

## **Session 4**

# **Cognitive Issues in Designing Web Surveys**

## Social Presence in Web Surveys

### **Mick P. Couper**

University of Michigan  
P.O. Box 1248, Ann Arbor,  
MI 48106 USA  
+1 734 647 3577

### **Roger Tourangeau**

University of Michigan  
P.O. Box 1248, Ann Arbor,  
MI 48106 USA  
+1 301 314 7984

### **Darby M. Steiger**

The Gallup Organization  
One Church Street, Suite 900  
Rockville, MD 20850  
+1 301 309 9439

### **ABSTRACT**

Social interface theory has widespread influence in the field of human-computer interaction. The basic thesis is that humanizing cues in a computer interface can engender responses from users similar to human-human interaction. In contrast, the survey interviewing literature suggests that computer administration of surveys on highly sensitive topics reduces or eliminates social desirability effects, even when such humanizing features as voice are used.

In attempting to reconcile these apparently contradictory findings, we varied features of the interface in a Web survey (n=3047). In one treatment, we presented an image of 1) a male researcher, 2) a female researcher, or 3) the study logo at several points. In another, we varied the extent of personal feedback provided. We find little support for the social interface hypothesis. We describe our study and discuss possible reasons for the contradictory evidence on social interfaces.

### **Keywords**

Social interfaces, Web surveys, social desirability

### **INTRODUCTION**

Social interface theory [8][11][21] appears to be generating much interest in the world of human-computer interaction. Much of the support for this perspective comes from laboratory-based studies.

A growing number of laboratory experiments suggest that relatively subtle cues (such as “gendered” text or simple inanimate line drawings of a face) in a computer interface can evoke reactions similar to those produced by a human, including social desirability effects. Nass, Moon and Green [17], for example, conclude that the tendency to stereotype by gender can be triggered by such minimal cues as the voice on a computer. Based on the results of a series of experiments that varied a number of cues in computer tutoring and other tasks, Nass and colleagues [9][16][17][18] argue that computer interfaces (even the words used in a text-based tutoring task) can engender reactions from subjects similar to those evoked by interactions with other people. Their central thesis is that people treat computers as social actors not as inanimate tools (see also [3]).

Additional support for the hypothesis that a computer interface can function as a virtual human presence comes from a study by Walker, Sproull, and Subramani [27]. They administered questionnaires to people using either a text display or one of two talking-face displays to ask the questions. Those interacting with a talking-face display spent more time, made fewer mistakes, and wrote more comments than did people interacting with the text display. However, people

who interacted with the more expressive face liked the face and the experience less than those who interacted with the less expressive face. In a subsequent experiment, Sproull and colleagues [23] varied the expression of a talking face on a computer-administered career counseling interview; one face was stern, the other more pleasant. The faces were computer-generated images with animated mouths. They found that: “People respond to a talking-face display differently than to a text display. They attribute some personality attributes to the faces differently than to a text display. They report themselves to be more aroused (less relaxed, less confident). They present themselves in a more positive light to the talking-face displays.” (p. 116) (see also [20]).

If the social interface theory is correct, it has important implications for the survey research industry for several reasons: 1) There is an increasing trend toward the use of computer-assisted interviewing, and especially the use of the World Wide Web, for administration of surveys [4][5]. 2) More and more surveys include sensitive questions (on sexual behavior, drug use, etc.), raising concerns about social desirability effects and interviewer influences. 3) Concomitant with the above, there is an increasing move towards the using of computer-assisted self-interviewing (CASI) methods, whereby the respondent interacts directly with the computer to answer questions. The most recent manifestation of this trend is the development of audio-CASI, in which the respondent listens to the questions read over headphones using a digitized voice, and enters the responses into the computer. A number of studies have compared CASI and audio-CASI to alternative approaches in field-based experiments. The general finding is that CASI methods (including audio-CASI) reduce social desirability distortions (i.e., increase reporting of sensitive information) over both interviewer-administered and paper-based self-administered methods [24]. Some have gone so far as to argue that voice does not matter when asking questions about sexual behavior (e.g., [25][26]), although these claims have not been empirically verified.

These results appear to contradict the findings of the social interface researchers. If subtle humanizing cues do indeed influence the behavior of computer users, we would fully expect the gender of the voice to affect the answers given to survey questions on topics such as gender attitudes and sexual behavior. Given the increasing use of multimedia tools on the Web, the addition of a variety of humanizing visual and/or aural cues, as is possible in Web surveys, may negate or at least mitigate the beneficial effects of self-administration, especially for items of a sensitive nature. It is thus important to explore the apparent contradiction between the social interface and survey methods work, and attempt to bring these two strands of research together.

There are several differences between the two literatures that could account for the discrepant results. For one, virtually all of the social interface research has been conducted in laboratory settings with students as volunteer subjects. In contrast, the survey-based findings are from probability samples of broader populations (e.g., teenage males, women 15-44, adult U.S. population). In the former, the number of subjects is typically measured in tens or scores while, in the latter, sample sizes go up to the thousands. The measurement settings also differ considerably. The social interface work is typically done in a laboratory setting, free from distractions and with privacy ensured. Most of the CASI surveys are conducted in the respondent’s home with an interviewer present, and sometimes with other family members home at the time. The perceived threat from disclosure varies greatly across the two settings. The more sterile, controlled environment of the laboratory may well focus subjects’ attention on the

experimental manipulation more than in an uncontrolled real-world setting with many potential distractors and less expectation of experimental manipulation. Furthermore, the measurement devices differ considerably between the two approaches. The social interface experiments often use subjects' performance on a computer task as the dependent measure. When questionnaire measures are used, they are typically self-reports of social desirability or impression management. The findings from the survey world are based on overt measures of highly sensitive behaviors (e.g., abortions, number of sex partners, engagement in high risk sexual behaviors, illicit drug use, etc.).

We obviously cannot address all these issues and resolve the controversy in a single study. We are engaged in a program of research to explore the issue of the effect of interface design and social interface features on survey responses. Work currently underway involves experiments on the effect of virtual interviewers (talking heads) on racial attitudes, manipulation of voice (male/female) in audio-CASI surveys, manipulation of privacy effects on self-disclosure in text-CASI versus audio-CASI surveys, and the effect of interface features on social desirability distortions in Web and interactive voice response (IVR) surveys. In this paper we report on the Web survey experiment we conducted as part of this broader research agenda.

## **METHODS**

We carried out two studies that examined the impact of characteristics of the interface on the responses obtained in a Web survey. Our first study compared six versions of a Web survey administered to 202 participants in a Web panel maintained by the Gallup Organization. The second study compared the same six versions of the survey in a much larger sample of Web users purchased from a commercial vendor, Survey Sampling, Inc. (SSI). Given that the design of the survey was identical across versions, and the findings were very similar, we focus on the larger sample from SSI here.

### **Experimental Manipulation**

The different versions of the Web questionnaire differed along two dimensions--the degree that the program presented personalizing cues and the degree that it seemed to interact with the respondent. At several points in the questionnaire, the personalized versions of the questionnaire displayed a picture of one of the male researchers, or one of the female researchers. A comparison version of the questionnaire presented the logo for the study, instead of the investigators' picture. Along with the pictures, the program displayed relevant statements from the investigator: "Hi! My name is Roger Tourangeau. I'm one of the investigators on this project. Thanks for taking part in my study." The high interaction versions of the questionnaire used the first person in introductions and transitional phrases (e.g., "Thanks, [name]. Now I'd like to ask you a few questions about the roles of men and women") and occasionally echoed back to the respondents their earlier answers ("According to your responses, you exercise once daily ..."). The low interaction versions used more impersonal language ("The next series of questions is about the roles of men and women") and gave less tailored feedback ("Thank you for this information"). Examples of these designs are shown in Figures 1-3 below.

This resulted in a 3×2 experiment, fully crossing the two dimensions of social presence we manipulated. We randomly assigned respondents to one of the six cells in the design, as shown in Table 1.

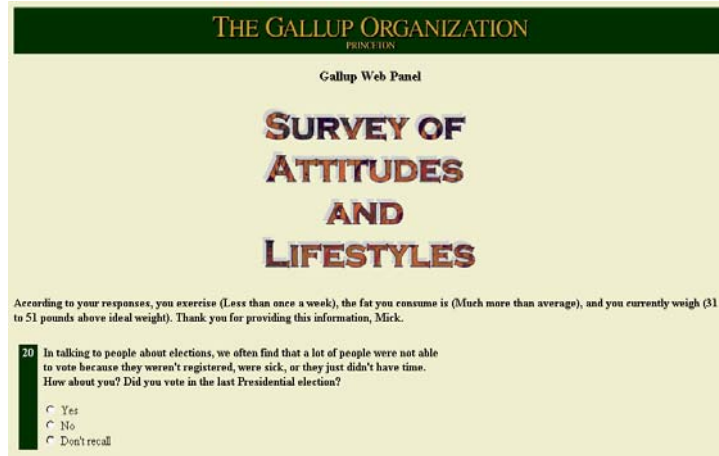


Figure 1: Logo and Personal Feedback

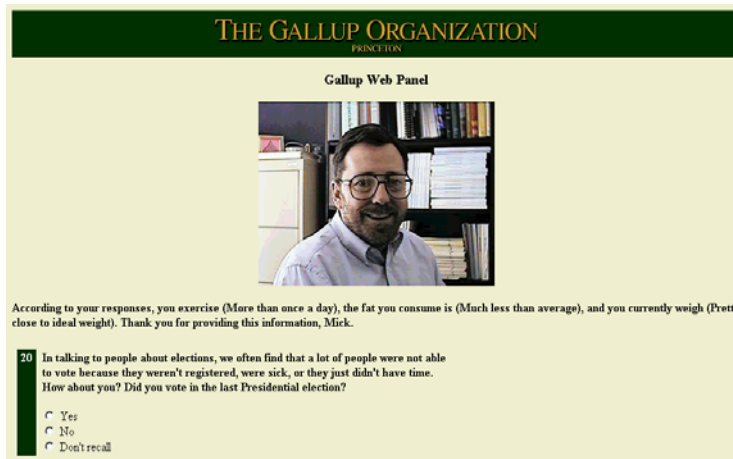


Figure 2: "Male" Interface and Personal Feedback



Figure 3: "Female" Interface

## Questionnaire

The survey questionnaire contained the following types of items:

- Gender attitudes: 8 items from Kane & Macauley's [10] study regarding the roles of men and women (e.g., *Thinking about men as a group, do you think men have too much influence, about the right amount of influence, or too little influence in society?*).
- Socially undesirable behaviors: 5 items on drinking and illicit drug use, 3 less-sensitive items on diet and exercise.
- Socially desirable behaviors: items on voting and church attendance.
- Self-reported social desirability: 16 items from the Marlowe-Crowne Social Desirability (SD) Scale [6] and the 20-item Impression Management (IM) scale from the Balanced Inventory of Desirable Responding (BIDR)[20].
- Trust: 3 items on trust (e.g., *Most people can be trusted*).
- Debriefing questions: 9 items to assess social presence and evaluate the interview experience (e.g., *How much was this interview like an ordinary conversation? How much was it like dealing with a machine?*).
- Demographic questions.

We included the gender attitude items to see whether our attempt to personalize the interface produced “deference” effects paralleling the gender-of-interviewer effects with actual interviewers – that is, more pro-feminist responses with the “female” than with the “male” interface. The items on diet, exercise, drinking, drug use, voting, and attendance at church were all included to test the hypothesis that humanizing the interface (both by personalizing it and by making it more interactive) would increase the number of socially desirable responses and decrease the number of socially undesirable responses. The SD and IM items have been used for similar purposes (to measure socially desirable responding) in the work by Nass and colleagues, and we included them in our studies for the sake of comparability. We included the trust items to see whether the impact of the experimental variables was greater among those low in trust (as found by Aquilino and LoSciuto [1]). The demographic items were included as a check on the randomization and to assess subgroup differences. On average, the questionnaire took about 15 minutes to complete.

## Hypotheses

Consistent with the social interface theory, our hypotheses were that increasing the social nature of the Web survey interaction, whether by personalization or interaction, would yield: 1) higher self-reports of social desirability and impression management, and 2) lower reports of socially undesirable behaviors (drug use, drinking, fat consumption) and higher reports for socially desirable behaviors (church attendance, voting, exercise). We also hypothesized that the “male” interface would elicit less positive attitudes toward women, while the “female” interface would yield more positive attitudes, with the neutral logo occupying a middle position.

## Sample Design and Implementation

The frame for the SSI sample consists of more than seven million e-mail addresses of Web users. SSI has compiled this list from various sources; in each case, visitors to specific Web sites

agreed to receive messages on a topic of interest. SSI selected a sample of 15,000 e-mail addresses and sent out an initial e-mail invitation to take part in “a study of attitudes and lifestyles.” The e-mail invitation included the URL of the Web site where our survey resided and a PIN number (which prevented respondents from completing the questionnaire more than once). After ten days, SSI sent a second reminder e-mail to sample persons who had not yet completed the survey. A total of 3,047 sample members completed the questionnaire, for a response rate of approximately 20%. (Less than 1% of the e-mails bounced back as invalid addresses.) Another 434 persons (3% of the sample) began the survey but broke off without finishing it. We focus here on the respondents who completed the survey. The number of completed cases per cell is shown in Table 1.

**Table 1. Number of Subjects per Cell**

<b>Personalizing Cues</b>	<b>Interaction</b>		<b>Total</b>
	<b>Low</b>	<b>High</b>	
Logo	492	502	994
Male picture	529	529	1058
Female picture	501	492	993
<b>Total</b>	<b>1522</b>	<b>1523</b>	<b>3047</b>

The number of cases we obtained far exceed that for most of the experimental studies on social interfaces (typically 10-20 subjects per cell). Statistical power to detect effects of the manipulations should not be a problem in our study. Furthermore, the respondents to our survey represent a much more diverse group than is typically found in laboratory-based experiments.

### **ANALYSIS AND RESULTS**

We created a number of scales for the key measures in our study. For the social desirability scale we assigned a score of 1 to every answer that represented socially desirable responding, and a 0 to every response that did not. This yielded a scale with a range of 0 to 16, with a high score indicating a greater tendency towards socially desirable responding. We used the same strategy for the impression management scale, creating a summary score ranging from 0 to 20, again with a high score indicating greater impression management. For the gender attitude items, we created a scale that combined responses across the eight items, by scoring responses to each item in a consistent direction and then summing across the items. The resultant scale ranged from 8 to 24, with a high score indicating pro-feminist or more egalitarian attitudes. Similarly, we created an index to combine answers to a number of the sensitive questions. Our index was the number of embarrassing answers given in response to those questions; the index varied from 0 to 7. Respondents got a point each if they reported they consumed more dietary fat than the average person, were 20 pounds or more over their ideal weight, drank alcohol almost every day (or more often), had smoked marijuana, had used cocaine, did not vote in the last election, and did not attend church in the last week.

The results for each of these scales by each of the two experimental conditions are presented in Table 2. None of the effects reach statistical significance ( $p > .10$ ) with the exception of the effect of personalization on gender attitudes, to which we return later. To perform a stronger test of the social interface hypothesis, we combined the two experimental conditions, and contrasted the high social interface group (high interaction, and male/female picture) with the low social interface group (low interaction, logo). The differences in means again do not approach statistical significance. We tried a variety of alternative specifications, including control variables, and interaction terms, but the findings essentially remain the same.

**Table 2. Scale Means by Condition (Standard Errors in Parentheses)**

	<b>Social Desirability</b>	<b>Impression Management</b>	<b>Gender Attitudes</b>	<b>Sensitive Admissions</b>
<b>Interaction</b>	n.s.	n.s.	n.s.	n.s.
Low interaction	7.87 (0.14)	8.84 (0.19)	18.25 (0.16)	3.27 (0.07)
High interaction	7.83 (0.10)	8.91 (0.13)	17.98 (0.11)	3.30 (0.05)
<b>Personalization</b>	n.s.	n.s.	$p < .05$	n.s.
Logo	7.95 (0.10)	8.87 (0.13)	18.09 (0.12)	3.27 (0.05)
Male Picture	7.77 (0.09)	8.73 (0.13)	17.77 (0.11)	3.21 (0.05)
Female Picture	7.85 (0.09)	8.84 (0.13)	18.19 (0.11)	3.31 (0.05)

**Table 3. Percentages on Behavior Variables by Condition**

	<b>% Used Cocaine in Lifetime</b>	<b>% Smoked Marijuana in Last Year</b>	<b>% Drink Daily or Almost Daily</b>	<b>% Attended Church Last Week</b>	<b>% Voted in Last Election</b>
<b>Interaction</b>	n.s.	n.s.	n.s.	n.s.	n.s.
Low interaction	14.2	10.7	7.8	23.3	53.2
High interaction	15.3	10.2	7.7	25.7	52.2
<b>Personalization</b>	n.s.	n.s.	n.s.	n.s.	$p < .05$
Logo	15.4	10.8	7.4	23.2	52.8
Male Picture	14.7	9.9	8.0	24.3	55.3
Female Picture	14.2	10.5	7.7	26.1	49.7

There were a few scattered findings for some of the individual sensitive items. We include a few examples of both socially undesirable and socially desirable behaviors in Table 3. For reports about voting, the personalization variable had a significant impact ( $X^2=6.35$ ,  $df=2$ ,  $p < .05$ ). Contrary to expectation, the respondents who got the female picture were least likely to say they had voted in the most recent election, while those who got the male picture were most likely to



say they had voted. In general, though, neither the level of personalization nor the level of interaction had much effect on reports about sensitive topics.

The only expected effect that found support in our data was related to gender attitudes (see Table 2). We expected respondents of both sexes to report the most pro-feminist attitudes when the program displayed pictures and messages from the female investigator and the least pro-feminist attitudes when the program displayed the pictures and messages from the male investigator. We expected the group who got the survey logo to fall in between the other two. This pattern was apparent, and reached statistical significance ( $F=5.52$ ,  $df=1,3028$ ,  $p<.05$ ).

One explanation for the significant gender effect could relate to the “mere presence” hypothesis from studies of prejudice. Research on race-of-interviewer effects [7][9] has found that racial stereotypes can be “primed” simply by presenting an image of the target group. This view is an alternative to the “racial deference” or “polite stranger” hypotheses [2][22] which suggest that people avoid articulating negative stereotypes in the presence of another person, particularly a member of the target group, out of politeness. This latter view is more akin to the social presence model. The fact that the female picture elicits the most pro-feminist attitudes, and the male picture the least, with the logo occupying a middle position, may suggest support for the “mere presence” theory of stereotypes, rather than for a social presence interpretation. This obviously deserves further research attention.

## **DISCUSSION AND CONCLUSIONS**

Our results were much weaker than the ones reported by Nass, Sproull, or their colleagues. We were puzzled by the discrepancy. We included some of the same measures used in the past work (e.g., the BIDR), and our sample sizes were much larger than in the earlier studies. Several explanations may account for the discrepancy. One could argue that our experimental manipulations were not sufficiently blatant to generate this hypothesized effect. We believe our manipulations to be at least as obvious as many of the social interface research studies which often use very subtle cues, such as a label on a computer monitor [14] or the shape of a mouth on a computer-displayed face [23] (see also [11][19]). Another explanation may relate to the use of college students in the experimental studies. In our study we had sufficient sample size to control for several variables--whether the respondent was currently a student, age, prior survey experience, and level of trust--that we thought might interact with the experimental variables and explain why our results differ from those of the earlier studies. For example, we tested the hypotheses that students are more sensitive to the characteristics of the interface and that respondents with prior experience with Web surveys would be less sensitive to them. None of these hypotheses received much support--we did not find any significant interactions between these individual differences variables and the experimental variables on the reporting of sensitive information or gender attitudes.

Another possible explanation, which we could not test, is that the demand characteristics of laboratory-based experimentation yield results that are not replicated in distraction-filled field-based surveys. In the experimental work, undergraduate students (often in psychology classes) typically are recruited for an experiment. They are aware of being in an experiment, and may be alert to any cues that might help them figure out the experimental manipulation. In contrast, survey respondents are typically unaware of being in an experiment, and believe the ostensible

reason for the survey is to elicit their views on particular topics. These differences may account for the failure of the social interface theory to replicate beyond the laboratory.

Given the influence of the social interface perspective in human-computer interaction (HCI) research and interface design, it is important to understand whether and how the findings from this work translate to the real-world experiences of those who interact with computers. In one such application (a Web survey) we appear to find little support for the social interface hypothesis.

## **ACKNOWLEDGMENTS**

This research is supported by the National Science Foundation, Award # SES-9910882. We thank Robert Tortora and Margrethe Montgomery of the Gallup Organization for their assistance with in planning and carrying out this project. This paper was originally presented at the CHI '01 Conference on Human Factors in Computing Systems in Seattle, Washington in April, 2001.

## **REFERENCES**

1. Aquilino, W., & LoSciuto, L. Effect of Interview Mode on Self-Reported Drug Use. *Public Opinion Quarterly*, 54 (1990), 362-395.
2. Athey, K.R., Coleman, J.E., Reitman, A.P., & Tang, J. Two Experiments Showing the Effects of the Interviewer's Racial Background on Responses to Questionnaires Concerning Racial Issues. *Journal of Applied Psychology*, 44 (1960), 562-566.
3. Couper, M.P. Review of Byron Reeves and Clifford Nass, 'The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places'. *Journal of Official Statistics*, 13 (1998), 441-443.
4. Couper, M.P. Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly*, 64 (2000), forthcoming.
5. Couper, M.P., & Nicholls II, W.L. The History and Development of Computer Assisted Survey Information Collection. In M.P. Couper, et al. (eds.), *Computer Assisted Survey Information Collection*. New York: Wiley, 1998.
6. Crowne, D., & Marlowe, D. *The Approval Motive*. New York: John Wiley, 1964.
7. Devine, P.G. Stereotypes and Prejudice: Their Automatic and Controlled Components. *Journal of Personality and Social Psychology*, 56 (1989), 5-18.
8. Dryer, D.C. Getting Personal with Computers: How to Design Personalities for Agents. *Applied Artificial Intelligence*, 13 (1999), 273-295.
9. Fogg, B.J., & Nass, C. Silicon Sycophants: The Effects of Computers That Flatter. *International Journal of Human-Computer Studies*, 46 (1997), 551-561.
10. Kane, E.W., & Macauley, L.J. Interviewer Gender and Gender Attitudes. *Public Opinion Quarterly*, 57 (1993), 1-28.
11. Kiesler, S. & Sproull, L. 'Social' Human-Computer Interaction. In B. Friedman (ed.) *Human Values and the Design of Computer Technology*. Stanford, CA: CSLI Press, 1997.

12. Kiesler, S., Sieff, E., & Geary, C. The Illusion of Privacy in Human-Computer Interaction. Carnegie Mellon University: Unpublished paper, 1992.
13. Kim, J., & Moon, J.Y. Designing Towards Emotional Usability in Customer Interfaces—Trustworthiness of Cyber-Banking System Interfaces. *Interacting with Computers*, 10 (1998), 1-29.
14. Moon, Y. Impression Management in Computer-Based Interviews: The Effects of Input Modality, Output Modality, and Distance. *Public Opinion Quarterly*, 62 (1998), 610-622.
15. Moon, Y. Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers. *Journal of Consumer Research*, 26 (2000), 323-339.
16. Nass, C., Fogg, B.J., & Moon, Y. Can Computers be Teammates? *International Journal of Human-Computer Studies*, 45 (1996), 669-678.
17. Nass, C., Moon, Y., & Carney, P. Are People Polite to Computers? Responses to Computer-Based Interviewing Systems. *Journal of Applied Social Psychology*, 29 (1999), 1093-1110.
18. Nass, C., Moon, Y., & Green, N. Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers with Voices. *Journal of Applied Social Psychology*, 27 (1997), 864-876.
19. Parise, S., Kiesler, S., Sproull, L. & Waters, K. Cooperating with Life-Like Interface Agents. *Computers in Human Behavior*, 15: (1999), 123-142.
20. Paulhus, D.L. Two-Component Models of Socially Desirable Responding. *Journal of Personality and Social Psychology*, 46 (1984), 598-609.
21. Reeves, B., & Nass, C. *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*. Cambridge: CSLI and Cambridge University Press, 1997.
22. Schuman, H., & Converse, J. The Effects of Black and White Interviewers on Black Responses in 1968. *Public Opinion Quarterly*, 35 (1971), 44-68.
23. Sproull, L., Subramani, M., Kiesler, S., Walker, J.H., & Water, K. When the Interface Is a Face. *Human-Computer Interaction*, 11, (1996), 97-124.
24. Tourangeau, R. & Smith, T. Collecting Sensitive Information with Different Modes of Data Collection. In M.P. Couper, et al. (eds.), *Computer Assisted Survey Information Collection*. New York: Wiley, 1998.
25. Turner, C.F., Forsyth, B.H., O'Reilly, J.M., Cooley, P.C., Smith, T.K., Rogers, S.M., & Miller, H.G. Automated Self-Interviewing and the Survey Measurement of Sensitive Behaviors." In M.P. Couper, et al., (eds.), *Computer Assisted Survey Information Collection*. New York: Wiley, 1998.
26. Turner, C.F., Ku, L., Rogers, S.M., Lindberg, L.D., Pleck, J.H., & Sonenstein, F.L. Adolescent Sexual Behavior, Drug Use and Violence: Increased Reporting with Computer Survey Technology. *Science*, 280 (May, 1998): 867-873.
27. Walker, J., Sproull, L., & Subramani, R. Using a Human Face in an Interface. *Proceedings of the Conference on Human Factors in Computers*. Boston (1994), ACM Press, pp. 85-91.

Comments on “Social Presence in Web Surveys” by Mick Couper, Roger Tourangeau & Darby Steiger.

Frederick Conrad  
Bureau of Labor Statistics

The paper by Mick Couper, Roger Tourangeau, and Darby Steiger presents exactly the kind of foundational work required to understand the consequences of moving to web surveys. Although this move is inevitable – web-based questionnaires will eventually be a common, if not the predominant, method for collecting Federal survey data – the differences between the web and other media are only just beginning to be identified and studied.

The web is actually similar to many other media but not the same as any of them. For example, it is similar to paper. The primary content is (still) presented in documents – hence the page metaphor. But it is different from paper in several ways, the most obvious being that the content in a web document is hyperlinked, making the document interactive. Web pages are similar to software in that users interact with both by clicking and typing; yet they are different in the sense that web pages are essentially static files (embedded JavaScript notwithstanding). The web resembles television: the content is sometimes partitioned into channels but the two media are different in the sense that web content is mostly text and graphics but not video and is available on demand while TV content is made available at prescribed times. And so on.

The point is that it is natural to apply what is already known about a similar seeming medium when adapting a task to web administration. Web surveys are a case in point. The 1999 book by Dillman, *Mail and Internet Surveys* is an update of his 1978 book *Mail and Telephone Surveys*, reflecting the fundamental similarities between paper and web-based questionnaires. In contrast, the Couper et al. study explores one of the novel aspects of the medium – for which I applaud the authors – namely, the interactive character of web surveys. In particular, the authors ask whether the interactivity of web surveys produces social presence effects – the tendency for respondents to behave as if the survey instrument was animate or administered by an interviewer. Social presence effects have been demonstrated in other interactive media by researchers such as Reeves and Nass and their colleagues. If web surveys do produce social presence effects then these must be reconciled with the increased sense of privacy that is apparently produced by self-administration – including self-administration of computerized survey instruments such as CASI and ACASI. In general, self-administration leads to what seems to be more honest reporting of sensitive behaviors (e.g. Turner, Forsyth, O’Reilly, Cooley, Smith, Rogers, & Miller, 1998; Tourangeau & Smith, 1996; Schaeffer, 2000).

In fact, Couper, et al. report no social presence effects in their two experiments on self-administered web-based surveys. This is potentially good news for researchers hoping to reap the benefits of self-administration on the web. After all, if the celebrated advantages of CASI and ACASI were to disappear when a remote computer is involved, then the move to web surveys would be a giant step backward, at least for collecting sensitive

information. The problem is that Couper et al. report *no effect*, not *the disappearance of an effect* under a particular experimental manipulation. So it is hard to interpret what is essentially a null result (albeit with a good sample and adequate power<sup>1</sup>). It could be that a different experimental manipulation would produce evidence of social presence. Alternatively, it could be that the experimental groups actually exhibit evidence of social presence but this is undetectable without a comparison group. Yet another possibility is that Couper et al. found no evidence of social presence in their experiments because such effects are confined to laboratory studies involving special tasks. I will take up each of these possibilities in turn.

Social presence effects may be restricted to situations in which the user (or respondent) is particularly aware of the agent-like character of the computer (or other medium) and it could be that Couper et al. did not sufficiently create this awareness among their respondents. The authors dismiss this kind of explanation because Reeves & Nass (1996) report many effects based on subtle manipulations such as those involving stick figures or gender of computer voices. But I think this kind of explanation may still apply because Reeves & Nass and their colleagues sometimes go to great lengths to make these manipulations effective. In a well known study by Nass, Moon & Carney (1999) (reported in the Reeves & Nass, 1996 book) users rated a computer's performance on a tutoring task more favorably and homogeneously when they registered their ratings using the same computer they were evaluating than when they used another computer or a paper questionnaire – as if they were being polite to the computer while interacting with it. During the tutoring task, the computer presented a series of facts to users; after reading each fact, users rated how much they knew about that fact. Users were led to believe that the more they knew about a fact, the fewer related facts they would be presented (in fact all users were presented the same facts). According to Nass et al. (1999, p. 1098) the goal was to “ensure that subjects felt they were interacting with the computer rather than simply being passive readers.”

It could be that what Couper et al. did to increase the sense of interactivity in their experiments, namely to fill the user's name and content of earlier responses into the questions, did not give individual respondents the sense that the computer was designing its interaction specifically for them. Computerized questionnaires are, in fact, highly interactive in the sense that the particular set of questions asked of any one respondent depends on previous answers and may be unique. Perhaps if this tailoring of question choice and sequence were made more explicit it would lead to more evidence of social presence.

Although Couper et al. did not detect differences in socially desirable responding among the various groups in their experiments, this does not necessarily mean there were no such effects. It just means that any effects were the same for all four groups. The fact that all of the groups completed the questionnaire on a remote computer could increase

---

<sup>1</sup>In fact, the scientifically constructed samples in the Couper et al. paper are a major advance over the convenience samples used in laboratory studies which report social presence effects, such as those of Reeves & Nass, 1996 and others.

socially desirable responding for all of them. Moon (1998) found more impression management and socially desirable responding (using the same scales<sup>2</sup> as Couper et al.) when the host computer appeared to very remote (3000 miles away) than when it appeared to be nearby (a few miles away) or a standalone machine. Respondents in the Couper et al. study surely perceived the server to be remote which may have led to greater impression management and socially desirable responding in all versions of the questionnaire than would have been observed in corresponding instruments administered on standalone computers. Thus an additional control condition in which it is clear to respondents that there is no network connectivity could help tease this apart. If the “unwired” groups scored lower on Impression management and socially desirable responding than their “wired” counterparts, this explanation would seem to hold.

Of course, it is possible that Couper et al. detected no evidence of social presence because there really is none. By this view, the kind of effects that fill the Reeves & Nass (1996) book are confined to laboratory studies in which convenience samples of subjects carry out special set-up tasks unlikely to occur under ordinary conditions of survey administration. I think this is partly right but that there is something to the Reeves & Nass (1996) kind of finding. One way to reconcile social presence effects in the laboratory with their absence in the current research is to acknowledge that people are ordinarily sensitive to the agent-like character of computers and in some ways treat them like people (e.g. pleading with computers not to crash before a save command is completed). But people know the difference between computers and people and under circumstances where this difference matters – such as reporting sensitive material – the inanimacy and privacy of the medium outweighs its social character. When web survey respondents report about sensitive topics they suspend the perception of the computer as a social agent.

Clearly this work has important implications for collecting information on sensitive topics but does that limit its usefulness for Federal statistical agencies? My sense was that Federal surveys overwhelmingly concern mundane facts and behaviors about which respondents are unlikely to be sensitive. However, if one steps through the “A to Z” topic index on the FedStats web site, many of these topics are potentially sensitive (see Table 1). Furthermore, self-presentation concerns may be more relevant for reporting mundane behaviors than is typically assumed. In a recent study we (Conrad & Schober, 1999) report that for mundane concepts like “more than one job,” “household furniture,” and “live in this house,” respondents were more likely to request clarification from a computer than from an interviewer. In part this may be because it is easier to click a mouse on highlighted text than to formulate and utter a question. But it may also reflect less shame in indicating confusion about everyday concepts to a computer than to an

---

<sup>2</sup>Self-reports of impression management and socially desirable responding are tricky. If someone tells you they are unlikely to be honest under the very circumstances in which they are telling this to you, it’s hard to know if they are currently being honest. For that matter, if they tell you they are likely to be honest under particular circumstances, they could be indicating this for reasons of impression management or social desirability. For current purposes, I accept the validity of these measures, but believe their use deserves additional scrutiny in the future.

interviewer. This was a small scale laboratory study and it's not clear whether the results will scale up to large web-based surveys. But the lack of social presence effects in the Couper et al. research bodes well for the benefit of computerized self-administration of both mundane and sensitive questions when asked in web surveys.

Table 1. Potentially sensitive topics and Federal agencies that collect information on those topics.

<u>Topic</u>	<u>Agency</u>
Abortion	NCHS
AIDS and STDs	NCHS
Crime victimization	BJS, OJJDP
Criminal offenses	BJS, OJJDP
Divorce	NCHS
Drug abuse	SAMHSA
Educational Assessment	NCES
Immigration status	INS
Income	BLS, Census
Job loss	BLS

#### References

- Conrad, F.G. & Schober, M.F. (1999). A conversational approach to text-based computer-administered questionnaires. In *Proceedings of the Third International ASC Conference*. Chesham, UK: Association for Survey Computing, pp. 91-101.
- Moon, Y. (1998). Impression management in computer-based interviews: The effects of input modality, output modality and distance. *Public Opinion Quarterly*, **62**, 610-622.
- Nass, B. & Reeves, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Palo Alto, CA and Cambridge, UK: CSLI Publications and Cambridge University Press.
- Reeves, C., Moon, Y. & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Social Psychology*, **29**, 1093-1110.
- Schaeffer, N.C. (2000). Asking questions about threatening topics: A selective overview. In A. A. Stone, J. S. Turkkan, A. A. Bachrach, J. J. Jobe, H. S. Kurtzman & V. S. Cain (Eds.), *The Science of Self-Report*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers, pp. 105-121.

- Tourangeau, R. & Smith, T.W. (1996) Asking sensitive questions: The impact of data collection mode, question format and question context. *Public Opinion Quarterly*, **60**, 274-304.
- Turner, C. F., Forsyth, B. H., O'Reilly, J.M., Cooley, P.C., Smith, T.K., Rogers, S.M., & Miller, H.G. (1998). Automated self interviewing and the survey measurement of sensitive behaviors. In M. Couper, R. Baker, J. Bethlehem, C. Clark, J. Martin, W. Nicholls, & J. O'Reilly (Eds.), *Computer Assisted Survey Information Collection*. New York: John Wiley & Sons, pp. 455-473.



Discussion of paper: Social Presence in Web Surveys  
by Roger Tourangeau, Mick P. Cooper, and Darby M. Steiger

David Mingay, Statistical Research Division, U.S. Census Bureau,  
Washington, DC 20233-9100. david.j.mingay@census.gov

## BACKGROUND

The research presented in this paper is just a small part of an important and innovative series of experiments that Tourangeau and Couper are conducting under their NSF grant. These projects should significantly expand our knowledge of how interface design affects survey responding. Knowing which design features have positive and negative effects on data quality should help Web designers to design survey instruments that collect high quality data and help those using the data to assess its quality. The paper follows the Cognitive Aspects of Survey Methods (CASM) tradition of taking a theory that was developed in a different domain and testing predictions derived from it when applied to a survey setting.

Web technology provides a substantial opportunity to add a variety of visual features to the survey instrument, including features that make the survey feel more personal. We know little about which features are helpful, harmful, or neutral in their effect on data quality. On the one hand, adding certain visual features to the Web questionnaire, including personalization features, may be helpful. Respondents may perceive the survey as being more interesting, make greater effort to answer the questions accurately, and be less likely to break off the interview. Those who do not complete the interview in a single setting may be more likely to complete it later.

But there can be disadvantages with adding visual features. For example, visual images take time to download, and the added time may frustrate the respondent—particularly if they have a 14K, 28K, or even 56K modem. Interactive features may require the respondent to have Flash or other software installed on their computer, which many people do not have. And adding features often introduces visual clutter and may distract the respondent from answering the questions in a considered manner. Most importantly, some features can affect the interpretation of questions or cause other response problems (Couper, in press).

A recent study of Web TV panel members conducted by the authors illustrates the last point (Kenyon, Couper, and Tourangeau, 2001). They looked at a feature that is being used in some commercial Internet surveys, namely pictures that illustrate the topic of the questions. Questions about the frequency with which panel members engaged in six activities (e.g., overnight trips in the last year) were accompanied by either no picture, a picture showing an example of one way of doing the activity (e.g., a businessman at an airport), a picture showing an example of a different way (e.g., a family in a station wagon), or both pictures. Nineteen of the 54 two-way comparisons showed statistically significant differences ( $p < .10$ ). Most notably, on all six comparisons between the two pictures illustrating an activity, one picture resulted in a significantly higher reported frequency for the activity than the other picture. In addition, asking the same question

with no picture versus one picture produced statistically significant mean scores in six of the twelve comparisons. These results suggest that the visual material accompanying a question can elicit different responses to that question.

In the paper presented in this volume, Tourangeau and Couper looked at two other types of features that are commonly found on questionnaires administered on the World Wide Web to see if they resulted in a more socially desirable response to questions. The features selected allowed the testing of predictions derived from social interface theory. This theory suggests that when the interaction with the computer has some of the features of an interaction with a human, responses similar to those elicited in a human-to-human interaction are obtained (Kiesler and Sproull, 1977). In contrast, the survey research literature suggests that computer-administration of surveys on highly sensitive topics reduces or eliminates the social desirability effects found with human administration of questions even when such humanizing features as a voice are used. This study explored the apparent contradiction between the social interface and survey methods findings.

Two characteristics of interactions were manipulated, personalization and interaction. The level of personalization was varied by presenting a study logo or an image of a male or female researcher along with text of relevant statements from the researcher. The level of interactivity was varied by changing the language used. For example, the high interaction version used the first person in introductions and transitional phrases and occasionally echoed back one of the respondent's earlier answers. Two Web surveys with identical designs were conducted. One survey was completed by 202 members of a Gallup Organization Web panel and the other by 3047 members of a sample of Web users purchased from Survey Sampling, Inc. (SSI).

## RESULTS

Noting that the smaller study had very similar findings, the authors only report the results of the larger study. Only one finding that offers some limited support for social interface theory is reported. Respondents of both sexes showed the most pro-feminist attitudes when the program displayed pictures and messages from the female investigator and the least pro-feminist attitudes when the pictures and messages were from the male investigator, with the attitudes reported by the group getting the survey logo falling between the two. Very few of the numerous other predicted interactions were observed (Tables 2 and 3 shows 16 of the results) and one—the effect of personalization on reports of having voted in the last election—was in the opposite direction to that predicted by social interface theory. Thus, neither level of personalization nor level of interaction demonstrably affected reports about sensitive topics. This is consistent with the survey literature, but not with social interface theory.

## DISCUSSION

Why were the results predicted by social interface theory not found in the two surveys? The authors appear to favor the idea that the sterile, controlled environment of the laboratory in social interface studies focuses subjects' attention on the experimental manipulation. In contrast, their two Web surveys involved a much more uncontrolled real world setting with many potential distracters.

I would suggest that respondents' lack of attention to the experimental manipulation might also be due to low involvement in the surveys. Tentative evidence for low involvement is provided by the fairly poor response rate: only 20% of people in the SSI panel who received an email invitation to participate and one email reminder did so. In addition, 12.5% of those starting the interview failed to complete it.

Uncontrolled real world settings with many potential distracters are intrinsic to Web surveys (Couper, in press) and respondent uninvolvement seems likely to be true for most, but not all, Web surveys. This suggests that these results may apply to other Web surveys as well. Several staff at commercial Web survey organizations have told me that their panel members often sign up to be on several other panels and receive numerous invitations to take surveys. They often chose to answer a questionnaire to be entered in a drawing for a prize or to get a small fee and want to finish as quickly as possible. In the terms of the Krosnick and Alwin (1980) model of survey responding, most respondents probably show considerable satisficing behavior. In addition, the experience of taking numerous prior surveys may affect how panel members respond to survey questions—perhaps making them less attentive to subtle cues, for example.

Of course, interactivity and personalization may commonly affect reporting, but the design of this study prevented the effects from being observed. The authors point out that with 3047 people completing the questionnaire in the second study, there was ample statistical power. However, the use of the first person and transitional phrases seems a relatively weak way to increase interactivity (although many of the manipulations used successfully in social interface research appear even more subtle). Other procedures might have resulted in more socially desirable responding in the high interactivity and/or personalization conditions.

I have some concern about using the method of echoing back certain answers (e.g., “According to your responses, you exercise once daily”) to increase the level of interactivity. While that does increase the interactivity of the interview, it also conveys other information to the respondent—for example, that the answers are particularly important. Respondents may then feel more pressure to answer accurately and may show fewer tendencies to bias their answers in a socially desirable manner. Thus it is possible that providing feedback as one of the interactive features may have somewhat reduced the effectiveness of the high interactive version of the interview for increasing social desirability biases.

There is also some evidence that the results of this research may not be entirely reliable or generalizable. In a recently completed third study that was mentioned in the talk, high interactivity and personalization was associated with more socially desirable reporting in a number of instances.

However, the findings are consistent with other evidence that computer-administration reduces social desirability effects even when there is quite a strong degree of personalization. For example, methodological research by RTI using telephone audio

computer-assisted self-interviewing (T-ACASI) and interactive voice response (IVR) administration methods for federal drug surveys found high levels of reporting of sensitive information despite the significant level of personalization associated with using a voice to administer the questions (Turner et al., 1998).

## CONCLUSIONS

This well-designed study addresses the important issue of whether greater personalization and interactivity in a Web survey increases social desirability effects in reporting. There was little evidence for this although an additional study that was mentioned in the talk provides somewhat contradictory findings. Several other studies being conducted as part of this NSF-funded project address related issues. Taken together, these studies should provide considerable insight into whether, as is postulated by social interface theory, humanizing cues in a computer interface can evoke reactions similar to those produced by a human, such as social desirability response effects, and, perhaps, under what conditions the effects are produced.

A particularly noteworthy feature of the research is that it investigated two important dimensions of surveys that have been neglected by researchers. Web survey designers often use personalization, interactivity, and other features in their surveys, believing that they may increase the respondent's involvement in the questionnaire, and thus improve reporting accuracy and reduce breakoffs. Research of this type is very important for determining whether introducing such features improves, harms, or does not affect the quality of the survey data.

## REFERENCES

- Couper, M. (in press). The promises and perils of Web surveys. In A. Westlake (ed.). *The challenge of the Internet*. Association for Survey Computing.
- Kenyon, K., Couper, M., and Tourangeau, R. (2001). *Experiments on visual effects in Web surveys*. Paper presented at the meeting of the American Association for Public Opinion Research, May 17-20, Montreal, Canada.
- Kiesler, S., and Sproull, L. (1997). 'Social' human-computer interaction. In B. Friedman (ed.). *Human values and the design of computer technology*. Stanford, CA: CSLI Press.
- Krosnick, J., and Alwin, D.F. (1987). *An evaluation of a cognitive theory of response order effects in attitude measurement*. *Public Opinion Quarterly*, 51, 201-219.
- Turner, C.F., B.H. Forsyth, J. O'Reilly, et al. (1998). Automated self-interviewing and the survey measurement of sensitive behaviors. In M. Couper (ed.). *Computer-Assisted Survey Interviewing Collection*. New York: Wiley.