Informational Social Influence, Belief Perseverance, and Conservatism Bias in Web Interface Design Evaluations

DANIEL S. SOPER

Department of Information Systems and Decision Sciences, California State University, Fullerton, CA 92831, USA

e-mail: dsoper@fullerton.edu

ABSTRACT The ability to accurately measure the quality of a web interface design is critical if organizations hope to make effective decisions about their websites. Since websites are often the most publically visible face of an organization, website-related decisions can have profound impacts on the organization's prospects for success. Prior research involving web interface quality assessments has typically considered those judging an interface to be independent actors who are making decisions in a situation that is free of social influence. In real-world settings, however, website designs are frequently evaluated by groups, rather than isolated individuals. To that end, this article seeks to provide insights into the effects that group-related cognitive biases may have on web interface design assessments. Specifically, by using a controlled, randomized experiment involving more than 500 research subjects, it is shown that judges' web interface ratings can be easily manipulated by social influence phenomena, and that people assessing web interfaces in a social context are highly susceptible to informational social influence, belief perseverance, and conservatism bias. Given that these phenomena can all serve as significant sources of measurement error, and given the importance of being able to accurately measure the quality of a website design, the findings of the study suggest that organizational web interface design evaluations should be deliberately performed by individual judges in an environment that is free from social influence.

INDEX TERMS Belief perseverance, conservatism bias, informational social influence, interface evaluation, web interface design.

I. INTRODUCTION

A. BACKGROUND AND MOTIVATION

In his highly influential work on software usability, Nielsen identified seven different methods of evaluating software interfaces, among which five involve interfaces being evaluated by independent, isolated judges [34]. It is perhaps not surprising, then, that the majority of studies focusing on web interface design assessments have traditionally characterized the person performing the assessment (*i.e.*, the judge) as an independent actor who makes her judgments in an environment that is free of social pressure or influence. The problem with this characterization, of course, is that real-world web interface design assessments are often carried out in a social

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

context (*e.g.*, by web design teams [18], [61] or by focus groups [30], [35]) using methods that Nielsen refers to as *pluralistic walkthroughs* or *consistency inspections* [34]. Such group-based evaluations are especially common in situations involving large organizations where creating a high-quality web experience for users or customers is deemed to be of particular importance.

It is certainly true that the "isolated judge" paradigm has helped human-computer interaction (HCI) researchers to learn much about factors that influence people's perceptions of a web interface. For example, this approach has been used to establish that the success of a website depends on the layout, sequencing, and arrangement of the various elements that together comprise the interface [38], and that attractiveness [23], [32], [56], consistency [1], [3], [42], mental models [53], [54], and cognitive biases such as the halo

IEEE *Access*

D. S. Soper: Informational Social Influence, Belief Perseverance, and Conservatism Bias in Web Interface Design Evaluations
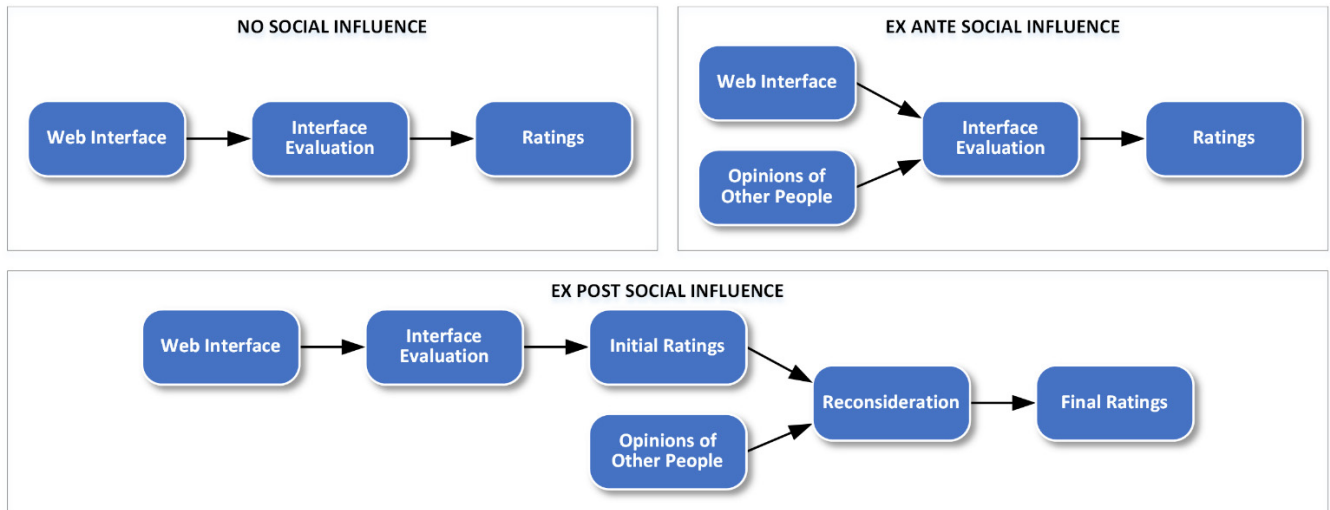
FIGURE 1. Different patterns of social influence in the web interface evaluation process. Existing research typically follows the "No Social Influence" pattern, which does not account for the potentially biasing effects of social phenomena.

effect [49], [55] all play important roles in shaping how a user perceives a web interface. Nevertheless, a certain disconnect remains between how web interfaces have traditionally been evaluated in the HCI literature and how such interfaces are often evaluated in real-world organizational settings.

Recognizing that groups are often involved in interface design and usability evaluations, several researchers have begun to develop techniques aimed at facilitating collaborative web interface assessments. Solano *et al.*, for example, have proposed methods and tools for collaboratively evaluating software interface usability [47], [48]. In none of these efforts, however, is consideration given to the influential effects that participating in a social-collaborative process may have on the opinions or perspectives of the individual judges. Many relevant phenomena have been identified in the social influence literature, but the potentially contaminating effects that these phenomena may have on interface design evaluations or broader software usability or user experience (UX) evaluations remains largely unknown. Consider, for example, that more than a quarter of a century has passed since the advent of the World Wide Web, and yet the HCI community still does not know whether or to what extent a person's opinions about the design of a web interface are influenced by knowledge of the opinions of others.

Intellectual curiosity notwithstanding, this situation is troubling for two interrelated reasons; namely, (1) because interactions with the Web are now an integral part of daily human life in much of the world, and (2) because websites now commonly serve as the most publically visible face of their underlying organizations. The extent to which a website is well-designed can hence directly influence an organization's prospects for success [54]. There are also many specific contexts such as medical decision-making and emergency response in which the quality of an interface design may quite literally have an impact on life or death decisions [22], [25], [26], [57]. It is for these reasons that

managers should be highly interested in ensuring that proposed designs for their organizations' websites are evaluated and measured as accurately and impartially as possible. In this spirit, the broad goal of the current paper is to investigate the role of *ex ante* and *ex post* social influence on judges' evaluations of the attractiveness of a web interface. This investigation is carried out using a controlled, randomized experiment in which subjects are asked to rate various interface designs. The current study can therefore be positioned as heuristic interface evaluation research, as originally defined by Nielsen and Molich [36].

B. PATTERNS OF SOCIAL INFLUENCE

To better understand and position the goal of the current paper, it is useful to consider the various patterns of social influence as they relate to web interface evaluations. Conceptually, *ex ante* and *ex post* social influence can be understood as informational cues to which a judge may be exposed at a particular point in time during the overall web interface evaluation process. These patterns of social influence are illustrated in Fig. 1 below, as is the pattern in which no social influence is present.

In Fig. 1, the "Interface Evaluation" and "Reconsideration" elements represent cognitive processes in which the judge considers the interface in question in light of her available information. Such information naturally includes the interface itself, as well as her past interface-related knowledge and experience (*i.e.*, her mental models of web interface design [53], [54]), which for purposes of clarity are not depicted in Fig. 1. Importantly for the current study, the judge may also be aware of what other people think about the interface, thus raising the specter of interference from social influence phenomena during her decision-making process. Within the context of the "Ex Ante Social Influence" pattern depicted above, the judge is aware of the opinions of others *before* she evaluates the interface [51]. By contrast, within

D. S. Soper: Informational Social Influence, Belief Perseverance, and Conservatism Bias in Web Interface Design Evaluations

IEEE *Access*

the "Ex Post Social Influence" pattern, the judge becomes aware of the opinions of others only *after* she has already evaluated the interface. In this scenario, the judge may subsequently reconsider her initial conclusions in light of the new information. In the context of the framework depicted in Fig. 1, almost all HCI research hitherto conducted in the area of web interface design evaluations belongs to the "No Social Influence" pattern, including all of the studies cited at the outset of this section.

C. RESEARCH QUESTIONS

The current study seeks to remedy several of the issues described above by extending knowledge about web interface design evaluations into the "Ex Ante Social Influence" and "Ex Post Social Influence" patterns shown in Fig. 1. Specifically, the current study relies on a controlled, randomized experiment involving three different web interfaces, five interface design characteristics, and more than 500 research subjects to provide insights into the following general research questions:

1. (*Ex Ante*) If a person is provided in advance with information about how other people evaluated a web interface, to what extent will that information influence her own ratings of the interface?
2. (*Ex Post*) If a person evaluates a web interface and is subsequently provided with information about how other people evaluated the same interface, to what extent will she be willing to revise her own initial ratings?

The balance of this article is organized as follows: the next section presents the theoretical framework upon which the study relies and develops the research hypotheses that are investigated herein. Section III describes the methods that were used to test the study's hypotheses and gain insights into the general research questions enumerated above. Section IV presents the results of the experiment and discusses the implications of the results for both scientists and practitioners alike. The paper concludes with Section V, which provides a brief summary, describes the limitations of the work, and offers a few final remarks.

II. THEORETICAL BACKGROUND AND RESEARCH HYPOTHESES

In this section, three key concepts that are critical to the current study are introduced and discussed. These three concepts – informational social influence, belief perseverance, and conservatism bias – originate from the cognitive and social psychology literatures, and collectively serve as both the theoretical foundation of the current study and the primary basis from which the study's research hypotheses are derived.

A. INFORMATIONAL SOCIAL INFLUENCE

Theory from the social psychology literature describes a common phenomenon known as *informational social influence* (or alternatively as *social proof*) in which a person

aligns her own beliefs, conclusions, or behaviors with those of the group [4], [10]. This phenomenon can exert a powerful, but often unconscious influence on human behavior, even when the group in question is comprised of complete strangers [5]–[7]. As a general principle, humans are social creatures who have an innate respect for and desire to belong to the group. Holding an unconscious regard for the group is believed to have helped our ancient ancestors survive in the unforgiving environment in which humans evolved, and is hence both adaptive and sensible from an evolutionary perspective [16]. Among modern humans, informational social influence continues to drive individuals to defer to the group opinion or behave according to group expectations.

This tendency of human beings to defer to the group can manifest itself under several conditions [17], including for purposes of simplifying decision-making, avoiding social ostracism, or both [4]. If, for example, a person is not confident in her own position or is faced with a difficult or ambiguous decision, she may defer to the group because she believes the group to have superior knowledge or a better understanding of the situation. A familiar and very modern example of this phenomenon can be found in the context of online shopping, wherein more than 80% of respondents to a large, recent survey indicated that they rely on product ratings when making a purchasing decision [46]. As this and countless other examples indicate, when a person is faced with a challenging task and is not entirely certain of or confident in her answer, she may rely on the opinion of the group as a basis for establishing her own position.

Informational social influence has been documented in a wide variety of real-world situations and contexts, and has been observed both across cultures and across time [4]. Further, while gender-based differences in susceptibility to social influence are by no means universal in the conformity literature, a minority of studies have observed the effects of group pressure to be significantly stronger among women [19], making gender a non-trivial consideration in any research examining informational social influence. In certain circumstances, age has also been found to be inversely related to social conformity in adults [39], hence making age another important consideration for the current study.

B. BELIEF PERSEVERANCE

In addition to informational social influence, the social psychology literature describes another behavioral phenomenon known as *belief perseverance* (or *belief perseverance error*) that may, in certain circumstances, act to counterbalance or even overwhelm the effects of social influence on a judge's decision-making processes. Belief perseverance has been defined as "…the tendency to cling to one's initial belief even after receiving new information that contradicts or disconfirms the basis of that belief [8]." As a practical example, consider that more than 97% of the climatological research on anthropogenic global warming (AGW) has concluded that AGW is a genuine and concerning phenomenon [15]. Despite this widespread scientific consensus, a disproportionately

IEEE Access

D. S. Soper: Informational Social Influence, Belief Perseverance, and Conservatism Bias in Web Interface Design Evaluations

large number of politicians, industrialists, and others exhibit belief perseverance by stubbornly clinging to the notion that global warming is an illusion or a scientific hoax. This example illustrates how after arriving at a position or a belief about a particular topic, many people are unwilling to revise their initial positions or beliefs in the face of new information, even when that information originates from an overwhelming group consensus. Refusing to revise one's views when presented with reliable and verifiable counterevidence is, of course, antithetical to the foundational principle of falsifiability in science [41], and as such, belief perseverance is orthogonal to the sort of behavior prescribed by the scientific method.

When considered together, the forces of informational social influence and belief perseverance can be conceptualized as exerting opposing psychological pressures on an individual's internal evaluation of his or her current beliefs. Although it is true that the status of human beings as social creatures imbues us with a natural esteem for the group consensus, it is also true that there are often psychological and social costs associated with admitting that one's beliefs are mistaken and need to be brought into alignment with those of the group. Both of these pressures, as manifested through informational social influence and belief perseverance, are at play when evaluating one's beliefs in the presence of new evidence. From a theoretical perspective, then, whether or not a person is willing to revise her beliefs hence depends in part on the magnitude of each of these pressures in a given situation. These concepts are metaphorically illustrated in Fig. 2 below.
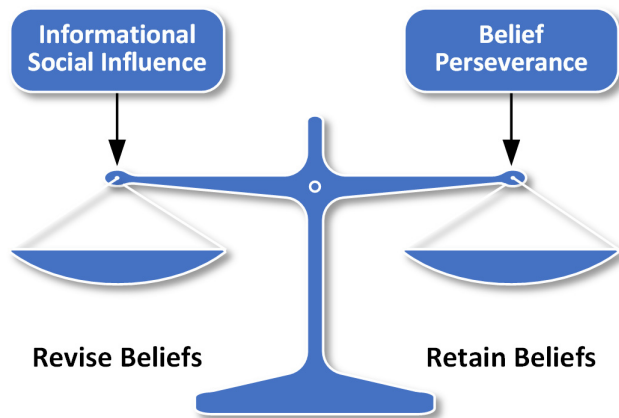

FIGURE 2. Informational social influence vs. belief perseverance in the decision to revise one's beliefs. These two phenomena are theorized to exert opposing cognitive pressures when a person evaluates a web interface in the context of social influence.

As shown in Fig. 2, when the immediate or cumulative pressure exerted on the psyche by informational social influence exceeds that of belief perseverance, then a person can be expected to revise her current beliefs. If, however, the pressure exerted on the psyche by belief perseverance exceeds that of informational social influence, then a person can be expected to retain her current beliefs. The magnitude of these opposing pressures depends, of course, on the nature

of the situation, and the predictions of informational social influence and belief perseverance theory must therefore be taken as situational; *i.e.*, whether and to what extent these phenomena manifest themselves will vary from person to person and from situation to situation [28].

C. CONSERVATISM BIAS

Having considered belief perseverance, we can next consider what happens when a person *does* revise her initial beliefs in light of new information. In this scenario, it has been observed that the degree to which an individual adjusts her beliefs is commonly less than would be expected in a Bayesian model of belief revision [20]. This phenomenon in which a person under-reacts to new information is known as *conservatism bias*. It is a specific type of belief perseverance error that is generally held to be an extension of Tversky and Kahneman's theory on anchoring and adjusting [58].

As a practical example of conservatism bias, consider a person who initially believes the value of her home to be €300,000. If home prices in her local area begin to decline and her neighbors inform her that the actual value of her home is now €200,000, she may be willing to lower her personal estimate of her home's value, but will be unlikely to revise her estimate all the way down to €200,000. Put more generally, a person's prior views and opinions often exert an unduly large influence on the extent to which she is willing to revise her beliefs in the face of new evidence. Conservatism bias has been identified and studied in a variety of contexts including decision science [20], investing [27], cognitive psychology [24], and economics [37].

D. RESEARCH HYPOTHESES

Together, the three theory-driven phenomena described above – informational social influence, belief perseverance, and conservatism bias – paint an interesting and nuanced portrait of how individuals can be expected to behave in the presence of social influence. With a solid understanding of these theoretical foundations, we are now equipped to return to the context of the current study and examine how these concepts apply to web interface evaluations in situations involving social influence.

To begin, consider the ''Ex Ante Social Influence'' pattern depicted in Fig. 1. When applied in the context of web interface evaluations, informational social influence theory predicts that if a person has prior knowledge of the opinions of others about a particular web interface, then those opinions can be expected to influence her own ratings of the web interface. Stated as a hypothesis, this becomes:

H1: *Ex ante* knowledge of the opinions of others significantly influences a judge's web interface design ratings.

If we next consider the ''Ex Post Social Influence'' pattern depicted in Fig. 1, the theoretical discussion above yields several additional predictions about how judges will behave when evaluating a web interface in the context of social influence. First, when provided on an *ex post* basis with other people's ratings of the quality of a web interface design,

D. S. Soper: Informational Social Influence, Belief Perseverance, and Conservatism Bias in Web Interface Design Evaluations

IEEE *Access*

belief perseverance suggests that many judges will stand by their initial ratings, even in situations where the evidence indicates that a judge's own ratings diverge markedly from those of the group. Put differently, and in comparison to the *ex ante* condition, once a judge has independently established her own beliefs about the quality of a web interface design, the opinion of the group will, on average, exert a weaker influence on the judge due to the countervailing effects of belief perseverance, particularly if there is no social penalty or risk of ostracism for holding a divergent point of view. This leads to the study's second hypothesis:

H2: *Ex post* knowledge of the opinions of others exerts a significantly weaker influence on judges' web interface design ratings than *ex ante* knowledge of the same opinions.

Although *ex post* knowledge of the group opinion can, on average, be expected to exert a weaker influence on web interface design evaluations than *ex ante* knowledge (due to the effects of belief perseverance), it is nevertheless important to consider the magnitude of the *ex post* social pressure to which the judge is subjected when making predictions about whether she will defer to the group opinion and subsequently revise her own beliefs. The theoretical considerations described previously suggest that when the force of informational social influence substantially exceeds that of belief perseverance, then a person may conclude that her beliefs are incorrect and need to be brought into closer alignment with those of the group. Thus, when a person is provided on an *ex post* basis with knowledge about the beliefs of the group, the extent to which the person is willing to revise her own beliefs can be expected to be positively related to the degree of misalignment between her initial opinions and the opinions of the group. Stated as a hypothesis, this becomes:

H3: When provided with *ex post* knowledge about the group's ratings of the quality of a web interface design, the extent to which a judge revises her own ratings is positively dependent on the degree of misalignment between her initial ratings and those of the group.

Finally, the discussion above about the theoretical predictions of conservatism bias provides further insights into the extent to which a judge is willing to revise her initial web interface design ratings after being provided with *ex post* knowledge about the corresponding ratings assigned by the group. Specifically, conservatism bias suggests that although a judge may revise her initial ratings, the extent of those revisions will, on average, be inadequate to fully close the gap between her own initial ratings and the ratings of the group. Put differently, the judge's prior views and opinions will exert an influence on the extent to which she is willing to revise her ratings, and the judge can hence be expected to under-react when provided with *ex post* knowledge about the ratings assigned to the web interface by the group. This leads to the study's fourth and final hypothesis:

H4: When a judge revises her ratings of the quality of a web interface design after being provided with *ex post* knowledge of the ratings assigned by the group, the magnitude of those revisions will, on average, be inadequate to fully close the gap between her own initial ratings and the ratings of the group.

III. METHODS

As noted briefly in Section I, the current study relied on human judges who were tasked with rating a variety of web interfaces. From a high-level methodological perspective, the study can thus be classified as heuristic interface evaluation research [36]. Although many different approaches have been used in interface evaluation research, empirical analysis has been identified as the leading method of evaluating user interfaces [34]. Further, among the various empirical methods, randomized controlled trials stand out as the gold standard due to their intrinsic ability to mitigate potential confounds among study participants [21]. For these reasons a controlled, randomized experiment was adopted as a basis for gaining insights into the current study's research hypotheses.

A. EXPERIMENT DESIGN AND SURVEY INSTRUMENT

The experiment used to test the study's research hypotheses relied on a three-group design, with each research subject being assigned to either a baseline (control) group, or to one of two experimentally manipulated treatment groups. The experiment itself was carried out using a custom, web-based software system that provided subjects with the ability to view and rate the characteristics of several different web interfaces. The specific characteristics for each interface were adopted from a five-item subscale that was created to measure the attractiveness of a web interface [2]. In the broader context of Nielsen's usability heuristics for user interface design [33], this focus on interface attractiveness maps to the *consistency* and *aesthetic* heuristics.

The specific survey instrument used in the study was chosen because it was subjected to rigorous validation procedures during its development, and because it has subsequently been very widely used in scientific studies examining web interface design. In accordance with the original instrument, subjects in the experiment were asked to respond to 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 original items in order to adapt those items to the context of the current experiment (see Table 1).

In total, three web interfaces were used in the experiment, with each interface being intentionally designed to align with the general mental model of web interface design identified by Soper and Mitra [53], [54]. Each subject was required to evaluate all three web interfaces along just one of the dimensions listed in Table 1 so as to minimize the possibility that the subject's ratings would be contaminated by halo error [55]. The specific design characteristic that each subject was asked to evaluate was determined using iterative assignment, and the order in which the three web interfaces were presented to each subject was randomized with a view toward mitigating any ordering or self-generated validity effects [9], [44].

IEEE Access

D. S. Soper: Informational Social Influence, Belief Perseverance, and Conservatism Bias in Web Interface Design Evaluations

TABLE 1. Original Subscale Items and Modified Subscale Items as Used in the Experiment.

Original Statement [2]	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.

B. RESEARCH SUBJECTS

The target population for the experiment was English-speaking, adult web users. The population was restricted to English speakers because the original survey instrument and the web interfaces used in the experiment were written in English, while the population was restricted to adults to ensure that informed consent could be obtained from each subject [59]. In light of these requirements, the leading global online advertising firm was engaged to craft a targeted campaign for the purpose of soliciting volunteers for the study. This approach was taken because the firm's demographic targeting technologies allowed subject recruitment to be explicitly constrained to the target population of English-speaking web users who were at least 18 years old. IP address restrictions were also enforced to ensure that each subject could participate in the experiment only once.

Upon providing their consent to participate in the experiment, subjects were asked to specify their age and gender. As noted in Section II, age and gender were explicitly included as covariates in the study because they have been identified by past research as being germane in the context of social influence, at least in certain circumstances [19], [39]. While other potential covariates such as culture, religion, or ethnicity were not explicitly considered in the study, any potentially confounding effects of these or other demographic characteristics were mitigated by the study's randomized design [21], [31].

Prior to data gathering, a formal *a-priori* sample size analysis was conducted in order to determine the appropriate number of subjects for the study. Given that linear models (discussed in the following subsection) would be used to evaluate several of the study's research hypotheses, and given that these models would contain a maximum of nine predictors, the *a-priori* sample size analysis revealed that a minimum of 879 observations would be required to detect a statistically significant f^2 effect size of 0.02 or above at a statistical power level of 0.85 [12], [52]. Based on the research design, observations from a maximum of two different experimental groups would be used to estimate any single linear model, yielding a minimum of approximately 440 observations per group. As described below, each subject in the study would yield three observations, meaning that each group would require a minimum of approximately 147 subjects in order to have a sufficient number of observations to detect a statistically significant f^2 effect size of 0.02 or above at a power level of 0.85.

In total, data were gathered from 503 subjects, of whom 254 (50.5%) were male and 249 (49.5%) were female.

Subjects ranged in age from 18 to 82 years, with the mean age being 34.02 years (*std dev* = 12.73). These demographic characteristics were observed to be consistent with the overall population of adult web users [40].

C. PROCEDURE

When data gathering began, the first 150 subjects who participated in the study were assigned to the baseline (control) group. The remaining 353 subjects were then iteratively assigned into the two treatment groups, with 177 subjects being assigned to Treatment Group 1 and 176 subjects being assigned to Treatment Group 2. This subject allocation method was employed because the baseline measurements obtained from the control group were used as inputs for the experimental manipulation of the treatment groups, hence making it necessary to acquire the baseline measurements in advance.

Subjects in the baseline group were simply shown the three interfaces and asked to rate each interface along their assigned dimension from Table 1. When aggregated, the responses from baseline subjects were regarded as the true, unadulterated ratings for each interface design characteristic, and served as the basis against which subject ratings from the treatment groups would be compared. As noted by Nielsen [34], this approach to combining the ratings of individual judges is common and appropriate when conducting heuristic interface evaluations.

The rating tasks and experimental process for subjects in the two treatment groups were identical to those of the baseline group, excepting that subjects in the treatment groups were provided with information about how other people rated the same interface and design characteristic that they themselves were currently considering. Subjects in Treatment Group 1 were provided with this information on an *ex ante* basis (*i.e.*, before submitting their own ratings), while subjects in Treatment Group 2 were provided with this information on an *ex post* basis (*i.e.*, after having already submitted their own ratings). In the case of the latter, subjects in Treatment Group 2 were given an opportunity to revise their original ratings after having been made aware of the average ratings of other people for a given interface and design characteristic. An illustration of the overall research design is provided in Fig. 3.

As shown in Fig. 3, 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 were not the true ratings given by others, however, but instead were experimentally manipulated
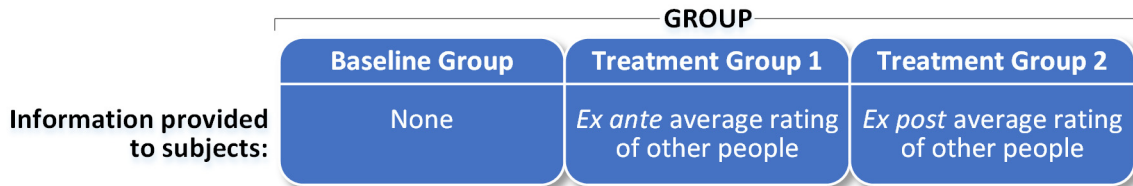
D. S. Soper: Informational Social Influence, Belief Perseverance, and Conservatism Bias in Web Interface Design Evaluations

IEEE Access

FIGURE 3. Overall research design. Responses from baseline subjects were used as the basis of comparison for responses obtained from subjects in the treatment groups. The three groups are differentiated based on whether and when information about the interface ratings of others was shared with subjects.

with a view toward gaining insights into the study's research hypotheses. Specifically, the artificial ratings supplied to subjects in the treatment groups were statistically derived from the distributions of the ratings obtained from the baseline group. To be more precise, the baseline mean rating and standard deviation for each combination of interface and design characteristic were used to compute the artificial score that was supplied to subjects in the treatment groups, with the artificial score being the value associated with a cumulative probability of 0.05 on the associated baseline rating's normal distribution. For example, imagine that the true rating obtained from baseline subjects for a particular interface characteristic were 4.0 (on a 1 to 7 scale), with a standard deviation of 1.0. By applying the normal distribution cumulative distribution function (CDF), it could be readily determined that 95% of subjects would naturally rate this interface characteristic at 2.36 or above, while only 5% of subjects would supply a rating lower than 2.36. In this case, treatment group subjects would be told that the artificially low score of 2.36 was the average rating given by other people when evaluating that particular interface characteristic. Using this approach, it would be statistically improbable for a subject in the treatment groups to assign a rating lower than 2.36 to the interface characteristic that she was evaluating. Any statistically significant differences between the true ratings given by the baseline group and the ratings given by the treatment groups could thus be attributed to informational social influence. This method of experimentally manipulating the interface ratings that were provided to subjects in the treatment groups is illustrated in Fig. 4.

After data gathering was complete, insight into the study's research hypotheses was gained through a combination of linear modeling, analysis of variance (ANOVA), and t-tests. Regression-based linear models were used to evaluate the extent to which subject ratings in the baseline group differed from those in the treatment groups. Each linear model was specified such that subject ratings were predicted by whether a subject belonged to the baseline group or to the model's associated treatment group, after controlling for the subject's age and gender, and the interface and design characteristic being evaluated. For this purpose, membership in the baseline or treatment group, subject gender, and the various interfaces and design characteristics were all appropriately coded using a series of binary dummy variables [13]. A one-way analysis of variance was used to ascertain whether there was a significant difference in the effects of *ex ante* and *ex post* social
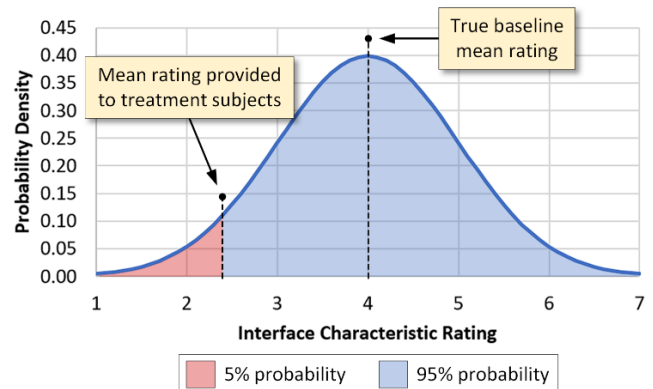

FIGURE 4. Illustrative example of experimental manipulation of mean interface characteristic ratings. Subjects in the treatment groups were provided with mean ratings that had been experimentally manipulated to be statistically improbable.

influence on subjects' interface ratings, after controlling for age and gender, and the interface and design characteristic being evaluated [45], [50]. Finally, a paired-sample t-test was used to compare the original and revised ratings for subjects in the *ex post* group (*i.e.*, Treatment Group 2), while a one-sample t-test was used to evaluate whether the average absolute distance between the revised ratings and baseline ratings was statistically different from zero [29]. The results of all of these analyses are presented and discussed in the following section.

IV. RESULTS AND DISCUSSION
A. HYPOTHESIS 1
The study's first hypothesis proposed that *ex ante* knowledge of the opinions of others would significantly influence a judge's web interface design ratings. In the experiment, the average ratings of other people that were provided to subjects in the *ex ante* group were manipulated to be artificially lower (*i.e.*, more negative) than the true ratings, with the hypothesis that the presence of this *ex ante* information would cause subjects to submit interface design ratings that were lower than they otherwise would have been. Initial estimation of the linear regression model that was used to test this hypothesis revealed that subject gender did not significantly affect interface design ratings. Gender was thus removed as a predictor, and the linear model was then duly reestimated. Although the overall model proved to be highly significant, the primary item of interest with respect to Hypothesis 1 was the parameter estimate for membership in the *ex ante* group.

IEEE *Access*

D. S. Soper: Informational Social Influence, Belief Perseverance, and Conservatism Bias in Web Interface Design Evaluations

After controlling for a subject's age and the significant effects of the different web interfaces and interface design characteristics being evaluated, the *ex ante* information provided about the opinions of others was found to exert a highly significant influence on subjects' interface design ratings, thus providing full support for Hypothesis 1. A summary of the results obtained from the test of this hypothesis is provided in Table 2.

TABLE 2. Results for the Test of Hypothesis 1.

Overall Linear Model Performance	Effect of *Ex Ante* Informational Social Influence on Interface Ratings
$F_{[4, 976]} = 48.743$	$\beta = -0.450$
$R^2_{Adjusted} = 0.163$	*std err* = 0.104
$p < 0.001$	$p < 0.001$

Since all of the interface characteristics in the experiment were rated on a 1 to 7, Likert-type scale, the parameter estimate β indicates that after controlling for other factors, the average rating given to an interface design characteristic by subjects in the *ex ante* treatment group was 0.450 units lower than the true, unadulterated rating for the same characteristic, as obtained from the baseline group. Put differently, when provided with an artificially low number that was represented to be the average rating of other people, subjects in the *ex ante* group succumbed to the informational social influence, and subsequently submitted interface design ratings that were significantly lower than the ratings that were obtained in the absence of such information. The actual mean interface design ratings for the baseline and *ex ante* groups were 4.864 and 4.382, respectively, indicating that the presence of the negative *ex ante* informational social influence yielded average ratings that were approximately 9.91% lower than the true ratings. This result should be of great interest to managers who would like to ensure that the design of their organization's website is evaluated as accurately and impartially as possible, since it plainly reveals that simply knowing in advance what other people think about a web interface can directly and significantly influence one's own opinions about that interface.

B. HYPOTHESIS 2

The study's second hypothesis addressed the behavior of subjects in the *ex post* condition, in which a judge begins by independently rating an interface design characteristic in the absence of any informational social influence. After recording her rating – and thus establishing her own beliefs about the quality of the interface design – the judge is then provided with the average rating of others for the same interface characteristic, and is given an opportunity to revise her original rating. The situation is thus characterized by a certain cognitive tension, with informational social influence (*i.e.*, the information about the ratings of others) impelling the judge to revise her original rating, while belief perseverance exerts a countervailing pull by impelling the judge to stand by her initial rating, even if the newly obtained information

indicates that the judge's own opinion diverges markedly from that of the group. In light of the behavioral predictions associated with belief perseverance, the second hypothesis proposed that *ex post* knowledge of the opinions of others would exert a significantly weaker influence on judges' web interface design ratings than *ex ante* knowledge of the same opinions.

Evaluating this hypothesis involved a deliberate, two-step process. First, the same regression-based linear modeling technique that was used for the *ex ante* condition was employed again for the subjects in the *ex post* condition, with a view toward obtaining a parameter estimate for the impact of *ex post* informational social influence on subjects' interface design ratings. Second, a one-way analysis of variance was used to determine if the *ex post* parameter estimate and its associated standard error differed significantly from their counterparts in the *ex ante* condition, thus allowing conclusions to be drawn regarding the relative impacts of *ex ante* vs. *ex post* informational social influence on web interface design ratings. As with the *ex ante* condition, initial estimation of the linear regression model revealed no significant impact of subject gender on interface design ratings. The model was hence reestimated after removing gender as a predictor, and the resulting overall model was observed to be highly significant. Further, after controlling for a subject's age and the significant effects of the different web interfaces and interface design characteristics being evaluated, the *ex post* information provided about the opinions of others was found to exert a highly significant influence on subjects' interface design ratings. These results are summarized in Table 3.

TABLE 3. Results for the Test of Hypothesis 2.

Overall Linear Model Performance	Effect of *Ex Post* Informational Social Influence on Interface Ratings
$F_{[6, 971]} = 33.129$	$\beta = -0.362$
$R^2_{Adjusted} = 0.165$	*std err* = 0.105
$p < 0.001$	$p < 0.001$
One-Way ANOVA Comparing Effects of *Ex Ante* and *Ex Post* Informational Social Influence	
$F_{[1, 1955]} = 346.944$	
$p < 0.001$	

The parameter estimate β suggests that on average, subjects in the *ex post* informational social influence group submitted interface design ratings that were 0.362 units lower (on a 1 to 7 scale) than the analogous ratings obtained from the baseline group. Thus, as with the *ex ante* group, the presence of informational social influence in the form of *ex post* knowledge of the opinions of others had a substantial impact on subjects' web interface design ratings. Since the actual mean interface design ratings for the baseline and *ex post* groups were 4.864 and 4.472, respectively, the presence of the negative *ex post* informational social influence yielded average ratings that were approximately 8.06% lower than the true ratings, as compared to nearly 10% lower for the *ex ante* group. As shown in Table 3, a one-way analysis of variance

D. S. Soper: Informational Social Influence, Belief Perseverance, and Conservatism Bias in Web Interface Design Evaluations

IEEE *Access*

revealed that the magnitude of the effect of *ex post* informational social influence on web interface design ratings was indeed significantly weaker than *ex ante* informational social influence, thereby providing full support for Hypothesis 2.

In light of the results reported above, it can be concluded that both *ex ante* and *ex post* informational social influence can lead to significant distortions in web interface design ratings. Belief perseverance, however, causes the magnitude of this effect to be significantly less pronounced in the context of *ex post* informational social influence. Put differently, once a judge has independently established her own beliefs about the quality of a web interface design, making the judge aware of the opinion of the group can, on average, be expected to influence the judge's beliefs, but the degree of this informational social influence will typically be significantly weaker than in the *ex ante* condition due to the countervailing pull of belief perseverance. Nevertheless, the effects of both *ex ante* and *ex post* informational social influence on web interface evaluations should be considered carefully by managers, since the results reported here demonstrate just how easily people's beliefs and opinions about a web interface can be manipulated or distorted by group pressure.

C. HYPOTHESIS 3

The results reported immediately above for the test of Hypothesis 2 reveal that on average, *ex post* knowledge of the group opinion significantly influences a judge's beliefs about the quality of a web interface, albeit to a lesser extent than *ex ante* knowledge (due to the effects of belief perseverance). The study's remaining two hypotheses explore questions relating to this *ex post* revision behavior in greater detail. First, to what extent is a judge willing to revise her beliefs in the milieu of *ex post* informational social influence? Drawing on theories of informational social influence and belief perseverance, the study's third hypothesis proposed that when provided with *ex post* knowledge about the group's ratings of the quality of a web interface design, the extent to which a judge revises her own ratings would be positively dependent on the distance between her initial ratings and those of the group. Thus, if a judge discovers that her initial ratings diverge greatly from those of the group, she can be expected to make more substantial revisions to her beliefs than if her initial ratings were closer to those of the group.

To test this hypothesis, a linear regression model was estimated for subjects in the *ex post* treatment group who had revised their initial web interface ratings after having been made aware of the group opinion. In the model, the distance between subjects' initial ratings and those of the group was used to predict the amount by which subjects ultimately revised their initial ratings, after controlling for age, gender, and the effects of the different web interfaces and interface design characteristics being evaluated. Initial estimation of the linear regression model revealed no significant impact of any of these control variables. All of the non-significant control variables were thus removed, and the model was reestimated. The resulting overall model was observed to be

TABLE 4. Results for the Test of Hypothesis 3.

Overall Linear Model Performance	Effect of Distance between Subject Ratings and Group Ratings on Rating Revision Behavior
$F_{[1, 36]} = 7.121$	$\beta = 0.253$
$R^2_{Adjusted} = 0.142$	*std err* $= 0.095$
$p < 0.05$	$p < 0.05$

statistically significant, as was the parameter estimate for the "distance" variable described above. These results are summarized in Table 4.

The results of the analysis thus revealed that the extent to which a judge is willing to revise her beliefs about a web interface after being informed of the group's opinion depends significantly on the distance between her initial opinion and that of the group. Further, the sign of the parameter estimate β indicates a positive relationship between these two variables, thus providing full support for Hypothesis 3. It can therefore be concluded that if a judge independently arrives at her own opinions about the quality of a web interface and subsequently discovers that her initial opinions diverge from those of the group, then the extent to which the judge revises her initial opinions will depend positively on the degree of that divergence.

D. HYPOTHESIS 4

As with Hypothesis 3, the study's final hypothesis addressed a judge's revision behavior as it relates to web interface design ratings in situations involving *ex post* informational social influence. Specifically, theory relating to conservatism bias suggests that even in the face of new evidence, a judge's prior views and opinions will continue to influence her beliefs such that the degree to which the judge revises her initial opinion in light of new information will be less than expected if she were to behave according to a Bayesian model. Thus, although being made aware of the opinion of the group may influence a judge to revise her initial ratings, the study's fourth hypothesis proposed that the extent of those revisions will, on average, be inadequate to fully close the gap between the judge's initial ratings and those of the group.

Evaluating the study's final hypothesis involved the use of both a paired-samples *t*-test and a one-sample *t*-test. First, among all of the subjects in the *ex post* group who revised their initial ratings, a paired-samples *t*-test was used to compare the distribution of initial ratings with the distribution of revised ratings, with a view toward determining if the subjects' revised ratings were statistically different from their paired initial ratings. Whereas the mean initial web interface design rating for these subjects was 5.47 (*std dev* = 1.572), the mean revised rating for the same subjects was 3.97 (*std dev* = 1.808). The paired-samples *t*-test revealed this difference to be statistically significant, indicating that subjects' revised ratings were indeed statistically different from their initial ratings after having been made aware of the opinion of the group. Although necessary, knowing that subjects revised

their ratings in a statistically significant manner was, however, insufficient in and of itself to draw any conclusions about the study's fourth hypothesis.

Having established that subjects in the *ex post* group significantly revised their web interface design ratings to accord more closely with the group opinion, the final question was whether the extent of those revisions was sufficient to fully close the gap between the subjects' initial ratings and the opinion of the group. To gain insights into this question, a one-sample *t*-test was used to compare the average distance between subjects' revised ratings and those of the group against a hypothesized mean distance of zero. Put differently, if the average distance between the subjects' revised ratings and those of the group was statistically equal to zero, then it could be concluded that the extent of the subjects' revisions was sufficient to fully close the distance between their initial ratings and the average group ratings, and that there was no evidence of conservatism bias in the subjects' behavior. Statistical evidence of the average distance between the subjects' revised ratings and those of the group being greater than zero, however, would reveal the presence of conservatism bias. The mean absolute difference between the subjects' revised ratings and those of the group was 1.867 (*std dev* = 1.539), and a one-sample *t*-test revealed this gap to be highly statistically significant. The results obtained from the *t*-tests for Hypothesis 4 are provided in Table 5 below.

TABLE 5. Results for the Test of Hypothesis 4.

Paired-Samples *t*-Test Comparing Initial and Revised Ratings
$t_{[37]} = 6.230$
$p < 0.001$
One-Sample *t*-Test Comparing Revised and Group Ratings
$t_{[37]} = 7.478$
$p < 0.001$

When considered together, the results reported in Table 5 indicate that although the subjects in the *ex post* group revised their initial ratings in the direction of the group opinion to a statistically significant degree, the extent of those revisions was insufficient to fully close the gap between their initial ratings and the average opinion of the group. These findings are consistent with the behavior predicted by the tenets of conservatism bias, and provide full support for Hypothesis 4.

V. SUMMARY, LIMITATIONS, AND CONCLUDING REMARKS

This article sheds light on the important roles of informational social influence and belief perseverance when people are evaluating the quality of a web interface. It is no exaggeration that websites are often the most publically visible face of modern organizations, and this fact alone should elevate the quality of the organization's website to a prominent position in boardroom discussions. Given that web interface assessments in organizational settings are commonly carried out by groups (*e.g.*, by focus groups [30], [35], web design teams [18], [61], *etc.*), understanding whether and to what

extent social influence phenomena affect judges' ratings is of critical importance for organizations that are seeking accurate and unadulterated information about the quality of their website designs. The current paper makes a novel and important contribution toward this understanding by rigorously demonstrating that judges' opinions about the quality of a web interface can be easily and significantly manipulated by informational social influence.

In brief, by means of a controlled, randomized experiment involving more than 500 subjects, this study's primary experimental findings show that both *ex ante* and *ex post* knowledge of the opinions of others significantly influence a judge's web interface design ratings. Additionally, the study reveals that *ex ante* knowledge exerts a substantially larger influence than *ex post* knowledge, with this difference being attributable to the effects of belief perseverance. The paper's findings further indicate that the extent to which a judge is willing to revise her own ratings on an *ex post* basis depends positively on the distance between her initial ratings and those of the group, and that when a judge does revise her ratings, the extent of those revisions is less than would be expected if she were to behave according to a Bayesian model of decision-making.

The primary experimental findings notwithstanding, this study also makes substantial contributions to theory through its integration and discussion of the competing effects of informational social influence and belief perseverance on human cognition, and ultimately on the way that people behave. As noted in the theoretical discussion in Section II, the opposing forces of informational social influence and belief perseverance suggest that the extent to which a judge is willing to revise her beliefs is a function of the comparative degree of cognitive pressure being exerted by these two competing forces. When the pressure to conform that is exerted by informational social influence exceeds the counteracting pressure to resist that is exerted by belief perseverance, then a judge can be expected to revise her beliefs, with the degree of revision depending on the degree of the situational imbalance between informational social influence and belief perseverance. These theoretical perspectives, supported by this study's experimental evidence, suggest a cognitive mechanism in which a person unconsciously weighs the pressure to conform to the group opinion against the pressure to resist and hold to her own views when evaluating a technology artifact such as a web interface in the context of social influence.

As with all works involving experimentation or social phenomena, this study has several limitations that merit acknowledgement. First, the study attempted to discern only whether *ex ante* and *ex post* informational social influence would cause subjects to unconsciously adjust their web interface ratings in the *negative* direction. Although there is no theoretical reason to expect that these forms of informational social influence would be incapable of also causing subjects to *raise* their ratings, influencing subjects to adjust their ratings in the positive direction was not tested in the current experiment. Indeed, phenomena that lead judges to artificially inflate their ratings of a technology artifact such as a web interface may

have very important implications for critical organizational decisions, and this represents a notable opportunity for future research.

Next, from a methodological perspective, the current study was constrained only to web interface evaluations, and it remains unknown whether the findings reported herein are generalizable to other varieties of human-technology interfaces. Again, there is no obvious reason why the theoretical tenets of informational social influence and belief perseverance should not be expected to apply similarly to evaluations of, say, mobile app interfaces or desktop application interfaces, but these predicted effects were not explicitly tested in the current experiment.

Finally, although the age and gender of the research subjects were included in the study, other demographic characteristics such as culture, religion, or ethnicity were not explicitly considered. There is, for example, some evidence in the literature indicating that cultural differences in the context of collectivism vs. individualism may be relevant to interface design [11], [14], but these ideas were not tested in the current study. While it is true that the experiment's randomized design served to inherently mitigate any potentially confounding effects of such characteristics on the overall results, the extent to which attributes such as culture, religion, or ethnicity influence web interface ratings in the presence of social influence remains unknown. Together, these limitations represent just a few of the many fruitful opportunities for future research in this area.

In the introduction to this article, it was noted that the preponderance of past heuristic interface evaluation research in the HCI literature has relied on judges who, either explicitly or implicitly, performed their interface rating tasks in environments free of social influence. The case was further built that because real-world interface evaluations are commonly carried out in social settings, social influence phenomena may cause judges to rate interfaces differently in social settings than they would in a non-social setting. Through careful experimental manipulation, the results reported in this article clearly indicate that informational social influence can exert a powerful impact on people's perceptions of a web interface. Given that this phenomenon can lead to substantial measurement error, and given the importance of being able to accurately measure the quality of a website design, the findings of this study suggest that organizational web interface design evaluations should be deliberately performed by individual judges in an environment that is free from social influence.

Although the current paper focused on social influence in the context of web interface design (specifically, interface attractiveness), interface aesthetics are but one component in the broader panoply of user experience (UX) research, wherein group-based evaluation methods and techniques are also common in both laboratory and real-world settings [4343, 60]. It is therefore reasonable to expect that social influence phenomena may also contaminate measurements in other forms of UX evaluations, and methods such as those described in this article may prove effective in identifying

and quantifying such contamination. It is also reasonable to expect that informational social influence and belief perseverance play significant roles in many other phenomena that lie at the intersection of human cognition and information technology, and this suggests a rich set of future research opportunities.

Ultimately, if our future is to be one in which humans extract maximum benefits from our technologies, then one of our guiding principles must be to find ways of making our interactions with those technologies as easy and as natural as possible. Achieving this goal will necessarily require that we develop ways of accurately measuring and evaluating the mechanisms that we create for interacting with our technologies. Identifying and eliminating sources of bias and error in these measurements will be critical to the success of this noble endeavor, and it is hoped that the current study and others like it will serve as accelerants that hasten the process.

REFERENCES

[1] A. Al Taboli and M. R. Abou-Zeid, "Effect of physical consistency of Web interface design on users' performance and satisfaction," presented at the Int. Conf. Hum.-Comput. Interact., 2007.

[2] A. M. Aladwani and P. C. Palvia, "Developing and validating an instrument for measuring user-perceived Web quality," *Inf. Manage.*, vol. 39, no. 6, pp. 467–476, May 2002.

[3] A. A. Ozok and G. Salvendy, "Measuring consistency of Web page design and its effects on performance and satisfaction," *Ergonomics*, vol. 43, no. 4, pp. 443–460, Apr. 2000.

[4] E. Aronson, T. D. Wilson, R. M. Akert, and S. R. Sommers, *Social Psychology*, 9th ed. New York, NY, USA: Pearson, 2016.

[5] S. E. Asch, "Effects of group pressure upon the modification and distortion of judgement," in *Groups, Leadership and Men*, H. Guetzkow, Ed., 1st ed. Pittsburgh, PA, USA: Carnegie Press, 1951, pp. 222–236.

[6] S. E. Asch, "Studies of independence and conformity: I. A minority of one against a unanimous majority," *Psychol. Monographs, Gen. Appl.*, vol. 70, no. 9, pp. 1–70, 1956.

[7] A. V. Banerjee, "A simple model of herd behavior," *Quart. J. Econ.*, vol. 107, no. 3, pp. 797–817, Aug. 1992.

[8] R. F. Baumeister and K. D. Vohs, *Encyclopedia of Social Psychology*, vol. 1. Thousand Oaks, CA, USA: Sage, 2007.

[9] P. Chandon, V. G. Morwitz, and W. J. Reinartz, "Do intentions really predict behavior? Self-generated validity effects in survey research," *J. Marketing*, vol. 69, no. 2, pp. 1–14, Apr. 2005.

[10] R. B. Cialdini, *Influence: The Psychology of Persuasion*. New York, NY, USA: William Morrow and Company, 1984.

[11] R. B. Cialdini, W. Wosinska, D. W. Barrett, J. Butner, and M. Gornik-Durose, "Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists," *Personality Social Psychol. Bull.*, vol. 25, no. 10, pp. 1242–1253, Oct. 1999.

[12] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ, USA: Lawrence Earlbaum Associates, 1988.

[13] P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Hove, U.K.: Psychology Press, 2014.

[14] C. A. Collazos and R. Gil, "Using cross-cultural features in Web design patterns," in *Proc. 8th Int. Conf. Inf. Technol., New Generat.*, Las Vegas, NV, USA, Apr. 2011, pp. 514–519.

[15] J. Cook, N. Oreskes, P. T. Doran, W. R. L. Anderegg, B. Verheggen, E. W. Maibach, J. S. Carlton, S. Lewandowsky, A. G. Skuce, S. A. Green, D. Nuccitelli, P. Jacobs, M. Richardson, B. Winkler, R. Painting, and K. Rice, "Consensus on consensus: A synthesis of consensus estimates on human-caused global warming," *Environ. Res. Lett.*, vol. 11, no. 4, Apr. 2016, Art. no. 048002.

[16] J. C. Coultas, "When in Rome. . . An evolutionary perspective on conformity," *Group Processes Intergroup Relations*, vol. 7, no. 4, pp. 317–331, 2004.

IEEE Access

D. S. Soper: Informational Social Influence, Belief Perseverance, and Conservatism Bias in Web Interface Design Evaluations

[17] M. Deutsch and H. B. Gerard, "A study of normative and informational social influences upon individual judgment," *J. Abnormal Social Psychol.*, vol. 51, no. 3, pp. 629–636, 1955.

[18] D. K. V. Duyne, J. Landay, and J. I. Hong, *The Design of Sites: Patterns, Principles, and Processes for Crafting a Customer-Centered Web Experience*. Reading, MA, USA: Addison-Wesley, 2002.

[19] A. H. Eagly, "Sex differences in influenceability," *Psychol. Bull.*, vol. 85, no. 1, pp. 86–116, 1978.

[20] W. Edwards, "Conservatism in human information processing," in *Judgment Under Uncertainty: Heuristics and Biases*, D. Kahneman, Eds. Cambridge, U.K.: Cambridge Univ. Press, 1982, pp. 359–369.

[21] M. W. Enkin and A. R. Jadad, *Randomized Controlled Trials*, 2nd ed. Oxford, U.K.: Blackwell Publishing, 2008.

[22] O. Gambino, L. Rundo, V. Cannella, S. Vitabile, and R. Pirrone, "A framework for data-driven adaptive GUI generation based on DICOM," *J. Biomed. Informat.*, vol. 88, pp. 37–52, Dec. 2018.

[23] J. Hartmann, A. Sutcliffe, and A. De Angeli, "Investigating attractiveness in Web user interfaces," presented at the SIGCHI Conf. Hum. Factors Comput. Syst., 2007.

[24] M. Hilbert, "Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making," *Psychol. Bull.*, vol. 138, no. 2, pp. 211–237, 2012.

[25] A. Iannessi, P.-Y. Marcy, O. Clatz, A.-S. Bertrand, and M. Sugimoto, "A review of existing and potential computer user interfaces for modern radiology," *Insights Into Imag.*, vol. 9, no. 4, pp. 599–609, Aug. 2018.

[26] W. Jorritsma, F. Cnossen, and P. M. A. van Ooijen, "Adaptive support for user interface customization: A study in radiology," *Int. J. Hum.-Comput. Stud.*, vol. 77, pp. 1–9, May 2015.

[27] P. Kadiyala and P. R. Rau, "Investor reaction to corporate event announcements: Underreaction or overreaction?" *J. Bus.*, vol. 77, no. 2, pp. 357–386, Apr. 2004.

[28] J. T. Kaplan, S. I. Gimbel, and S. Harris, "Neural correlates of maintaining one's political beliefs in the face of counterevidence," *Sci. Rep.*, vol. 6, no. 1, Dec. 2016, Art. no. 39589.

[29] G. Keller, *Statistics for Management and Economics, Abbreviated*. Boston, MA, USA: Cengage Learning, 2015.

[30] I. Küster and N. Vila, "Successful SME Web design through consumer focus groups," *Int. J. Qual. Rel. Manage.*, vol. 28, no. 2, pp. 132–154, Jan. 2011.

[31] D. C. Montgomery, *Design and Analysis of Experiments*, 8th ed. Hoboken, NJ, USA: Wiley, 2017.

[32] I. Nakarada-Kordic and B. Lobb, "Effect of perceived attractiveness of Web interface design on visual search of Web sites," presented at the 6th ACM SIGCHI New Zealand Chapter's Int. Conf. Comput.-Hum. Interact., Making CHI Natural, 2005.

[33] J. Nielsen, "Heuristic evaluation," in *Usability Inspection Methods*, J. Nielsen and R. L. Mack, Eds., 1st ed. New York, NY, USA: Wiley, 1994, pp. 25–62.

[34] J. Nielsen, "Usability inspection methods," in *Proc. Conf. Companion Hum. Factors Comput. Syst.*, Boston, MA, USA, 1994, pp. 413–414.

[35] J. Nielsen, "The use and misuse of focus groups," *IEEE Softw.*, vol. 14, no. 1, pp. 94–95, Jan. 1997.

[36] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Seattle, WA, USA, 1990, pp. 249–256.

[37] J. Oechssler, A. Roider, and P. W. Schmitz, "Cognitive abilities and behavioral biases," *J. Econ. Behav. Org.*, vol. 72, no. 1, pp. 147–152, Oct. 2009.

[38] J. W. Palmer, "Web site usability, design, and performance metrics," *Inf. Syst. Res.*, vol. 13, no. 2, pp. 151–167, Jun. 2002.

[39] M. Pasupathi, "Age differences in response to conformity pressure for emotional and nonemotional material.," *Psychol. Aging*, vol. 14, no. 1, pp. 170–174, 1999.

[40] P. R. Center, "Internet user demographics," in *Pew Research Internet Project Survey*, 1st ed. Washington, DC, USA: Pew Research, 2014.

[41] K. R. Popper, *The Logic of Scientific Discovery*. London, U.K.: Routledge, 2014.

[42] C. Rhee, J. Moon, and Y. Choe, "Web interface consistency in e-learning," *Online Inf. Rev.*, vol. 30, no. 1, pp. 53–69, Jan. 2006.

[43] V. Roto, M. Obrist, and K. Väänänen-Vainio-Mattila, "User experience evaluation methods in academic and industrial contexts," in *Proc. Workshop UXEM*, Uppsala, Sweden, 2009, pp. 1–5.

[44] W. E. Saris and I. N. Gallhofer, *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Hoboken, NJ, USA: Wiley, 2007.

[45] H. Scheffe, *The Analysis of Variance*. New York, NY, USA: Wiley, 1999.

[46] A. Smith and M. Anderson, *Online Shopping and E-Commerce*. Washington, DC, USA: Pew Research Center, 2016.

[47] A. Solano, J. C. Cerón, C. A. Collazos, H. M. Fardoun, and J. L. Arciniegas, "ECUSI: A software tool that supports the usability collaborative evaluation of interactive systems," in *Proc. 18th Int. Conf. Hum. Comput. Interact.*, Vilanova i la Geltrú, Spain, 2015, pp. 1–2.

[48] A. Solano, C. A. Collazos, C. Rusu, and H. M. Fardoun, "Combinations of methods for collaborative evaluation of the usability of interactive software systems," *Adv. Hum.-Comput. Interact.*, vol. 2016, pp. 1–16, 2016.

[49] D. S. Soper, "User interface design and the halo effect: Some preliminary evidence," presented at the 20th Americas Conf. Inf. Syst. (AMCIS), Savannah, GA, USA, 2014.

[50] D. S. Soper. (May 13, 2020). *Analysis of Variance (ANOVA) Calculator—One-Way ANOVA from Summary Data*. [Online]. Available: https://www.danielsoper.com/statcalc/

[51] D. Soper, "The effects of ex ante informational social influence on Web interface design ratings," in *Proc. 53rd Hawaii Int. Conf. Syst. Sci.*, Maui, HI, USA, 2020, pp. 4195–4203.

[52] D. S. Soper. (Apr. 20, 2020). *A-Priori Sample Size Calculator for Multiple Regression*. [Online]. Available: https://www.danielsoper.com/statcalc/

[53] D. S. Soper and S. Mitra, "An inquiry into mental models of Web interface design," presented at the 19th Americas Conf. Inf. Syst. (AMCIS), Chicago, IL, USA, 2013.

[54] D. S. Soper and S. Mitra, "The nature, antecedents, and impacts of visuospatial mental models of Web interface design," *IEEE Access*, vol. 4, pp. 7930–7939, 2016.

[55] D. S. Soper and F. Piepkorn, "Halo effect contamination in assessments of Web interface design," *Open J. Inf. Syst.*, vol. 5, no. 1, pp. 1–23, 2018.

[56] A. Sutcliffe, "Assessing the reliability of heuristic evaluation for Web site attractiveness and usability," presented at the 35th Annu. Hawaii Int. Conf. Syst. Sci. (HICSS), 2002.

[57] M. Turoff, M. Chumer, B. V. de Walle, and X. Yao, "The design of a dynamic emergency response management information system (DERMIS)," *J. Inf. Technol. Theory Appl.*, vol. 5, no. 4, pp. 1–35, 2004.

[58] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *J. Risk Uncertainty*, vol. 5, no. 4, pp. 297–323, Oct. 1992.

[59] U.S. Department of Health Human Services, "Title 45, public welfare; part 46, protection of human subjects," in *Code of Federal Regulations*. Washington, DC, USA: U.S. Dept of Health and Human Services, Jul. 2018.

[60] A. P. O. S. Vermeeren, E. L.-C. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Väänänen-Vainio-Mattila, "User experience evaluation methods: Current state and development needs," in *Proc. 6th Nordic Conf. Hum.-Comput. Interact.*, Reykjavík, Iceland, 2010, pp. 521–530.

[61] P. Zaphiris and S. Kurniawan, *Human Computer Interaction Research in Web Design and Evaluation*. Hershey, PA, USA: IGI Global, 2007.

DANIEL S. SOPER was born in Aurora, CO, USA, in 1978. He received the B.Sc. degree in business administration with a second major in sociology and the M.Sc. degree in computer information systems from Colorado State University, in 2001 and 2004, respectively, and the Ph.D. degree in information systems from Arizona State University, in 2008.

After his Ph.D. degree, he joined the Faculty of the Mihaylo College of Business and Economics, California State University, Fullerton, where he currently serves as a Professor of information systems and decision sciences and the Vice-Chair of the Information Systems and Decision Sciences Department. He also served for several years as the Mihaylo College's Associate Dean and the Director of M.B.A. and Graduate Programs. His current research interests include the realms of machine learning and artificial intelligence, computational linguistics and natural language processing, human cognition, and human–computer interaction. His research has appeared in a wide variety of information systems and computer science journals and conference proceedings, including a number of different IEEE journals. His educational videos have received millions of views on YouTube, and he is also the creator of one of the world's most popular statistics websites, which has been used by more than 65 million people.

● ● ●