Changing Minds: Evidence of Social Conformity in Evaluations of Web Design

Emergent Research Forum Paper

Daniel S. Soper
Mihaylo College of Business and Economics
California State University, Fullerton
dsoper@fullerton.edu

Arghavan Salehian
Neudesic, LLC
arghavan.salehian@gmail.com

Abstract

Despite the large array of behavioral phenomena that have been linked to the desire for social conformity, the willingness to engage in such behavior in the context of web interface design evaluations has attracted very little attention in information systems research. This paper therefore reports on the progress of a study aimed at investigating the role of *ex post facto* social influence on evaluations of the quality of a web interface design. Using a controlled, randomized experiment involving 854 subjects, three different web interfaces, and five interface design characteristics, the preliminary findings reveal that information provided to subjects about the opinions of others can cause those subjects to revise their own initial web interface evaluations, such that their revised ratings more closely conform with those of the group. The observed effects are particularly pronounced among younger subjects and among subjects whose initial ratings differ widely from those of the group.

Keywords

Social conformity, web interface design, interface evaluation

Changing Minds: Evidence of Social Conformity in Evaluations of Web Design

Emergent Research Forum Paper

Introduction and Research Questions

The desire to conform to the values, beliefs, and attitudes of the group exerts a powerful, albeit unconscious influence on human behavior, even when the group in question is comprised of complete strangers (Asch 1951; Asch 1956; Banerjee 1992). The theory of normative social influence suggests that this desire to conform evolved in humans in response to our need to successfully live and work together in order to maximize our chances of survival and prosperity (Aronson et al. 2005). Despite the large array of behavioral phenomena that have been linked to the desire for social conformity, the willingness of humans to engage in such behavior in the context of web interface design evaluations has attracted very little attention in the information systems (IS) research literature. One notable exception is the work of Soper (2015), in which it was demonstrated that *a priori* knowledge of the opinions of others exerts a strong and highly significant influence on web interface design assessments. Nevertheless, whether and to what extent people are willing to revise their initial evaluations of a web interface design *after* being made aware of the collective opinion of the group remains unknown. For this reason, we would like to report in this emergent research forum paper on the progress of a large study aimed at investigating the role of *a posteriori* social influence on people's evaluations of the quality of a web interface design. In its current preliminary phase, the study seeks to provide insights into the following three research questions:

1. To what extent are we willing to revise our web interface design ratings after being made aware of the opinions of others?
2. Which factors (if any) influence our willingness to revise our web interface design ratings after being made aware of the opinions of others?
3. As the perceived degree of similarity between ourselves and the members of a reference group increases, are we more willing to revise our web interface design ratings such that they conform to the collective opinion of the group?

Although this study is certainly still a work in progress, we nevertheless hope that the reader will find the brief report provided herein to be of interest.

Research Design and Methodology

A controlled, randomized experiment was conducted in order to gain insights into the research questions developed above. Given that the target population for the experiment was adult, English-speaking web users, the leading global online advertising firm was contracted to carry out a targeted campaign for the purpose of soliciting volunteers for the study. The firm's demographic targeting technologies allowed subject recruitment to be explicitly constrained to English-speaking web users who were at least 18 years old. IP address restrictions were also employed to help ensure that subjects could participate only once in the experiment. After agreeing to participate, subjects were asked to specify their gender and age, and were then iteratively assigned into one of five experimental groups. Gender and age were explicitly included in the study because they have been identified by past research as the primary features by which people unconsciously judge the degree of similarity between themselves and others (Brewer and Lui 1989). In total, data were gathered from 854 subjects, of whom 433 (50.7%) were female and 421 (49.3%) were male. Subjects ranged in age from 18 to 80 years, with the mean age being 32.79 years (std dev = 11.93), signifying an age distribution skewed in the direction of youth. These demographic characteristics were observed to be consistent with the overall population of adult web users (Pew Research Center 2014).

A custom, web-based software system was used to conduct the experiment. As their primary task, subjects were required to evaluate the design characteristics of three different web interfaces, each of which intentionally followed the general mental model of web interface design identified by Soper and Mitra

(2013). The specific interface design characteristics that subjects evaluated were adopted from a pre-validated, five-item subscale designed to assess web interface quality (Aladwania and Palvia 2002). In accordance with the original instrument, subjects were asked to respond to the five evaluative statements using a seven-point Likert-type scale anchored at 1 = *strongly disagree* and 7 = *strongly agree*. Minor modifications were made to the wording of the items in order to accomodate the current experiment.

Original Statement (Aladwania and Palvia 2002)	Modified Statement Used in Current Experiment
____'s website looks attractive.	This website looks attractive.
____'s website looks organized.	This website looks organized.
____'s website uses fonts properly.	This website uses fonts properly.
____'s website uses colors properly.	This website uses colors properly.
____'s website uses multimedia features properly.	This website uses multimedia features properly.

Table 1. Original and modified subscale items.

Consistent with the recommendations of Soper (2014), each subject evaluated all three web interfaces along just one of the dimensions listed in Table 1 in order to minimize the possibility that the subject's ratings would be contaminated by halo error. Iterative assignment was used in order to determine the specific design characteristic that each subject was asked to evaluate, and the order in which the three web interfaces were presented to each subject was randomized to mitigate the possibility of ordering or self-generated validity effects (Chandon et al. 2005; Saris and Gallhofer 2007).

As noted above, subjects were iteratively assigned into one of five groups, including a baseline group and four experimentally manipulated "treatment" groups. Subjects in the baseline group were simply shown the three web interfaces and asked to rate each interface along their assigned dimension. When aggregated, the responses from subjects in the baseline group were regarded as the true, unadulterated ratings for each web interface design characteristic, and served as the basis of comparison for subjects in the treatment groups. The rating tasks and experimental process for subjects in the four treatment groups were identical to those of the baseline group, except that treatment group subjects were given an opportunity to change their minds and revise their ratings, thus allowing the study's research questions to be evaluated. More specifically, after having rated a web interface along a particular dimension, subjects in each treatment group were provided with experimentally manipulated information about how other people rated the same design characteristic, and were then asked if they would like to revise their initial rating. Subjects who expressed a desire to revise their initial ratings were granted the opportunity to do so, after which their revised rating was recorded.

Importantly, the only difference among the four treatment groups was the degree of similarity between these "other people" and the subject herself. As a clarifying example, imagine that a 31-year-old female subject had been asked to rate the extent to which a website uses fonts properly. If the subject had been assigned to the first treatment group, she would be provided with a generic statement such as "The average response given by other people for this question is 2.59 out of 7.00". By contrast, if the same subject had been assigned to the fourth treatment group, she would be provided with an age- and gender-specific statement such as "The average response given by other *31-year-old women* for this question is 2.59 out of 7.00" (emphasis added). A more complete illustration of the research design is provided in Figure 1 below.

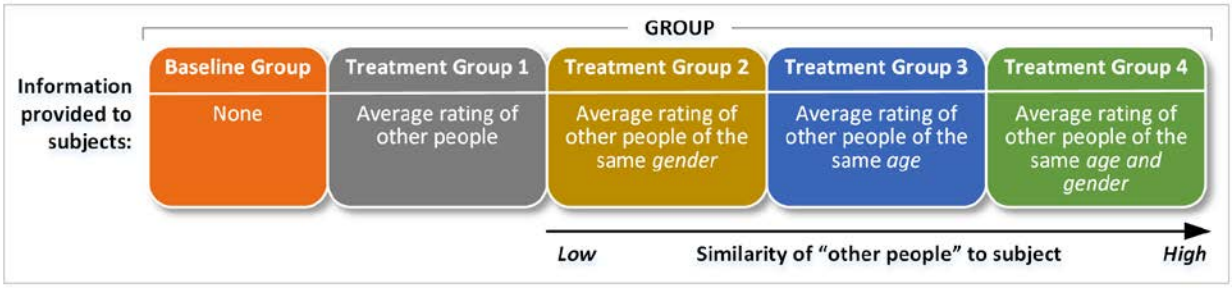

Figure 1. Research design.

Past research has concluded that in the absence of specific knowledge, gender and age are the primary unconscious cues that human beings use when judging how similar they are to others (Brewer and Lui

1989). The figure above thus illustrates how a subject's perceived degree of similarity between herself and the people with whose average rating she has been supplied increases as one proceeds from Treatment Group 2 to Treatment Group 4. Intuitively, approximately 50% of a large group of randomly chosen people would share the subject's gender (Treatment Group 2), while a much smaller percentage would share the subject's age (Treatment Group 3). The least likely combination of all, of course, would be to randomly choose people who share both the subject's age *and* her gender (Treatment Group 4), and in the absence of other information, it is with these people that the subject can be expected to most closely identify.

With five different groups, three different web interfaces, and five different interface characteristics, a total of 75 different configurations were possible. For purposes of statistical validity, a minimum of 30 responses were required for each possible configuration, thus bringing the total minimum number of responses to 2,250, or 450 per group. Since each subject provided three responses, the preliminary minimum sample size was determined to be 150 subjects per group. Given that the linear models (discussed below) that would be used to evaluate the study's research questions would contain a maximum of 12 predictors, a formal *a priori* sample size analysis revealed that for an anticipated medium effect size (f^2) of 0.15 and a statistical power level of 0.80, a minimum of 127 subjects would be required per group (Cohen 1988; Soper 2016). In light of these analyses, the more conservative minimum sample size of 150 subjects per group was adopted for the experiment. The final dataset contained 2,562 web interface ratings from 854 subjects, distributed by group according to the values in Table 2 below.

Group	Subjects	Responses
Baseline group	150	450
Treatment group 1 (other people)	176	528
Treatment group 2 (gender)	176	528
Treatment group 3 (age)	176	528
Treatment group 4 (age and gender)	176	528
Total:	**854**	**2,562**

Table 2. Distribution of subjects and responses by research group.

As noted previously, subjects in the treatment groups were supplied with the average rating of other people for the interface design characteristic that they were currently considering. These ratings, however, were not the true ratings given by others, but instead were experimentally manipulated with a view toward gaining insights into the study's research questions. Specifically, the artificial ratings supplied to subjects in the treatment groups were statistically derived from the distributions of the true, unadulterated ratings obtained from the baseline group. To be more precise, the baseline mean rating and standard deviation for each combination of web interface and interface design characteristic were used to compute an artificial score that would be supplied to subjects in the treatment groups, with that artificial score being the value associated with a cumulative probability of 0.05 on the associated baseline rating's normal distribution. For example, the true rating obtained from the baseline group for the extent to which the third web interface used fonts properly was 5.70 (on a 1 to 7 scale), with a standard deviation of 1.32. Applying the cumulative distribution function, we can readily determine that 95% of subjects would naturally rate the usage of fonts on this interface at 3.53 or above, while only 5% of subjects would supply a rating lower than 3.53. In this case, subjects in the treatment groups would be told that the artificially low score of 3.53 was the average rating given by other people when evaluating font usage on that interface. Using this approach, it would be statistically improbable ($p < 0.05$) for a subject in the treatment groups to naturally assign such a low rating to the interface design characteristic that she was evaluating. Any statistically significant differences in the ratings given by subjects in the baseline group and subjects in the treatment groups who chose to revise their initial ratings could thus be attributed to social conformity behavior.

In light of the data described above, preliminary insights into the study's research questions were gained through correlation analysis and through the estimation of four linear models, each of which evaluated the extent to which subjects in the four treatment groups were willing to revise their initial ratings. Each of the four linear models was specified such that a subject's willingness to revise his or her ratings was predicted by the treatment group to which the subject belonged, after controlling for the distance between the subject's initial ratings and those of the group, the subject's age and gender, and the interface and interface design characteristic being evaluated by the subject. For this purpose, treatment group membership, subject

gender, and the various interfaces and design characteristics were all appropriately coded using a series of binary dummy variables. The results of the correlation analysis and the four linear regression analyses are presented and discussed in the following section.

Preliminary Results and Discussion

A total of 2,112 web interface ratings were collectively obtained from the four treatment groups, and among these, 1,687 were separated by more than one full point (on a 1 to 7 scale) from the associated average group rating, thus making them reasonably subject to revision for purposes of social conformity. Of these, a total of 220 ratings (13.0%) were voluntarily revised after subjects were provided with the associated opinion of the group. This result provides insight into the study's first research question, and reveals that in approximately 87% of the rating tasks, subjects chose to stand by their initial ratings rather than bowing to social pressure.

The study's second research question inquired into the factors that influence one's willingness to revise his or her web interface design ratings after being made aware of the opinions of others. To gain preliminary insights into this question, a correlation analysis was conducted among the responses obtained from subjects in the four treatment groups. The results of this analysis revealed that a subject's willingness to revise her ratings depends significantly on the distance between her initial rating and the opinion of the group (r = 0.174, $p < 0.001$), her age (r = -0.062, $p < 0.01$), and whether she was told the group opinion originated from "other people" (r = 0.068, $p < 0.01$) or from people of the same age and gender as herself (r = -0.057, $p < 0.01$). Among these significant correlations, the distance between a subject's initial rating and the opinion of the group exhibited the strongest association with a willingness to revise one's initial rating, with the direction of the correlation coefficient indicating that an initial rating that is very different from the group average is much more likely to be revised than an initial rating that is closer to the group average. This, we believe, provides compelling preliminary evidence of a desire for social conformity in assessments of web interface design. A subject's age was also observed to be associated with a willingness to revise one's initial rating, with the direction of the correlation coefficient indicating that younger people are more likely to revise their ratings in an effort to conform to the group than are older people. It is interesting to note that despite popular clichés about the obstinacy of men and the greater need among women for social participation, gender was not observed to be related to one's willingness to conform to the group by revising his or her ratings. It is also interesting to note that the collective opinions of "other people" were found to exert a positive influence on willingness to revise one's ratings, while the opinions of others of the same age and gender were found to discourage people from revising their ratings.

Initial estimation of the four linear regression models described in the previous section revealed that neither a subject's gender nor the interface or interface characteristic that the subject was evaluating significantly affected his or her willingness to revise his or her initial ratings. These control variables were thus removed as predictors, and the four linear models were then duly reestimated. After controlling for the effects of a subject's age and the distance between her initial rating and the opinion of the group, the artificially manipulated information regarding the opinions of others was found to exert a highly significant impact on a subject's willingness to revise her ratings in the "other people" treatment group ($p < 0.001$), but not in the other three treatment groups. These results are summarized in Table 3 below.

Group	Parameter estimate (β)	Percentage of initial ratings which were revised	t	Probability
Treatment group 1 (other people)	0.074	14.02%	5.438	< 0.001
Treatment group 2 (gender)	0.024	10.23%	1.799	0.072
Treatment group 3 (age)	0.017	10.04%	1.245	0.213
Treatment group 4 (age and gender)	-0.018	7.39%	-1.320	0.187

Table 3. Effects of *a posteriori* knowledge of group opinions on willingness to revise web interface design ratings.

The study's third research question inquired into the extent to which the perceived degree of similarity between a subject and the members of a reference group influences the subject's willingness to revise his or her web interface design ratings such that they better conform to the collective opinion of the group. As

shown in the table above, only members of Treatment Group 1 – who were provided with the collective opinion of "other people" – exhibited a statistically significant willingness to revise their initial web interface ratings after having been provided with the associated average of their group. The perceived degree of similarity between a subject and the group to which the subject belongs (at least in terms of age, gender, or both) therefore does not seem to significantly influence the subject's willingness to alter his or her opinions about the quality of a web interface design after the subject has independently established those opinions. This stands in stark contrast to situations in which subjects know the group's opinion *before* providing their own ratings, insofar as such ratings have been found by past research to be highly influenced by the perceived degree of similarity between the subject and the group (Soper 2015).

Concluding Remarks and Future Research Directions

Despite being a work in progress, the findings reported herein have important implications. A website now commonly serves as an organization's public face, and the design of an organization's website has critical consequences for how the organization is perceived, thus influencing its prospects for success. Eliminating bias from web interface design assessments should therefore be of particular interest to managers who wish to align their organization's website with the needs and expectations of their users.

There is clearly much more to be learned about the role of social conformity in assessments of interface design, and several additional facets of this phenomenon will be explored as this project matures. The current study, for example, only attempted to discern whether the desire to conform would cause subjects to lower their web interface ratings. Can social pressure also cause ratings to increase? Although questions such as this remain to be answered, the desire to conform clearly exerts a powerful influence on assessments of user interface design, and it seems likely that social conformity also plays an important role in many other phenomena which lie at the intersection of technology and human behavior. It is hoped that the preliminary work reported here will serve as a point of embarkation for a long and fascinating stream of research in this area.

References

Aladwania, A.M., and Palvia, P.C. 2002. "Developing and Validating an Instrument for Measuring User-Perceived Web Quality," *Information & Management* (39:6), pp. 467-476.

Aronson, E., Wilson, T.D., and Akert, A.M. 2005. *Social Psychology (5th Ed.)*. Upper Saddle River, NJ: Prentice Hall.

Asch, S.E. 1951. "Effects of Group Pressure Upon the Modification and Distortion of Judgement," in *Groups, Leadership and Men,* H. Guetzkow (ed.). Pittsburgh, PA: Carnegie Press.

Banerjee, A.V. 1992. "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics* (107:3), pp. 797-817.

Brewer, M.B., and Lui, L.N. 1989. "The Primacy of Age and Sex in the Structure of Person Categories," *Social Cognition* (7:3), pp. 262-274.

Chandon, P., Morwitz, V.G., and Reinartz, W.J. 2005. "Do Intentions Really Predict Behavior? Self-Generated Validity Effects in Survey Research," *Journal of Marketing* (69:2), pp. 1-14.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences (2nd Edition)*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Pew Research Center. 2014. "Internet User Demographics," in *Pew Research Internet Project Survey*. Washington, DC: Pew Research.

Saris, W.E., and Gallhofer, I.N. 2007. *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Hoboken, NJ: John Wiley & Sons.

Soper, D.S. 2014. "User Interface Design and the Halo Effect: Some Preliminary Evidence," in: *Proceedings of the 20th Americas Conference on Information Systems (AMCIS)*. Savannah, GA.

Soper, D.S. 2015. "Herd Behavior in Assessments of Web Interface Design: Some Preliminary Evidence," *21st Americas Conference on Information Systems (AMCIS)*, Fajardo, Puerto Rico.

Soper, D.S. 2016. "A-Priori Sample Size Calculator for Multiple Regression [Software]." Retrieved 03 Feb 2016, from http://www.danielsoper.com/statcalc

Soper, D.S., and Mitra, S. 2013. "An Inquiry into Mental Models of Web Interface Design," in: *Proceedings of the 19th Americas Conference on Information Systems (AMCIS)*. Chicago, IL.