

Social Presence in Web Surveys

Roger Tourangeau
University of Michigan &
University of Maryland

Mick P. Couper
University of Michigan

Darby M. Steiger
The Gallup Organization

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.

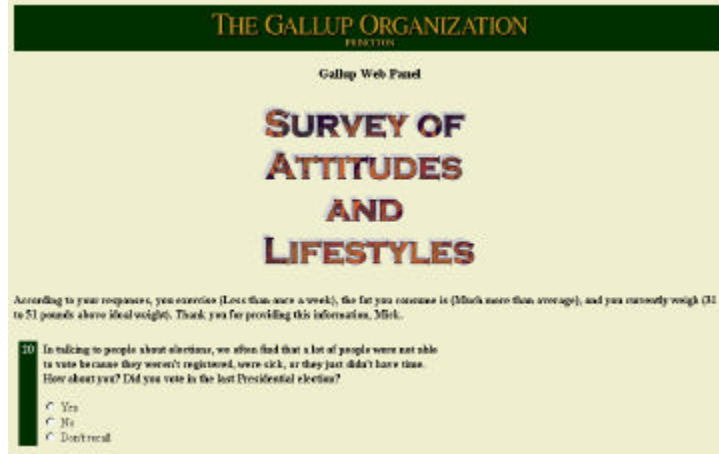


Figure 1: Logo and Personal Feedback

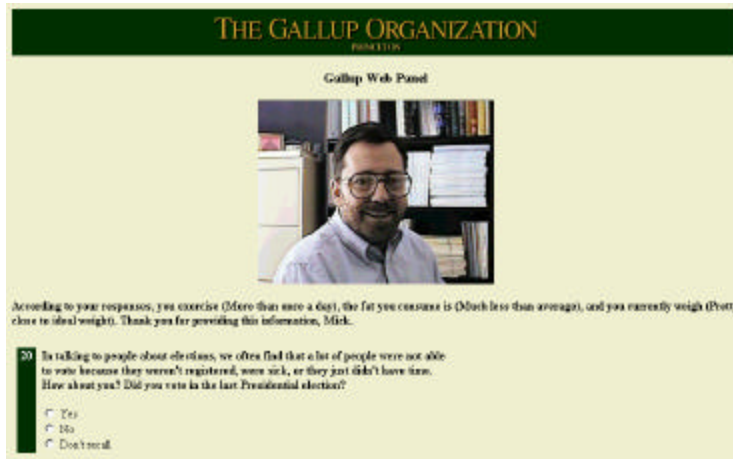


Figure 2: "Male" Interface and Personal Feedback



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

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

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)

	Social Desirability	Impression Management	Gender Attitudes	Sensitive Admissions
Interaction	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)
Personalization	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

	% Used Cocaine in Lifetime	% Smoked Marijuana in Last Year	% Drink Daily or Almost Daily	% Attended Church Last Week	% Voted in Last Election
Interaction	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
Personalization	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.