

2004

Using Ratings and Response Latencies to Evaluate the Consistency of Immediate Aesthetic Perceptions of Web Pages

Noam Tractinsky

Ben-Gurion University of the Negev, noamt@bgu.ac.il

Avivit Cokhavi

Ben-Gurion University of the Negev, cohavi@bgu.ac.il

Moti Kirschenbaum

Ben-Gurion University of the Negev, kirchenb@bgu.ac.il

Follow this and additional works at: <http://aisel.aisnet.org/sighci2004>

Recommended Citation

Tractinsky, Noam; Cokhavi, Avivit; and Kirschenbaum, Moti, "Using Ratings and Response Latencies to Evaluate the Consistency of Immediate Aesthetic Perceptions of Web Pages" (2004). *SIGHCI 2004 Proceedings*. 13.

<http://aisel.aisnet.org/sighci2004/13>

This material is brought to you by the Special Interest Group on Human-Computer Interaction at AIS Electronic Library (AISEL). It has been accepted for inclusion in SIGHCI 2004 Proceedings by an authorized administrator of AIS Electronic Library (AISEL). For more information, please contact elibrary@aisnet.org.

Using Ratings and Response Latencies to Evaluate the Consistency of Immediate Aesthetic Perceptions of Web Pages

Noam Tractinsky, Avivit Cokhavi, Moti Kirschenbaum

Department of Information Systems Engineering
Ben-Gurion University of the Negev
{noamt, cohavi, kirchenb}@bgu.ac.il

ABSTRACT

Using explicit (subjective evaluations) and implicit (response latency) measures, this study replicated and extended the findings by Fernandes et al (2003), who found that immediate aesthetic impressions of web pages are remarkably consistent. Forty participants evaluated 50 web pages in two phases. The degree to which web pages were regarded, on average, as attractive after a very short exposure of 0.5 sec. was highly correlated with attractiveness ratings after an exposure of 10 seconds. Extreme attractiveness evaluations (both positive and negative) were faster than moderate evaluations, providing convergent evidence to the hypothesis of immediate impression. Overall, the results provide direct evidence in support of the premise that aesthetic impression of the IT artifacts are formed quickly. Indirectly, the results suggest that visual aesthetics can play an important role in users' evaluations of the IT artifact.

Keywords

Attractiveness, aesthetics, web pages, response latency, human-computer interaction.

INTRODUCTION

First impressions color subsequent search for information and sway judgment and choice processes. One of the most notable sources of impression is the visual appearance of objects. In a seminal paper, Dion, Berscheid and Walster (1972), demonstrated that a person's physical appearance influences other aspects of the social interaction. Beautiful people earn more (Hamermesh and Biddle, 1994), and receive higher teaching evaluations (Hamermesh and Parker, 2003). People are affected by the aesthetics of nature or of artifacts (Nasar, 1988; Norman, 2004; Porteous, 1996; Postrel, 2002).

One of the reasons for influence of aesthetics on our judgment may stem from its immediate effect on our senses and, consequently on our affective system. This is important because the affective system and the cognitive system are intertwined (Bargh, in press ; Ortony, Norman

and Revelle, in press; Russell, 2003). Thus, while previous research in MIS and in HCI largely presumed that human decision making relies entirely on cognitive processes, current research on decision making portrays a different picture. One of the key characteristics of the affective system is that some of its reactions are very rapid (Norman, 2004; Pham et al, 2001). Other affective responses often involve considerable cognitive mediation and are decidedly slower. Recent research suggests that first aesthetic impressions (as opposed to more reflective and cognitively driven aesthetic judgments) are affective and may be formed immediately and thus precede cognitive processes (Fernandes et al., 2003; Pham et al, 2001). Hence, the immediate affective reactions may color and potentially sway successive cognitive processes (Duckworth et al., 2002; Pham et al, 2001).

First impression of web pages

Aesthetics has been shown to influence users' perceptions of interactive systems in general (Tractinsky et al., 2000) and of web pages in particular (van der Heijden, 2003; Karvonen 2000; Kim et al., 2003; Lavie and Tractinsky, 2004; Schenkman and Jonsson, 2000; Zhang and von Dran, 2000; Zhang et al., 2001). However, it is not clear whether this influence stems from immediate, first impression, or from a more elaborated examination of the web site. The causal relationship between aesthetic perceptions and other attitudes towards the IT artifact is also not clear. For example, while Tractinsky et al (2000) suggest that "beautiful is usable," implying that aesthetic perceptions color other perceptions of the interaction, Hassenzahl (2004) raises the idea of "usable is beautiful," suggesting that a fluent interaction causes the system to be perceived as more beautiful. Studies that demonstrated the immediacy of certain affective reactions have mostly been based on simple stimuli (e.g., Duckworth et al, 2002). But the visual appearance of computer software, and of web pages in particular, is considerably more complex. Can we assume, then, that users form immediate, consistent aesthetic judgments of fairly complex stimuli such as web pages? Fernandes et al (2003), suggest that the answer is "Yes." In their study,

participants saw images of 100 web pages for 0.5 sec. in a random order. This exposure time, they reasoned, was long enough to form a first impression, yet not sufficiently long to evaluate other features of the web site, such as its semantic content. After each page was shown, the participants rated its visual attractiveness by using a continuous rating scale, ranging from 0 (for very unattractive web pages) to 100 (for very attractive web pages). Then, each participant viewed the 100 pages for a second time in a newly randomized order. The correlation between the mean evaluation of the visual attractiveness of web pages in the first phase and the mean evaluations in the second phase was .97, indicating that, when aggregating individual evaluations, even very short exposure resulted in remarkably consistent aesthetic evaluation.

Study Objectives

The objectives of this study are threefold:

1. Replicate Fernandes et al's (2003) findings using a different set of web pages, a different rating scale and with participants from a different culture (Israelis instead of Canadians).
2. Validate the results using converging evidence from an implicit measure (response latency) in addition to the use of an explicit measure (subjective rating). We elaborate on the measure of response latency in the next subsection.
3. Study two types of evaluation consistency. One type refers to the consistency by which ratings of web pages are averaged over users in the two rounds of evaluations. Monk (2004) suggests that for HCI designers, the interesting question is whether products are rated consistently higher or lower relative to other products. This type of consistency was demonstrated by Fernandes et al (2003). However, it is also of interest to study a second type of consistency, one that gauges the degree to which individuals are *internally* consistent in evaluating aesthetic stimuli (e.g., Hassenzahl, 2004) and in particular stimuli that were presented for a very short duration. For lack of space, the analysis of the internal consistency is not presented here. It is available from the authors.

Response latency

Fernandes et al (2003) used subjective measures to study whether very short exposure can elicit aesthetic response. In this study we add another measure, response latency, to validate their results. Response latency is the length of time taken by a respondent to answer a question. As a measure, response latency has several general advantages: It is unobtrusive and is very easy to collect over computerized systems. In addition, there are three specific reasons for our interest in this measure.

1. Response latency is an obvious and natural selection for a measure of immediate reactions.
2. Response latency has been demonstrated to be a measure of strength of preferences (e.g., Aaker et al,

1980; MacLachlan et al., 1979; Tyebjee, 1979). Most studies employing response latency measures to infer preferences have used binary choice tasks in which the psychological distance between the two stimuli was manipulated and its effects on response latencies observed. Because of the large space required to present web pages for users' evaluations, our study is limited to registering latencies from judgments of single stimuli (i.e., a web page). Relatively little research has been done using response latencies in this mode. However, it is reasonable to expect that the more extreme evaluations of web page attractiveness will be associated with shorter response latencies. For example, Pham et al (2001) found that extreme ratings of pictures – whether positive or negative – were associated with lower latencies than were more moderate ratings. Their results suggest an inverted U shape relationship between response latencies and rating extremity. Similarly, Bassili (1996) reports an inverted U-curved pattern of the relation between response latency and extremity of opinion.

3. Finally, Aaker et al.'s (1980) suggest that measuring response latency can be helpful in assessing the construct validity of preference measurements. As a very different measurement method relative to explicit ratings of stimuli, it can be used to test for convergence across different measures of the same "thing" (Cook and Campbell, 1979), given that, as mentioned above, we expect it to covary with rating extremity.

METHOD

Sample. 40 students (25 female, 15 male) volunteered to participate in the study for course credit. They were 19-28 years old (average = 23.7).

Stimuli. 50 web pages were selected for this experiment based on two criteria: (1) They did not belong to well known web sites (to reduce the possible influence of familiarity on evaluations). (2) We considered 25 of them to be relatively attractive and the other 25 to be relatively unattractive. The web pages came from a large variety of domains. Screen shots of the web pages were captured at a resolution of 1024 X 768 pixels in 24-bit true color, but were compressed to JPG format with a resolution of 800X600 before being presented in the experiment.

Procedure. Participants were briefed about the study's general purpose and were given written instructions regarding the experimental task. In addition, each experimental phase was preceded by online instructions.

The participants interacted with a computer system that included a P4 1.7MHz processor and a 19 inch display. A C# program in .NET environment was built to control the procedure, to present images of web sites, to control the display time of the images and to collect user data, including ratings of the web site images and response latencies. The study consisted of two main phases. In Phase 1, each web page image was displayed for 500 ms, after which the rating scale was displayed on the screen

and the participants could rate the page which they just saw. The rating scale was represented by 10 radio buttons arranged in order from left to right. A "Very Unattractive" and a "Very Attractive" verbal anchors were placed below the "1" button and the "10" button respectively. There were no instructions or time limits regarding speed of rating.¹ Before the experimental stimuli, a block of 10 trial images was administered to get the participants acquainted with the rating method and the short display times. Next, the 50 web page images were presented in a random order; participants rated each image in turn, and pressed a "Continue" button when they were ready to proceed to the next image. After that phase, an instructions page informed the participants that they are done with the first phase and that the second phase is about to begin. The stimuli and the procedure in Phase 2 were identical to those of Phase 1, with the exception that images were presented for 10 seconds rather than for 0.5 sec. The order of presentation of web page images was again randomized for each participant.

RESULTS

Overall, 2000 attractiveness ratings and response latencies were collected (40 participants X 50 web pages). The distribution of attractiveness evaluations suggests a quasi-normal distribution where more evaluations concentrate at the middle of the scale and fewer pages are evaluated as extremely attractive or unattractive. Also, extreme positive ratings were rarer than extreme negative ratings.

Attractiveness evaluations of web pages

Evaluations were somewhat more favorable in Phase 2 (mean = 5.28, SD = 1.31) compared to Phase 1 (mean = 5.05, SD = 1.25). In both cases, the average ratings were just slightly below the middle of the rating scale, indicating that the set of web pages chosen for this study was quite balanced in terms of the pages' attractiveness. In Phase 1, the average attractiveness of the web pages ranged from 2.5 for the least attractive page to 7.98 for the most attractive page. In Phase 2 the attractiveness evaluations ranged from 3.3 to 8.3. For each participant, the ratings of the web pages in each phase were transformed into z-scores to control for individual rating tendencies. The means of the raw scores and of the z-scores for each web page were calculated separately for each of the two experimental phases. The correlation of the mean z-scores of the 50 pages between the two phases was .92. It was almost identical to the correlation between the mean raw rating of visual appeal for each web page on both phases ($r = .92$). The relation between the raw ratings in both phases is depicted in Figure 1. The high explained variance ($R^2 = 0.85$) indicates that even with minimal

exposure (i.e., only 0.5 sec.), the evaluation of web page attractiveness (averaged over participants) was very consistent. These findings replicate the results of Fernandes et al (2003) despite the use of different web pages, a different sample from a different culture, and a slightly modified methodology (e.g., recording participants' ratings using 10 radio buttons rather than a slider indicating a range of 0 to 100).

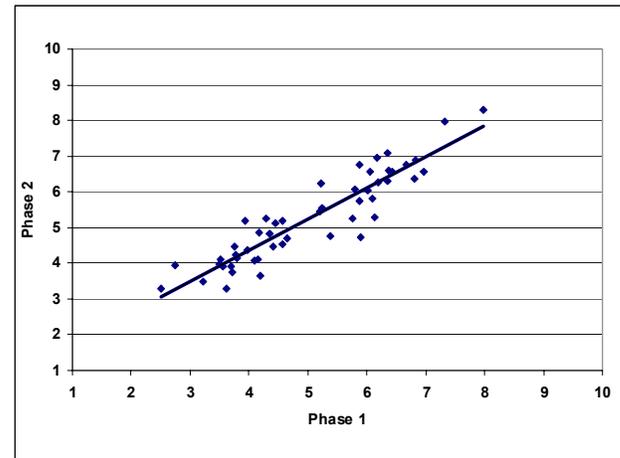


Figure 1. Average raw rating for each web page in Phases 1 and 2 (each dot reflects the mean rating over 40 participants of each of 50 web pages)

Relation between rating extremity and response latency

The mean and the median latencies of very attractive or very unattractive web pages were shorter than latencies of ratings that were placed at the middle of the scale. In other words, with as little exposure as 0.5 seconds, users respond more quickly to very attractive or very unattractive web pages than to web pages that are more ordinary beauty-wise.

To test the relation between extremity of attractiveness rating and reaction times, an ANOVA was performed with ratings as random factors and the transformed latencies as a dependent variable. Ratings were treated as random factors with 5 levels based on their distance from the scale's mid-points. For example, ratings of 5 and 6 – which are the scale's mid-point – belong to Category 0, whereas ratings of 1 and 10 – which are the most extreme ratings – belong to category 4. There was a significant effect of extremity of the rating on response latency ($F(4, 1995) = 10.815, p < .001$). Table 1 displays the post-hoc comparisons between response latencies of the five pooled rating categories. The comparisons support the premise that the more extreme the rating, the shorter the response latency. The only exception is the lack of significant difference in response latencies between Categories 1 and 2.

¹ Practice differs regarding the use of speed instructions between the preference measurement paradigm (no such instructions) and the implicit attitude measurement paradigm. In both cases, though, the results are quite consistent regarding the relation between response latency and attitude or preference strength.

Rating Category	Vs. Cat 3	Vs. Cat. 2	Vs. Cat. 1	Vs. Cat. 0
4 (ratings of 1 and 10)	*	***	***	***
3 (ratings of 2 and 9)	-	*	*	***
2 (ratings of 3 and 8)		-	ns	**
1 (ratings of 4 and 7)			-	**
0 (ratings of 5 and 6)				-

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

Table 1. Pairwise (LSD) post-hoc comparisons between transformed (ln) latencies of web page ratings. Categories 0 through 4 represent the extremity of evaluations of web page attractiveness

DISCUSSION AND CONCLUSION

Our findings demonstrate that users are able to form immediate and consistent evaluation of the attractiveness of web pages. These evaluations were very consistent across web pages. That is, the degree to which web pages were regarded, on average, as attractive after a very short exposure remained stable given a considerably longer exposure, lending support to the proposition that the relative attractiveness of web pages is determined quickly.

Convergence between the explicit rating measure and the implicit response-latency measure was demonstrated for both attractive and unattractive evaluations, in line with similar findings from different contexts (e.g., Bassili, 1996; Ostrom and Gannon, 1996; Pham et al., 2001). In addition to the support these findings lend to the proposition that evaluations of web page aesthetics are immediate, it also demonstrates the potential of response-latency as an easy-to-collect, unobtrusive measure of preferences and attitudes in MIS and HCI research.

Some promising work has already been done on the question of what design characteristics affect evaluations of web pages (Fernandes et al, 2003; Kim and Moon, 1998; Kim et al., 2003; Lavie and Tractinsky, 2004). There is ample room for future research to elaborate on this question and on related issues. For example, cultural and individual differences may play an important role in moderating such effects.

While the findings of this study relate to a presumably minor issue of web site design, the implications may be far reaching. Recent studies have argued that positive affect improves decision making, trust and social interactions (e.g., Isen (2001). Other studies found that decisions consist of a mix of conscious and nonconscious processes (Bargh, in press) and that the degree to which nonconscious processes influence choice processes "is much greater than most choice researchers believe" (Fitzsimons et al., 2002). Studies have also shown that one's affective states are related to aesthetics of one's

environment, be it in the working place (Rafaeli and Vilnai-Yavetz, in press), the home or the neighborhood (Nasar, 1988), the store (Russell and Pratt, 1980), or the web site (Kim et al., 2003; Zhang and Li, 2004). The findings of this study point to the potential influence of aesthetics on subsequent attitudes towards the IT artifact. Since aesthetic information is evaluated immediately, it is largely responsible for the users' first impressions. Subsequently, new information tends to be processed in a way that is biased towards those first impressions (Fitzsimons et al., 2002). This may help explain why the visual appeal of IT has been found to strongly affect overall impressions of the interaction and color other aspects of the IT artifact (van der Heijden, 2003; Lindgaard and Dudek, 2003; Schenkman and Jonsson, 2000; Tractinsky et al., 2000). Taken together, these studies attest to the important role of aesthetics in human-computer interaction.

REFERENCES

1. Aaker, D.A., Bagozzi, R.P., Carman, J.M., MacLachlan, J.M. (1980) On using response latency to measure preference, *Journal of Marketing Research*, May 1980, 17, 237-244.
2. Bargh, J.A. (in press) Bypassing the Will: Towards demystifying the nonconscious control of social behavior. In Hassin, R. R., Uleman, J. S., & Bargh, J. A. (Eds.) *The New Unconscious*. New York: Oxford University Press.
3. Bassili, J.N. (1996) The how and why of response latency measurement in telephone surveys. In N. Schwarz and S. Sudman (eds.) *Answering Questions*. San Francisco: Jossey-Bass Publishers.
4. Cook, T.D. and Campbell, D.T. (1979) *Quasi-Experimentation*, Boston:Houghton Mifflin Co.
5. Dion K., Berscheid E. and Walster E. (1972), What is beautiful is good, *Journal of Personality and Social Psychology*, vol. 24, NO. 3, 285-290.
6. Duckworth, K.L., Bargh, J.A., Garcia, M., and Chaiken, S. (2002) The automatic evaluation of novel stimuli, *Psychological Science*, 13(6), 513-519.
7. Fernandes, G., Lindgaard, G., Dillon, R. and Wood, J. (2003) Judging the appeal of web sites, *Proceedings 4th World Congress on the Management of Electronic Commerce*, McMaster University, Hamilton, ON, 15-17 January, 2003.
8. Fitzsimons G.J., Hutchinson J.W., Williams P., Alba J.W., Chartrand T.L., Huber J., Kardes F.R., Menon G., Raghurir P., Russo J.E., Shiv B., and Tavassoli N.T. (2002) Non-conscious influences on consumer choice, *Marketing Letters*, August 2002, vol. 13(3), 269-279.
9. Hamermesh, D. and Biddle, J. (1994) Beauty and the labor market," *American Economic Review*, 84 (December): 1174-94.

10. Hamermesh, D. and Parker, A.M. (2003) Beauty in the classroom: Professors' pulchritude and putative pedagogical productivity," NBER Working Papers 9853, National Bureau of Economic Research, Inc.
11. Hassenzahl, M. (2004) Things that make me shine: the beauty in interactive products, *Human-Computer Interaction*, 19(4), (in press).
12. van der Heijden, H. (2003) Factors influencing the usage of websites: the case of a generic portal in the Netherlands. *Information and Management*, 40, 541–549.
13. Isen, A.M. (2001) An influence of positive affect on decision making in complex situations: Theoretical issues with practical implications, *Journal of Consumer Psychology*, 11(2), 75–85.
14. Karvonen, K. (2000) The beauty of simplicity. *Proceedings of the ACM Conference on Universal Usability (CUU 2000)*, November 16-17, 2000, Washington D.C., USA
15. Kim, J., Lee, J., and Choi, D. (2003) Designing emotionally evocative homepages: an empirical study of the quantitative relations between design factors and emotional dimensions, *International Journal of Human-Computer Studies*, 59(6), 899–940.
16. Kim, J. and Moon, J.Y. (1998) Designing towards emotional usability in customer interfaces -- trustworthiness of cyber-banking system interfaces, *Interacting with Computers*, 10 (1), 1-29.
17. Lavie, T. and Tractinsky, N. (2004) Assessing dimensions of perceived visual aesthetics of Web sites, *International Journal of Human-Computer Studies*, 60, 269-298.
18. Lindgaard, G. and Dudek, C. (2003) What is this evasive beast we call user satisfaction? *Interacting with Computers*, 15: 429-452.
19. MacLachlan, J., Czepiel, J., and LaBarbera, P. (1979) Implementation of response latency measures, *Journal of Marketing Research*, Nov 1979, 16, 573-577
20. Monk (2004) The product as a fixed effect fallacy, *Human-Computer Interaction*, 19(4) (in press).
21. Nasar, J.L. (ed.) (1988) *Environmental Aesthetics: Theory, Research, and Applications*, Cambridge University Press.
22. Norman, D. A. (2004) *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic Books.
23. Ortony, A., Norman, D. A., & Revelle, W. (In press). The role of affect and proto-affect in effective functioning. In J.-M. Fellous & M. A. Arbib (Eds.), *Who Needs Emotions? The Brain Meets the Machine*. New York: Oxford University Press.
24. Ostrom and Gannon (1996) Exemplar generation: Assessing how respondents give meaning to rating scales. In N. Schwarz and S. Sudman (eds.) *Answering Questions*. San Francisco: Jossey-Bass Publishers.
25. Pham, M.T., Cohen, J.B., Pracejus, J.W., Hughes, G.D. (2001) Affect monitoring and the primacy of feelings in judgment, *Journal of Consumer Research*, 28, 167-188.
26. Porteous, J.D., (1996). *Environmental Aesthetics: Ideas, Politics and Planning*, Routledge, London.
27. Postrel, V. (2002) *The Substance of Style*. Harper Collins.
28. Rafaeli, A. and Vilnai-Yavetz, I. Emotion as a connection of physical artifacts and organizations, *Organization Science* (in press).
29. Russell, J.A. (2003) Core affect and the psychological construction of emotion, *Psychological Review*, 110(1), 145-172.
30. Russell, J.A., and Pratt, G. (1980) A description of the affective quality attributed to environments, *Journal of Personality and Social Psychology*, 38(2), 311-322.
31. Schenkman, B.N., Jonsson, F.U. (2000) Aesthetics and preferences of web pages, *Behavior and Information Technology*, 19(5), 367-377.
32. Tractinsky, N., Shoval-Katz, A. Ikar, D., (2000) What is beautiful is usable, *Interacting with Computers*, 13, 127-145.
33. Tyebjee, T.T. (1979) Response latency: A new measure for scaling brand preference, *Journal of Marketing Research*, Feb 1979, 16, 96-101.
34. Zhang, P. and von Dran, G. M. (2000) Satisfiers and dissatisfiers: A two-factor model for website design and evaluation, *Journal of the American Society for Information Science* 51(14):1253–1268.
35. Zhang, P., von Dran, G., Blake, P. and Pipithsuksunt, P. (2001) Important design features in different website domains: An empirical study of user perceptions, *e-Service Journal*, Vol. 1, no. 1.
36. Zhang, P. & Li, N. (2004) The importance of affective quality, *Communications of the Association for Computing Machinery* (forthcoming).