support this perception. The Design Division's website did not evoke strong comments on content and was perceived to be the least aesthetic and usable.

Judgement between subject backgrounds differed for training, in aesthetics, reputation, and overall preference. The difference in rating of reputation is curious since all three sites shared the same brand (Stanford University); however, the inter-group differences provide a clue. The DSchool may be perceived as a separate sub-brand; hence it receives a higher rating. Judgement of individual qualities did not seem to differ between cultural groups. However, the overall preference did differ for the DSchool's website, where the Chinese participants rated it significantly lower than Western subjects ($p < .01$).

Feature Analysis

The most frequent design features cited by the subjects were in descending order: colours which were mainly positive for the DSchool, disapproval for the HCI Group, and mixed comments for the Design Division; and use of images which was positive for the DSchool but negative for the Design Division (too few images). Next were the content comments reported earlier, followed by clear / clean layout, a positive aspect for the DSchool, and hover / pop up links that were approved of on the DSchool site. The Design Division was criticised for poor use of space, followed by poor readability of text spread across all three sites. Absence of a clear Stanford University logo was also noted, followed by several other features cited by < 3 users.

Decision Making in Scenarios

Subjects were asked to rank the three departments according to the situation described in two scenarios. Most Western subjects (82%) changed their preference order. For the summer internship 90% of the subjects selected the DSchool as their first choice, whereas for the PhD research-studentship 52% of subjects preferred the HCI Group (Table 5). When the scenario preferences were examined by subject training background, all the technical subjects changed their preferences, whereas 46% of the designers did not and remained loyal to the DSchool.

	DSchool	HCI Group	Design Divn
Internship			
1 st	26	1	2
2 nd	3	17	9
3 rd	0	11	18
PhD			
1 st	9	15	5
2 nd	13	7	9
3 rd	7	7	15

Table 5: Frequency of Western subjects expressing ranked preferences for each site for the Internship and PhD study scenarios.

Of the subjects who changed their preference, all swapped the first rank position apart from one designer and one technical subject. All the loyal subjects chose the DSchool as their first preference in both scenarios and indicated that the look and feel or aesthetics of the website were their main driver in determining their choice. The subjects who did show the scenario framing effect mentioned as a basis for their decision that they were looking for a more "fun" place for the summer internship, and a more "serious" or "research-oriented" department for a PhD research studentship. They then chose the department with the website that matched their expectation for the scenario.

This agrees with the subjects' general impressions recorded after browsing each website at the beginning of each session. Table 6 summarises the most frequently cited keywords for each website. The DSchool was perceived as being a more fun, friendly, and welcoming place to study, whereas the other two departments were perceived as more hard-working, serious, and research-oriented.

When asked to express the basis for their decision in relation to the qualities of the website, 87% of the subjects who showed the framing effect indicated that the look and feel of the website was the most important determinant for the summer internship, whereas for the PhD research-studentship the content of the website was decisive.

DSchool	HCI Group	Design Division
fun / friendly / welcoming (16)	research oriented (6)	research oriented / dedicated to work (12)
innovative / creative (12)	hard working (6)	organised / efficient (9)
"show-off" / arrogant / basing everything on reputation (6)	organised (5)	boring / dull / not interesting (8)
	technical (6)	cold / distant (7)
		experienced / high reputation / many affiliations (6)

Table 6: Keywords used to describe the department; cut-off 6 or more subjects.

	DSchool	HCI Group	Design Divn
Internship			
1 st	6	5	3
2 nd	5	4	5
3 rd	3	5	6
PhD			
1 st	4	6	4
2 nd	4	6	4
3 rd	6	2	6

Table 7: Chinese subjects expressing ranked preferences for each site for the Internship and PhD study scenarios.

However, the Chinese subjects were much less sensitive to the scenario (Table 7). 71% of Chinese students were loyal to their first place choice. Of these, 80% indicated that either content (2), reputation (3), or a combination of both (3) was decisive for their decisions in both scenarios. Concerning “reputation” the following criteria were mentioned: age/history/background of department, industry affiliations, size of department, perceived importance of department within university, perceived “wealth” / amount of funds available.

General Importance

Subjects were asked to rank the quality dimensions aesthetics, content, customisability, reputation/identity, and usability in order of importance for their overall judgement when evaluating websites more generally. Figure 8 shows the distributions of ranks for each measure. Content was the first or second most important component for 81% of the subjects (mean rank = 1.43, SD = .79) with a significant difference by participant background ($F_{(2,42)} = 3.61, p < .05$). In contrast, customisability was the least important for 78.6% of the subjects (mean = 4.71, SD = .60). Usability occupied the second place overall (mean = 2.71, SD = 1.01), with 74% of the subjects ranking usability in second or third place.

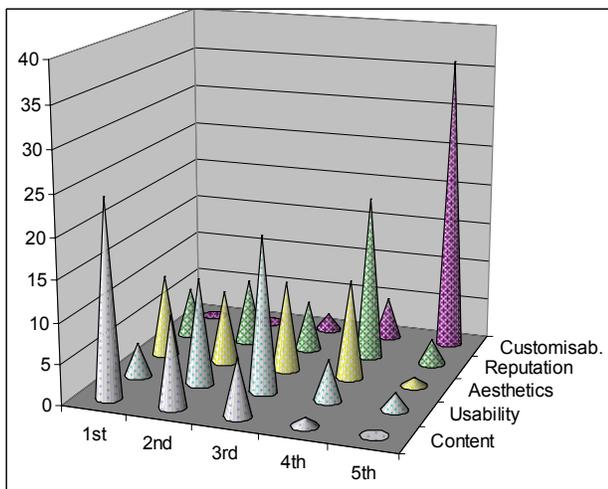


Figure 8: Ranking of framework components according to relative importance for overall judgement, frequency by rank where 1 = most important, 5 = least important

Aesthetics overall occupied third place (mean = 3.25, SD = .89) with an almost equal distribution among the first four places (1st place = 10, 2nd = 9, 3rd = 11, 4th = 12) and a strong effect by participant background ($F_{(2,42)} = 22.55, p < .001$). As expected, designers rated this component as more important than other groups. When the distribution by sub-groups was examined, aesthetics appeared to be very important for designers, two-thirds of them ranking it in first place; whereas for technical subjects, aesthetics rankings were evenly distributed over second, third, and fourth places. Chinese participants gave the least importance to aesthetics, with 64% ranking aesthetics on one of the last two places (Figure 9).

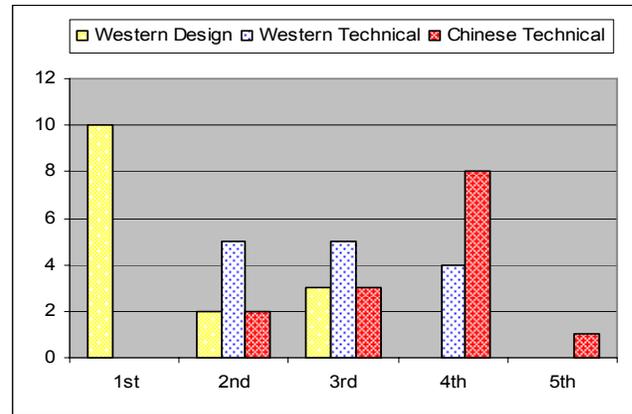


Figure 9: Ranking of importance of aesthetics by participant background.

Reputation was in fourth place (mean = 2.86, SD = 1.27), with a weak inter-group effect ($F_{(2,42)} = 3.75, p < .05$). Analysing the ranking orders of reputation by participant background shows a possible hybrid background-culture effect. 57% of Chinese technical participants ranked reputation in first or second place, compared to 57% of Western technical and 80% of Western design students who ranked reputation in the last two places (Figure 10).

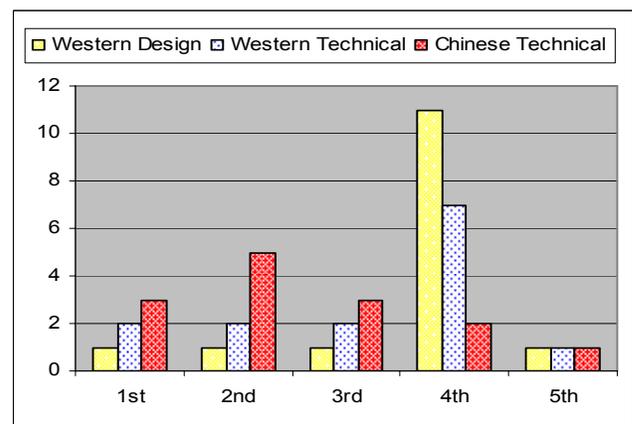


Figure 10: Ranking of importance of reputation by participants background

DISCUSSION

Our findings agree with previous studies that reported a correlation between perceived usability and aesthetics [12, 28, 29]. However, we found a difference between objective usability in reported problems and subjective usability ratings, with the positive perceived usability showing a halo effect to override users' poor usability experience. This suggests that users' overall impression of a website could be a determinant of user satisfaction and system acceptability, even overcoming poor usability experience and poor content.

To revisit our hypotheses, H1 was partially supported, since we did find a strong interaction between training background and aesthetics for the design students. The effects of culture were weak with a small negative effect on judgement of content and aesthetic ratings for the DSchool. However, there was a difference in overall preference for the DSchool, where although Chinese students recognised the DSchool being superior in aesthetics, it may have been less influential for their overall judgement than for Western participants, leading to the lower overall rating, analogous to the differences in the importance ranking and the scenarios.

H2 was supported since the framing scenario did change the subjects' intent and preferences from the DSchool for a "less critical" summer internship to the HCI Group for the "more critical" PhD study. Furthermore, the evidence from the subjects' comments suggested that aesthetics played a more important role in the less critical scenario, while content was more important for the more critical PhD scenario.

Finally, the result on H3 was complex with several interactions between usability, aesthetics, content and overall preference. The direction of influence of possible halo effects is difficult to determine from our data. If we take the importance ratings as a guide then we might expect judgement of content to be the dominant variable influencing the less important aesthetics and usability. However, if content were dominant then we might have expected the HCI Group to receive a better overall rating. Instead the conflict between a positive content rating and many negative comments suggests that aesthetics may have been the more dominant variable. Given the DSchool's consistent positive rating on usability, aesthetics and overall preference, we argue that a combination of criteria may influence overall preference rather than any one variable having a dominant effect. Our results suggest that the task scenario has the strongest influence since it overturned preference for the DSchool; background may have a less strong but important influence when the subjects' background area and decision criteria are related, as for aesthetics and our design subjects. Culture, on the other hand appears to have only a weak influence on judgement of quality, although all these conclusions have to be qualified by the limits of the study to one set of websites, with only two comparisons of background and culture.

The attractiveness framework we propose gives a more comprehensive view of design quality and extends a range of subjective measures of emotive and aesthetic factors [6, 12]. The importance rankings indicate that contents and services, implicitly utility, are the dominant factor, in agreement with website development guidelines [18]. However, content was not a major discriminant between the websites in our study. Instead we believe a more subtle interaction between aesthetic, brand and content explains the preference shown for the DSchool site. Aesthetics and usability vied for second and third place in the importance rankings. The incongruence between importance rankings and actual choice resemble results of Tractinsky and Zmiri [30] on preferences of media player skins.

Our study suggests that the relative importance of aesthetics and usability is sensitive not only to the users' background but also to the users' goal or task. The inter-group differences in judgement for aesthetics were marked. The scenario framing effect was strong; however, design students were more resistant to changing preferences, since the DSchool probably matched their requirements for both scenarios; furthermore, there may have been a halo effect from aesthetic judgement influencing their overall preference. An implication for these findings is to reinforce the well known advice "know your audience" [14, 25], but to refine it as a heuristic "know your audience's preferences and expectations". We argue that design priorities for aesthetics, usability, content or other components in our attractiveness framework should be matched to the user profile and application domain.

In conclusion, we have demonstrated that while aesthetics is an important component of design quality, perception of aesthetics is susceptible to the users' background and task. Usability is important but good aesthetic design can overcome some deficit of usability problems. The strength of the halo effect needs further research. In our study the usability problems were not serious; with more severe problems the halo effect from favourable aesthetics may evaporate if users lose trust in the system. Another area for further research is extending the concept of attractiveness to encompass interaction, and engagement (see also [7]). We have demonstrated a similar halo effect for engagement metaphors influencing overall preference in face of worse usability in educational websites with a similar scenario framing effect [27]. We will refine our evaluation instruments to measure the relative contribution of interactive as well as presentation aesthetics in the future.

REFERENCES

1. Blom, J., and Monk, A. 2003. Theory of personalization of appearance: Why users personalize their pcs & mobile phones. *Human-Computer Interaction*. 18. 3, 193-228.
2. Djajaningrat, J. P., Overbeeke, C. J., and Wensveen, S. A. G., 2000. Augmenting fun and beauty: a pamphlet. In *Proceedings Designing Augmented Reality Environments, DARE 2000* (Elsinore, Denmark). 131-134.

3. DeAngeli, A., Sutcliffe, A., and Hartmann, J., Interaction, usability and aesthetics: what influences users' preferences?, Proc. DIS 2006, ACM Press (2006), 271-280.
4. Dourish, P. 2004. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, Cambridge MA.
5. Green, W. S., and Jordan, P. W. 2001. *Pleasure with Products: Beyond Usability*. Taylor & Francis, London.
6. Hallnäs, L., and Redström, J. 2002. From use to presence: On the expression of aesthetics of everyday computational things. *ACM Transactions on Computer-Human Interaction*. 9. 2, 106-124.
7. Hassenzahl, M. 2004. The interplay of beauty, goodness and usability in interactive products. *Human-Computer Interaction*. 19. 4, 319-349.
8. Hassenzahl, M., Platz, A., Burmester, M., and Lehner, K., 2000. Hedonic and ergonomic quality aspects determine a software's appeal. In *CHI 2000 Conference Proceedings Conference on Human Factors in Computing Systems* (The Hague) T. Turner, G. Szwillus, M. Czerwinski and F. Paterno (Eds), ACM Press, New York, 201-208.
9. ISO. 1997. *ISO 9241: Ergonomic Requirements for Office Systems with Visual Display Terminals (VDTs)*. International Standards Organisation
10. Ivory, M., and Hearst, M. 2001. The state of the art in automated usability evaluation of user interfaces. *ACM Computing Surveys*. 33. 4, 173-197.
11. Kim, J., Lee, J., and Choi, D. 2003. Designing emotionally evocative home pages: An empirical study of the quantitative relations between design factors and emotional dimensions. *International Journal of Human-Computer Studies*. 59. 6, 899-940.
12. Lavie, T., and Tractinsky, N. 2004. Assessing dimensions of perceived visual aesthetics of websites. *International Journal of Human-Computer Studies*. 60. 3, 269-298.
13. Lindegaard, G., and Dudek, C. 2003. What is this evasive beast we call user satisfaction? *Interacting with Computers*. 15. 3, 429-452.
14. Lynch, P. J., and Horton, S. 2001. *Web Style Guidelines* (2nd ed.). Yale University Press, New Haven CT.
15. McCarthy, J., and Wright, P. 2005. *Technology As Experience*. MIT Press, Cambridge MA.
16. Merrilees, B. and Fry, M.-L. 2002. Corporate branding: A framework for e-retailers. *Corporate Reputation Review*, 5, 213-225.
17. Mullet, K., and Sano, D. 1995. *Designing Visual Interfaces: Communication Oriented Techniques*. SunSoft Press, Englewood Cliffs NJ.
18. Nielsen, J. 2000. *Designing Web Usability: The Practice of Simplicity*. New Riders, New York.
19. Nielsen, J., and Molich, R. 1990. Heuristic evaluation of user interfaces. *SIGCHI Bulletin*. April: special issue, 249-256.
20. Norman, D. A. 2004. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books, New York.
21. Park, S., Choi, D., and Kim, J. 2004. Critical factors for the aesthetic fidelity of web pages: Empirical studies with professional web designers and users. *Interacting with Computers*. 16. 2, 351-376.
22. Payne, J. W., Bettman, J. R., and Johnson, E. J. 1993. *The Adaptive Decision Maker*. Cambridge University Press, Cambridge.
23. Petersen, M. G., Iversen, O. S., Krogh, P. G., and Ludvigsen, M., 2004. Aesthetic interaction: a pragmatist's aesthetics of interactive systems. In *Proceedings 2004 Conference on Designing Interactive Systems: processes, practices, methods, and techniques* (Cambridge MA).
24. Shusterman, R. 1992. *Pragmatist Aesthetics, Living Beauty, Rethinking Art*. Blackwell, Oxford.
25. Spool, J. M., Scanlon, T., Snyder, C., Schroeder, W., and DeAngelo, T. 1999. *Website Usability: A Designer's Guide*. Morgan Kaufmann, San Francisco.
26. Sutcliffe, A. G., 2002. Heuristic evaluation of website attractiveness and usability. In *Proceedings 8th Workshop on Design, Specification and Verification of Interactive Systems* (Glasgow), C. W. Johnson, Ed. Springer-Verlag, Berlin, 188-199.
27. Sutcliffe, A. G., and De Angeli, A., 2005. Assessing interaction styles in web user interfaces. In *Proceedings Human Computer Interaction - Interact 2005* (Rome) M. F. Costabile, F. Paterno (Eds). Springer Verlag, Berlin, 405-417.
28. Tractinsky, N., 1997. Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In *Human Factors in Computing Systems CHI 97 Conference Proceedings* (Atlanta GA), Pemberton S. (Ed.) ACM Press, New York, 115-122.
29. Tractinsky, N., Shoval-Katz, A., and Ikar, D. 2000. What is beautiful is usable. *Interacting with Computers*. 13. 2, 127-145.
30. Tractinsky, N., and Zmiri, D., 2006. Exploring Attributes of Skins as Potential Antecedents of Emotion in HCI. In *Aesthetic Computing*, Fishwick, P. (Ed), MIT Press Cambridge, MA, 405-421