PERGAMON

Computers in Human Behavior 16 (2000) 553–574

Computers in Human Behavior

www.elsevier.com/locate/comphumbeh

Interactivity in human–computer interaction: a study of credibility, understanding, and influence

J.K. Burgoon [a], J.A. Bonito [a,*], B. Bengtsson [b], C. Cederberg [b], M. Lundeberg [c], L. Allspach [d]

[a] *Department of Communication, University of Arizona, Communication Building #25, Room 209, Tucson, AZ 85743, USA*
[b] *Umeå University, Sweden*
[c] *Royal Institute of Technology, Stockholm, Sweden*
[d] *Western Illinois University, USA*

Abstract

Advancements in computer technology have allowed the development of human-appearing and -behaving virtual agents. This study examined if increased richness and anthropomorphism in interface design lead to computers being more influential during a decision-making task with a human partner. In addition, user experiences of the communication format, communication process, and the task partner were evaluated for their association with various features of virtual agents. Study participants completed the Desert Survival Problem (DSP) and were then randomly assigned to one of five different computer partners or to a human partner (who was a study confederate). Participants discussed each of the items in the DSP with their partners and were then asked to complete the DSP again. Results showed that computers were more influential than human partners but that the latter were rated more positively on social dimensions of communication than the former. Exploratory analysis of user assessments revealed that some features of human–computer interaction (e.g. utility and feeling understood) were associated with increases in anthropomorphic features of the interface. Discussion focuses on the relation between user perceptions, design features, and task outcomes.

Keywords: Human–computer interaction; Interactivity; Decision making; Anthropomorphism; Contingency

Modern information technology now affords organizations, businesses, individuals, and institutions of learning a variety of options for engaging in communication

* Corresponding author. Tel.: +1-520-621-1025.
E-mail address: jbonito@u.arizona.edu (J.A. Bonito).

0747-5632/00/$ - see front matter
PII: S0747-5632(00)00029-7

and information exchange. One such innovation is the augmentation of human sources of information with intelligent computer agents — computer interfaces that come in a variety of guises and that present and process information according to a set of predefined algorithms. Agents may be designed to appear more anthropomorphic by fitting them with distinctly human-like (virtual) features such as voice recognition, synthesized voices, and computer animation that simulates human facial expressions and gestures. Doing so often compels human partners to think of the computer in more social terms, for example, as being a member of a team (e.g. Nass, Fogg & Moon, 1996).

Other advances in artificial intelligence enable more adaptive responses to human input which will no doubt increase the perceived social character of computer agents. For example, Nongjian (1999) describes how computers might be fitted with a "core ego" whereas Brazier and Treur (1999) discuss methods for building a reflective agent. Other researchers envision the day when computers will be able to recognize a user's intention from his or her linguistic input, making the computer more responsive to user commands and desires (McKevitt, Partridge & Wilks, 1999). Clearly, making the computer more human externally, internally, and in action will affect how humans use and work with them.

While design options for computer agents are rapidly increasing, few studies have systematically and comprehensively compared multiple human–computer interaction (HCI) interfaces to face-to-face (FtF) interaction to determine which features affect the process and products of work. This report, one from a planned program of research (Bonito, Burgoon & Bengtsson, 1999; Burgoon, Bonito, Bengtsson, Ramirez & Dunbar, 2000), tests the properties of *interactivity* that interfaces may afford. We explicate the multidimensional nature of interactivity then present results of an experiment manipulating three of its properties — contingency, modality richness, and anthropomorphism — for their impact on task-related interaction and outcomes. Specifically, we examine: (1) if human or computer partners (with varying interfaces) are deemed more or less credible; (2) if understanding and information processing differ as a function of interacting with a human or computer partner; and (3) if human or computer partners are more or less influential.

1. Credibility, understanding, and influence

Our analysis of persuasion process and outcomes is based on McGuire's (1985) model of persuasion in which understanding mediates the effect of source credibility on influence. Credibility refers to the recipient's judgment that a message and/or its source are believable and convincing. Understanding a message and assigning credibility to it or its information source are thought to be prerequisites to message or information acceptance. Recipients who fail to comprehend or recall a piece of information or who believe that it or its source is untrustworthy are unlikely to incorporate such information and arguments into their own decision-making.

Understanding comprises both the accurate decoding of information contained in a message and appropriate inference making — drawing the information intended by the source. Whereas the former may be assessed according to a recipient's recall of specific factual details of denotative content, the latter may be evaluated by examining the recipient's ability to paraphrase the *meaning* of a message, to recapitulate its connotative implicature. The latter represents a deeper level of understanding than the former although both are doubtless necessary conditions for long-term message retention and influence.

Credibility is also multidimensional; it is a constellation of judgments related to perceived competence, character, composure, dynamism, and sociability. Typically, communicators are more likely to influence others if they are more credible (Burgoon, Pfau & Birk, 1990), but different dimensions of credibility may operate in different ways. A communicator or message may acquire its credibility by virtue of conveying expertise and authoritativeness, in which case the competence dimension is responsible. A communicator may be believable because he or she is seen as truthful, trustworthy, and reliable, in which case the character dimension is implicated. Credibility may be gained by virtue of the sender appearing dominant and extroverted (part of the dynamism dimension) or instead as seeming likable and friendly (part of the sociability dimension).

The applicability of this model to HCI depends on the notion that participants make credibility judgments about computer interfaces in much the same way as they judge human partners. This implies that humans think of computers in fundamentally social ways and that such assessments influence communication processes by mediating between communication inputs and outcomes. In the following section we describe current thinking and research on the relationship between humans and their computer partners.

2. Computers as social actors

People respond to computer artifacts, media, and other communication technologies in fundamentally social ways. Marakas, Johnson and Palmer (2000) argued that people develop and maintain attributions of computers in roughly the same way as people make attributions about each other. Inputs to the computer attribution process are: (1) the social character of the technology; (2) an individual's core self-evaluations (e.g., self-esteem or locus of control); (3) the nature of and the context in which interaction with a computer takes place; and (4) the availability of attributional information cues. The degree to which attributions are made and their quality (e.g., internal or situational) depends in part on how actors see themselves vis-à-vis the computer.

Several studies highlight the way in which persons attribute social or human characteristics to computers. For example, Nass and colleagues (e.g., Nass et al., 1996; Nass, Steuer, Henriksen & Dryer, 1994a; Nass, Steuer & Tauber, 1994b) conducted a set of studies that examined a host of user perceptions in HCI. Among the findings were that users applied politeness norms, notions of "self" and "other," and gender stereotypes when interacting with computers. Also, subjects who were told they were interdependent with a computer formed a connection with the computer as a team member, not because of a mistaken belief that computers are human-like or that they act as proxies for human programmers, but because it is a natural psychological tendency to process such social information in the same way as it is processed when emanating from humans.

Furthermore, a limited set of interface characteristics normally associated with humans provides sufficient cues to encourage users to exhibit social behavior toward their computer partner. Morkes, Kernal and Nass (1999) noted that computers that were designed to be more humorous were perceived as likable, and participants were more likely to joke with and act more sociable toward the computer. Other researchers found that people were more likely to exhibit impression management concerns with a talking face than with computer text (Sproull, Subramani, Kiesler, Walker & Waters, 1996). In an interview conducted by a computer "counselor," participants revealed less to the talking face than to the text and evaluated it less positively. These evaluations were made on personality attributes that research has shown are affected by people's physical appearance and voice (Ekman, 1982; Warner & Sugarman, 1986). Finally, expert systems that were perceived to be more confident were more influential on a problem solving task (Jiang, Klein & Vedder, 2000).

In a similar vein, research within mass media has shown that people frequently develop a personal sense of connection and involvement with media celebrities, television characters, and others who are portrayed or featured in the media (Alperstein, 1991; Turner, 1993). Through parasocial relationships, people come to rely on their exposure to or interactions with these figures for some of their social needs and in so doing may subconsciously lose sight of the fact that characters are not real people or that they are not acquainted with these people personally. In this way, what is presented in the media becomes "real". For example, even skeptical viewers have been shown to form their construction of reality by blending what they actually witness on the media with prior media exposures and gossip (Perrolle, 1987). From this we can surmise that computer agents may be the beneficiaries of similar tendencies.

Other research has documented the power of media to confer credibility and exert influence (Shapiro & McDonald, 1992). The tendency for people to imbue media with undue credibility is perhaps nowhere more evident than with inexperienced computer users, who seem to attribute more validity to computer-generated and computer-presented information than is warranted (e.g. Harmon, 1996). It is as if the "ghost in the machine" has conferred special authoritative status on such sources, perhaps because it is presumed that any information appearing in broadcast or computer-mediated form has already been through appropriate gatekeepers who have verified its accuracy, relevance, validity, and appropriateness. Therefore, its very availability in mediated form is taken as prima facie evidence of its authenticity. These findings imply that mediated communication forms in general should be the beneficiary of a positivity bias whereby information that is delivered through a mechanical or electronic medium is regarded de facto as credible.

Other media research alternatively suggests a gradient of increasing credibility as a medium becomes "richer" in sensory channels and amount of social information it supplies; credibility increases as more modalities become available (a property we discuss shortly as a feature of interactivity). People trust pictures more than the printed word and they are more likely to trust a television image than a newspaper article (Rubin & McHugh, 1987). Thus, there is reason to believe that the addition of computer graphics will increase trust in computer-generated and computer-presented information.

In sum, there is a tendency for persons to attribute credibility to computer artifacts and other aspects of media. It stands to reason that such attributions are responsible for outcomes in HCI simply because interpersonal and social assessments are generally consequential for human interaction and because computer agents unconsciously elicit the same kinds of social judgments that human sources and other technologies do. This raises the question of whether the credibility attributed to computers, media, and mediated sources is due to heuristically attributing human-like characteristics to them. If so, then it would follow that the more anthropomorphic or richer a computer agent becomes, the more likely that humans will understand and believe it and thus be influenced by it. However, it is also plausible that computer agents may gain influence by departing from the human prototype and capitalizing on those special properties that confer the impression of authoritativeness and expertise on anything related to technology. To address these alternative possibilities requires a closer look at the properties of communication, influence, and the kinds of social judgments that might be enabled or encouraged by each. This brings us to the principle of interactivity.

3. Structural affordances in HCI: the role of interactivity

An important basis for human interaction and for HCI is the structural affordances that interactive situations offer communicators. These are features that are "built into" or intrinsic to a given communication mode or format. They are also the kinds of properties that might be engineered into a given communication system that make interaction possible. St. Amant (1999) opined that affordances are: (1) the properties between an agent and his or her surrounding environment; (2) potential actions supported by the environment; (3) perceived properties of an object; and (4) mental constructs that are subjective in nature. All these characteristics of affordances are important to interaction processes in general because they allow or inhibit the transmission and reception of important information on which assessments of the interaction and the persons involved in it are based.

Varying conceptualizations have been offered for what is meant by interactivity (e.g., Bonito et al., 1999; Burgoon et al., 2000). We conceive of interactivity both according to a number of structural properties that afford or permit interactivity and according to the processes themselves that are the phenomenological experience of interacting. Structurally, communication formats may be arrayed along a continuum from high to low interactivity such that formats affording more potential for interactivity are:

(1) *contingent*: each contribution is dependent on prior contributions, so that the flow of discourse is sequenced and coherently related but somewhat

unpredictable, with meanings managed "locally" (Sacks, Schegloff & Jefferson, 1978);

(2) *unmediated*: interactants are physically co-present and messages are exchanged orally rather than transmitted via some electronic or mechanical medium;
(3) *propinquitous*: participants are in the same location rather than geographically dispersed;
(4) *synchronous*: interaction occurs in real time rather than being delayed or asynchronous;
(5) *participative*: all actors are both senders and receivers of verbal and nonverbal messages and feedback, rather than senders transmitting one-way messages or receivers passively witnessing another's communication;
(6) *modality- and information-rich*: participants have full access to a wide array of environmental, visual, audio, verbal, and other sensory information;
(7) *anthropomorphic*: the guise in which an interface appears resembles humans; and
(8) *identified*: participants are known to one another rather than anonymous.

Traditional FtF dialogue is the current prototype for highly interactive encounters because it is contingent, unmediated, same-place, same-time, participative, informationally rich and complex, and identified.[1] Moving down the continuum would be various mediated and distributed formats, with asynchronous electronic correspondence among unacquainted members of distributed work teams being a less interactive mode of interaction. The degree of interactivity may also vary within a given communication format. Televised broadcasts, for instance, are usually non-interactive, but the call-in talk show is an exception.

An alternative way to conceptualize interactivity is according to how it looks and feels, that is, according to the indicators by which we "know" that a situation is interactive. In essence, these are qualities of the interaction process itself that flow from the structural affordances. From this perspective, a communication environment or format is perceived as more interactive to the extent that it manifests greater:

(1) *individual involvement* (high cognitive, sensory, visceral, and motor engagement, i.e. a sense of presence, of "here and now");
(2) *mutuality between individuals* (a sense of "connectedness," interdependence, receptivity, collective sense-making, shared understandings, and coordinated interaction); and
(3) individuation (well-defined notions of "me," "you" and "us" rather than vague identities and pseudo- or imagined relationships).

These properties are interrelated. For example, normal interaction is comprised of the identities of individuals involved in interaction (Goffman, 1967). Identity creates

[1] It should be noted that although FtF interaction is considered prototypical of high interactivity, technologies of the future, such as immersive virtual reality, may eventually exceed the capacity of FtF interaction to maximize certain interactive properties.

an impression of the social, which in turn engenders feelings of engagement or connectedness. It allows individuals to create important representations of their interlocutors, and such representations influence subsequent interaction. Features such as participation, nonmediation, propinquity, synchronicity, modality richness, identification, and contingency are presumed to promote greater involvement, mutuality, and individuation.

4. Hypotheses and research questions

We begin this section by asking if humans and computers differ in terms of credibility and influence. As noted above, there is a tendency for humans to attribute human qualities to media and that in some forms of HCI computers can be as influential as human partners. In addition, because people often ascribe a sense of infallibility to information presented through or by computers, it is possible that computers are more influential than human partners, all other things being equal. Therefore, our two research questions are stated as follows:

RQ1: Do human and computer partners differ in terms of credibility?
RQ2: Do human and computer partners differ in terms of the amount of influence they exert on a decision-making task?

Assuming that human and computer partners differ in the amount of credibility attributed to them as well as in the amount of influence they exert, we are concerned with identifying the underlying characteristics responsible for these differences. Our hypotheses are derived from three of the eight features of interactivity listed above. First, we surmise that social judgments and interaction outcomes are affected by contingency. Human interaction is contingent such that a given utterance or message is relevant on several levels to previous turns of talk (Sanders, 1987, 1997). For example, interactants generally take speaking turns in orderly fashion (i.e., they avoid talking "over" one another), provide structurally relevant forms (e.g., follow questions with answers), and can produce and respond to messages whose meanings and functions are indirect or concealed (e.g., by offering to close a window as a response to "It's cold in here"). We assume that contingent interaction, because it meets expectations regarding interaction, leads to higher credibility, better understanding, and greater influence than less contingent formats:

H1a: Judgments of credibility will be greater in highly contingent interaction formats than in minimally contingent formats.
H1b: Understanding in highly contingent interaction formats will be greater than in minimally contingent formats.
H1c: Influence in highly contingent interaction formats will be greater than in minimally contingent formats.

Modality richness is the second feature of interactivity upon which our analyses are based. Humans supply words, visual cues, and vocal cues, among others. When they are all present in HCI, computers may simulate human interaction which in turn should lead to higher judgments of credibility, greater understanding, and increased influence. There is some evidence for this assertion in a recent survey of the literature on the effects of animated agents by Dehn and van Mulken (2000). They noted that some features of interaction (e.g. problem solving and believability) are enhanced when the agent is animated compared to when it is static. Our second set of hypotheses concerns the addition of modalities to computer interfaces.

H2a: Participants' credibility ratings of the computer agent increase as modalities are added to the interface.
H2b: Participants' understanding increases as modalities are added to the computer interface.
H2c: Computer agents will be more influential as modalities are added to the interface.

Configuration is the third feature of interactivity used in this study and it drives our last set of hypotheses. An issue is whether adding text to the most anthropomorphic interfaces (thereby making them less human-like) affects HCI processes and outcomes. The addition of text might in fact make the information environment too complex; participants would have difficulty simultaneously attending to aural, nonverbal, and textual data. On the other hand, text might actually enhance the HCI experience under these circumstances by adding redundancy of information, the "recoverability" of information that texts affords, and multiple options for information reception that cater to individual differences in channel preference. This question necessitates a nondirectional set of hypotheses:

H3a: Participants' credibility ratings of an anthropomorphic computer agent will differ when text is present or absent in the interface.
H3b: Participants' understanding will differ when text is present or absent from an anthropomorphic interface.
H3c: Influence by anthropomorphic computer agents will differ when text is present or absent from the interface.

5. Materials and method

5.1. Subjects

Subjects ($n = 70$) were male undergraduate students recruited via advertisements from the social science department at Umeå University, Sweden. The advertisement described the research as a study of alternative problem-solving methods. Participants were paid 100 Swedish crowns (approximately $10) for their participation.

5.2. Independent variables

The experimental conditions represent our manipulations of the three independent variables.

5.2.1. HCI conditions

Five HCI conditions were created. Importantly, although the interfaces differed in terms of modality or level of anthropomorphism, the information presented in all five conditions did not vary; it was presented according to a fixed interaction script. Participants entered information via the keyboard into a text box. The computer produced its next turn only after the participant pressed the "enter" key to submit his or her contribution.

The screen configuration for all five interfaces is presented in Fig. 1. The five conditions were: (1) text-only; (2) text and synthesized voice; (3) text, voice, and still image (Fig. 2); (4) voice and animation (with the same image as in the previous condition but with facial features and a mouth that "moved" in sync with the synthesized voice); and (5) text, voice, and animation.[2] Thus, the five conditions represent increases in modalities and/or anthropomorphism. The last was the richest in having these modalities present but less anthropomorphic than the fourth because it had less resemblance to human interaction by virtue of the presence of text. It should be noted

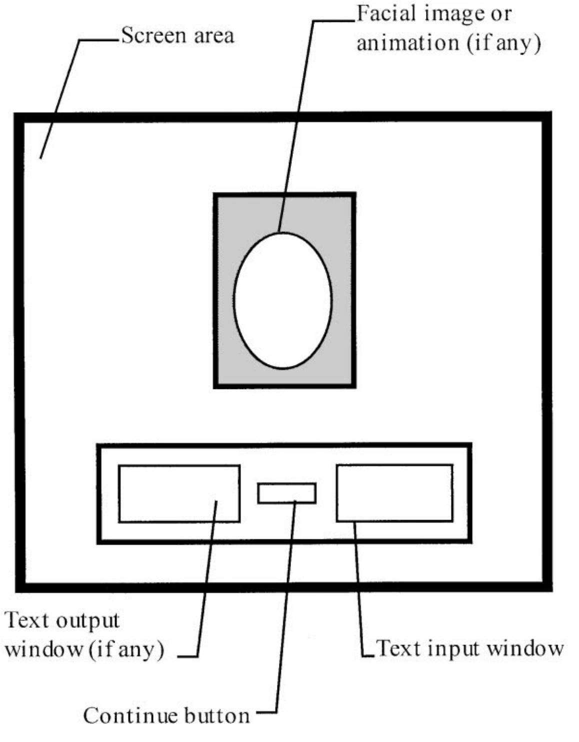

Fig. 1. Screen configuration for human–computer interaction conditions.

[2] The speech synthesis software used in this study was a development of the KTH text-to-speech synthesis project, described in Carlson, Granström and Hunnicut (1982). The animation software used in this study was developed by Jonas Beskow and Magnus Lundeberg at KTH (Royal Institute of Technology), Stockholm, Sweden, and is further described in Beskow (1995) and Lundeberg (1997).

Fig. 2. Facial image for the relevant human–computer interaction conditions.

that these conditions were not crossed with. Because we had no real or convincing way to create interfaces that were fully contingent because responses were processed in advance; thus, the HCI conditions were minimally contingent. However, the FtF conditions, discussed next, covered the contingency comparisons.

5.2.2. *Face-to-face conditions*

We employed experimental confederates in the two FtF conditions. The confederates presented the same information in the same order that was used in the HCI conditions. In the *minimally contingent* condition, the confederates were instructed to follow the script (the same one programmed into the computer agents) as closely as possible and to disregard irrelevant or off-task remarks from the participants. This created the greatest comparability to the HCI conditions but was highly unnatural in the FtF interaction. In the contingent condition, confederates were instructed to follow the script as closely as possible but were also asked to respond relevantly to any remarks made by the participants.

The analysis plan, discussed below, contrasted various combinations of these seven conditions. Four of the five HCI conditions (excluding text, voice, and animation) and the minimally contingent FtF condition represent the modality richness manipulation (with the latter being the richest). Because one of the speech and animation interfaces had text and the other did not, those two conditions represent our anthropomorphism or configuration manipulation. Finally, the FtF conditions, in addition to representing high richness and anthropomorphism, constituted the contingency manipulation.

5.3. *Dependent measures*

Dependent measures were derived from the post-interaction questionnaires. They included three different assessments of influence, partner's credibility, understanding, assessment of the communication format, assessment of the interaction, and task attraction. Each is described below.

5.3.1. *Influence*

The first measure of influence, *absolute influence*, was computed by calculating the difference between each subject's pre-ranking and post-ranking from expert rankings. The mean of the differences for the post-ranking score was subtracted from the mean of the differences from the pre-ranking to arrive at the absolute distance subjects moved toward the confederate's position.[3] A large distance indicated high absolute influence. The second measure was *relative influence* which reflected the proportion of movement relative to the original distance between the subject and the partner. It was computed as 1 minus the ratio of absolute pre-rank difference to absolute post-rank difference. As with absolute influence, a large distance indicated high relative influence. The reasoning behind calculating this second measure was that, for some conditions, subjects might happen to start out closer to their partner's ranking than in others. If the distance between rankings was small from the start there would be no possibility of a great shift towards the partner's rankings, in absolute terms. Finally, *decision quality* was computed as the mean absolute distance between subject and expert rankings on each item. A small distance indicated high decision quality.

5.3.2. *Credibility*

Participants rated partner credibility along five dimensions: competence, character (which includes trust and honesty), sociability, dynamism, and dominance. Semantic differential items were selected from previously validated measures developed by Burgoon, Johnson and Koch (1998), McCroskey and Young (1981), McCroskey, Hamilton and Weiner (1974), and Wheeless and Grotz (1977).

The competence subscale was comprised of adjectives such as intelligent, informed, experienced, expert, clever, insightful, and imaginative. The character subscale consisted of adjectives such as responsible, sincere, trustworthy, truthful, straightforward, credible, reliable. The sociability subscale consisted of ratings of friendliness and likability. Dynamism items were also subdivided into two separate dimensions because previous research has found dominance to be a judgment relevant to computer agents (Burgoon et al., 2000). The two subscales included adjectives related, respectively, to dominance and confidence and to dynamism, energy, and talkativeness. These measures were scored as the mean item rating (ranging from 1 to 7) for each dimension. Interim reliabilities, calculated with Cronbach's alpha, were 0.81 for character, 0.68 for truthfulness, 0.75 for competence, 0.59 for dominance, 0.68 for dynamism, and 0.86 for sociability.

In addition to the indexes used above, we measured *task-partner attraction* by combining items from Moon and Nass's (1996) questionnaire and items from the

---

[3] The script given to the confederates contained "expert" arguments developed by NASA experts. Therefore, the confederate was an expert whose comments were consistent with the ranking to which participants' rankings were compared. Participants were not told, and therefore unaware, that the confederates were experts on the matter.

task attraction dimension of McCroskey and McCain's (1974) measure of interpersonal attraction (alpha = 0.79).

5.3.3. Understanding

Participants' understanding of the information presented by confederates was evaluated along two dimensions. First, *accuracy of recall* was assessed by asking subjects to recall the partner's ranking for the three top-ranked and three bottom-ranked items. Accuracy was scored as the number of correct matches. Second, *content understanding* was measured by asking subjects to state in writing what they believed to be their partner's position on the six middle-ranked items (selected because they might be less obvious and less biased by subjects' own preferences). They were also asked to paraphrase their partner's reasons for their rankings. These responses were rated by two independent coders, after which the results were subjected to an inter-judge reliability test ($r = 0.92$). The individual ratings were averaged to produce a single understanding score.

5.4. Procedure

Upon arrival at the research site, participants completed consent forms and were thanked for agreeing to be recorded. Study participants were randomly assigned to one of the seven communication conditions (text-only; text and voice; text, voice, and picture; text, voice, and lip-synched animation; voice and lip-synched animation; minimally contingent FtF interaction; and contingent FtF interaction). There were 10 subjects in each condition. In the FtF conditions, the partner was a male confederate.[4] In the computer conditions, the partner was a computer named Holger (Fig. 2).

The Desert Survival Problem was chosen as the task because it allows a fair amount of experimental control while still approximating features of normal conversation. In it, participants were asked to imagine that their jeep had crashed in the Kuwaiti Desert, with no sign of potable water but some salvageable items from the wreckage. They rank-ordered 12 items for their survival value: a $20' \times 20'$ piece of blue canvas, a set of ground cloths, a gun, a box of matches, a set of jackets, a flashlight, a knife, a map, a magnetic compass, one bottle of water for each survivor, a book titled "Edible Plants of the Desert", and a rearview mirror. Subjects were further told that a group of survival experts had come up with a set of rankings, based on their expertise, and that subjects would be evaluated on the quality of their final rankings and those of their partner. The instructions informed them that they were "free to convince your partner to change [his/its] ranking when you think it is incorrect." Those in the computer conditions were also told, "This computer program, like all other computer programs subjects will use, does not have access to this information."

[4] This was because the facial images and animations in the computer conditions were those of a male character.

The instructions, description of the Desert Survival Problem, initial rankings, post-rankings, and all other post-measures were posted on the World Wide Web and collected via a Macintosh computer. Participants were seated at a table and first worked alone, generating their own ranking and notes on their selections, then entering their initial rankings in the computer.

The discussions then commenced. In the five computer conditions, the interaction took place on a Hewlett-Packard workstation placed on a table in the middle of the room. In the FtF conditions, there was no other computer on the table. Instead, a human confederate entered the room upon the subject's completion of his or her initial ranking, and was seated on the opposite side of the table. In both the computer and FtF conditions, subjects were videotaped. The video camera was located on a side-table, fully visible to the subjects. The discussion consisted of confederates and subjects alternating turns in discussing each of the 12 objects. The content of the confederates' contributions was strictly controlled and drew upon scripts generated by Nass and colleagues (Nass et al., 1994a, b). To increase the sense of contingent turn-taking, questions were used to elicit subject rankings and reasons, and the interaction concluded with confederates asking subjects if they had any further thoughts to offer. After having ranked the items a second time, subjects completed other post-measures, were debriefed, and received their payment.

6. Results

Our tests of the research questions and the hypotheses were conducted as a series of planned contrasts (Table 1) within a one-way analysis of variance. We evaluated the two research questions using the first two contrasts. Contrast 1 compares the average of the two FtF conditions to the average of the computer conditions whereas contrast two omits the noncontingent FtF condition from the comparison so as to compare HCI to a "normal" FtF condition (given the very artificial nature of the noncontingent FtF interaction). The third contrast was used to evaluate

Table 1
Contrast coefficients for evaluation of hypotheses and the research questions

Contrast	Richness						
	Text-only	Text and voice	Text, voice, and image	Voice and animation	Text, voice, and animation	Minimally contingent face-to-face	Contingent face-to-face
1	−2	−2	−2	−2	−2	5	5
2	−1	−1	−1	−1	−1	0	5
3	0	0	0	0	0	−1	1
4	−3	−1	1	3	0	0	0
5	−1	−1	−1	3	0	0	0
6	−1	−1	2	0	0	0	0
7	−1	1	0	0	0	0	0
8	0	0	0	1	−1	0	0

hypothesis one; it compares the two FtF conditions to examine the effect of contingency (without confounding it with HCI). Contrasts 4 through 7 were used to evaluate the second hypothesis. Contrast 4 corresponded to the trend associated with increased modalities, whereas contrasts 5 through 7 are a series of reverse Helmert comparisons within the four modality conditions. Finally, our evaluation of hypothesis three was based on contrast 8, which asks if adding text to voice and animation alters results. We chose to evaluate the research questions with $\alpha = 0.10$ because of the relatively low power caused by the small sample sizes. The first two hypotheses were evaluated with directional tests at $\alpha = 0.05$, whereas the third hypothesis was again evaluated at $\alpha = 0.10$ because of the nondirectional nature of the tests (Cohen, 1988; Keppel, 1991; Lipsey, 1990).

The variables used in the examination of credibility were task-partner attraction, sociability, expertise, dependability, dominance, and trust. For the influence variables, absolute and relative influence were evaluated, as was decision quality. Content understanding and accuracy of recall were used to assess understanding.

6.1. Research question 1

Our first research question asks if human and computer partners are viewed as more or less credible. There is some evidence that human partners are judged higher on credibility than computer partners. (Means for the credibility measures across the seven interaction conditions are presented in Table 2.) The analysis corresponding to contrast 1 revealed that human partners were rated as more sociable than computer agents, $t(63) = 2.60$, $P < 0.01$, $R^2 = 0.07$. It is important to note that this difference remained unchanged when the second contrast was used (which ignores the

Table 2
Means and standard deviations for credibility measures by discussion condition[a]

	Condition						
	Text-only	Text and voice	Text, voice, and image	Voice and animation	Text, voice, and animation	Minimally contingent face-to-face	Contingent face-to-face
Competence	4.74 (0.87)	4.49 (0.53)	4.78 (0.84)	5.03 (0.63)	4.88 (1.03)	4.46 (0.76)	5.06 (0.67)
Dominance	5.35 (1.08)	4.60 (0.81)	4.95 (0.83)	5.40 (1.26)	5.35 (1.13)	4.95 (1.50)	4.80 (1.14)
Sociability	4.60 (0.88)	4.05 (1.01)	3.75 (1.01)	4.05 (0.86)	4.10 (1.15)	4.65 (0.85)	5.10 (1.26)
Task-partner attraction	4.68 (1.25)	4.96 (1.08)	4.61 (0.80)	4.88 (0.98)	5.29 (0.55)	4.04 (1.05)	4.40 (1.21)
Expertise	4.62 (1.07)	4.32 (0.41)	4.55 (1.25)	4.72 (0.61)	4.58 (1.32)	4.08 (1.06)	5.05 (0.76)
Trust	4.78 (1.15)	4.72 (0.88)	4.85 (0.86)	4.90 (0.63)	5.20 (1.07)	4.83 (0.77)	5.13 (0.64)

[a] Standard deviations are in parentheses.

noncontingent FtF condition), $t(63) = 2.82$, $P < 0.01$, $R^2 = 0.06$. Interestingly, computers were found to be more attractive task-partners than human partners, $t(63) = -2.47$, $P < 0.05$, $R^2 = 0.09$; unlike sociability, this difference disappeared when the noncontingent FtF condition was dropped from the comparison, $t(63) = -1.37$, $P < 0.15$. None of the other contrasts were significant for the other dimensions of credibility.

6.2. Research question 2

The second research question asks if human and computer partners differ in terms of influence. Because the test for heterogeneity of variance was significant for absolute influence (at alpha = 0.25 as recommended by Keppel, 1991), we used the Welsh test to evaluate the contrasts. Doing so revealed that computers are more influential than human partners, $t(19.68) = -1.80$, $P < 0.10$, $R^2 = 0.03$.[5] Neither of the contrasts was significant for relative influence or decision quality.

6.3. Hypothesis 1

The first hypothesis states that contingent interaction leads to higher credibility, understanding, and influence than minimally contingent interaction. Results indicate a modicum of support for the hypothesis.

6.3.1. Credibility

We examined the six dimensions of credibility and found that contingency affected only expertise. Partners in the contingent condition were rated as more expert than partners in the minimally contingent FtF discussions, $t(63) = 2.21$, $P < 0.05$, $R^2 = 0.01$.

6.3.2. Understanding

Because of a technical error we were unable to gather data on understanding from the minimally contingent FtF condition. (Means for understanding for the remaining conditions and for the three influence measures are presented in Table 3.)

6.3.3. Influence

Contingency had no bearing on any of the three measures of influence.

6.4. Hypothesis 2

According to hypothesis two, increased richness in the computer interfaces leads to higher social judgments, greater understanding, and more influence.

[5] The denominator for the Welsh test is derived by summing the products of the squared contrast coefficients and their respective variances and then dividing that quantity by the sum of the squared constrast coefficients (Keppel, 1991). The result in some cases is to allow the variances associated with large coefficients, as is the case here with the contingent FtF condition, to dictate the size of the error term. We caution the reader that the relatively small variance associated with the contingent FtF condition may have overcompensated for the heterogeneity in the variances, making the test somewhat liberal.

Table 3
Means and standard deviations for influence and understanding by discussion condition[a]

	Condition						
	Text-only	Text and voice	Text, voice and image	Voice and animation	Text, voice and animation	Minimally contingent face-to-face	Contingent face-to-face
Decision quality	3.14	2.87	3.23	3.24	3.15	3.34	3.54
	(1.19)	(1.27)	(1.10)	(0.77)	(1.15)	(1.09)	(0.87)
Absolute influence	(0.80)	1.09	1.02	0.91	0.86	0.98	0.51
	(0.87)	(1.24)	(1.07)	(0.95)	(0.73)	(1.13)	(0.61)
Relative influence	0.20	0.26	0.24	0.21	0.23	0.22	0.12
	(0.19)	(0.25)	(0.24)	(0.21)	(0.21)	(0.25)	(0.13)
Understanding	4.80	5.00	5.10	5.18	4.88		5.08
	(0.98)	(0.85)	(0.58)	(0.57)	(1.53)		(0.90)

[a] Standard deviations are in parentheses.

6.4.1. *Credibility*

There is limited evidence that credibility is a function of increased richness across the computer conditions. The contrast (4) for a linear trend in sociability approached significance, $t(63)=-1.36$, $P<0.09$, $R^2=0.02$. Of the reverse Helmert contrasts for sociability, the comparison in contrast 6 revealed that the addition of an image to the interface approached significance, $t(63)=-1.46$, $P<0.08$, $R^2=0.02$. In addition, regarding dominance, we found a significant effect for contrast 7, $t(63)=-1.77$, $P<0.05$, $R^2=0.01$. The addition of speech to the text interface was perceived as less dominant than text alone. No other contrasts were significant for any of the credibility dimensions.

6.4.2. *Understanding*

None of four contrasts for the effect of modality richness on understanding was significant.

6.4.3. *Influence*

The three influence measures were unaffected by modality richness.

6.5. *Hypothesis 3*

The final hypothesis compared the most anthropomorphic interface (the one with voice and synched animation) with one to which text was added. That contrast (8) was not significant for any of the credibility dimensions, understanding, or influence.

7. Discussion

The rapidly accelerating development and diffusion of new communication and information technologies places a premium on assessing the consequences of employing such technologies. Research that systematically disentangles relevant properties of such technologies can inform future design and utilization so as to maximize advantageous uses, minimize undesirable effects, and capitalize on unanticipated benefits. It is toward that end that this investigation was undertaken.

The current results are informative in some senses and mystifying in others. Consistent with the principle of interactivity, partners were seen as more credible, at least in terms of sociability, when participants engaged in FtF than HCI. However, computerized agents were somewhat more influential on the whole than were human partners. Contrary to the hypothesized benefits of greater richness and anthropomorphism, there was a nearsignificant tendency for greatest influence and high-quality decisions to occur in the text and voice condition. This severing of the relationship between influence on the one hand, and credibility and understanding on the other, is puzzling, to say the least. We offer speculation below.

7.1. Credibility

FtF interaction generated higher perceptions of credibility (on the sociability dimension) than did HCI, whereas contingent interaction was rated as more expert than was noncontingent interaction. We also noted a marginal effect of increased richness on sociability (with the main difference occurring when an image was added to the interface). Interestingly, the text-only condition earned the highest ratings on dominance, and the other credibility dimensions showed highly variability across conditions.

Although these findings are puzzling, we offer two explanations. One possibility is that the black sheep effect comes into play as representational richness increases. From this perspective, text-only could be considered neutral, not demonstrating any idiosyncrasies, and therefore not giving rise to any negative attributions by the subjects. The synthesized voice, however, is unique and particular, making possible all kinds of attributions on the part of the subject.

A second possibility — or perhaps a more detailed account of the black sheep effect — is based on Walther's (1996) notion of "hyper-personal communication". In a study of computer-mediated communication, he found that mediated interaction created such a high sense of personalism that actually exceeded FtF communication in creating connectedness, mutuality, and involvement. Such interactions, it is suggested, may result from situations where parties are self-aware, physically separated, and communicating via a limited-cues channel that allows them to selectively self-present and edit. Interaction with limited information allows participants to construct and reciprocate representations of their partner and relationships without the interference of environmental reality; this may be amplified when communication is asynchronous and when the computer-mediated communication link is the only one there is.

7.2. Influence and understanding

The finding that the collective computer versions were somewhat more influential than a human partner is consistent with prior findings that people are more susceptible to influence from mediated messages and information. This cannot be ascribed to computers being seen as more competent per se, because there were no significant differences across conditions in competence ratings. It is possible, however, that even though participants rated computer partners on a par with humans in terms of competence, they were willing to defer to the judgments of the computer partner because they regarded the computer's arguments as better informed than their own opinions.

We have documented elsewhere (Bonito et al., 1999; Burgoon et al., 2000) that several of the credibility judgments are correlated with influence, including task-partner attraction. The difference in influence between human and computer partners might be indirect such that the degree to which an interface fosters higher credibility manifests itself as increased influence. Moreover, our other studies have shown that features of interaction (e.g., perceived mutuality and involvement) are correlated with credibility and influence, suggesting that the relation is even more complex than initially thought. In general, partners that engender a sense of mutuality or involvement are perceived as more credible and are, in turn, more influential.

It is possible that the task rather than the nature of the communication interface produced limited variability in understanding. If participants feared they did not understand something important or were in awe of the partner's ability to offer complicated and obscure technical information, they may have been overly prone to defer to their partner's judgment. Other tasks might introduce more variability in responses and hence, a better opportunity to detect the impact of the communication format on understanding.

7.3. Implications for HCI interface designers

How does one design a "better" interface? Interface designers ought to be concerned with facilitating clear and accurate information exchanges, efficient transactions, and high-quality collaborative work. Therefore, the solution for designers is to understand how features of computer interfaces relate to features of the interaction between human users or between a human user and computer agent and then match the interface design to the desired outcomes. Our results suggest that choice of interface design is influenced by the outcomes one wishes to emphasize for human participants in computer interactions. Three global outcomes stand out as most relevant: passive involvement, collaboration, and relationship building. Optimal matching of these outcomes to computer interfaces leads to better task outcomes and partner assessments of credibility.

Interface designers have a variety of options to incorporate into new interfaces, such as animated agents, yet these options are not always matched to the desired task and social outcomes. In the case of passive involvement, the participant is cast as a receiver of information but takes little or no role in creating it. In addition, the participant is not expected to be influential in the sense that his or her contributions will affect the nature of the encounter. This situation might be desirable when the goal is to have the partner merely receive information (e.g., from a web search engine) or when interaction needs to take place over a fairly short period of time and the likelihood of users being misled is low. For those interested in fostering passive involvement, interfaces need only have text-based interfaces or text-based with audio, since they tend not to foster complete involvement. (This statement applies only to HCI, since text-based computer-mediated communication does not produce the same effects.) A common example would be a university library database for students to search for books or journal articles. Incorporating anthropomorphic options into an interface for the sake of including them will not create a "better" interface. Quite the opposite, our findings suggest that mismatching of the interface with the task reduces utility and task outcomes.

Collaborators, on the other hand, are more or less equal partners who are expected to participate actively in generating information and knowledge, and whose contributions are consequential for how the encounter proceeds. Interfaces fostering an increased sense of involvement and mutuality would be an optimal design for human–computer collaboration, such as in assisting in the decision making process. Anthropomorphic interfaces incorporating animated characters, speech synthesis or artificial intelligence are more appropriate when collaborative encounters are desired in as much as participants take a more active role in the interaction. Such interfaces also are rated high in utility and promote the most attraction for the task outcome.

Finally, relational building concerns creating and maintaining useful and positive assessments of interactional partners. Anthropomorphic interfaces would seem appropriate for relational building because respondents tend to rate their partners and interactions higher when more human features are provided. To date, this type of computer interface is still an unachieved goal except in the futuristic prophesies of MIT computer scientist Dertouzos (1997) and the virtual butler in Apple Computer's famous Knowledge Navigator video. Undoubtedly, advancements in computer technology will increase the capabilities of computers to use more anthropomorphic features. The challenge for interface designers is to resist the temptation of burdening the user with an overloaded interface not matched to the desired task or outcome.

7.4. Future work

This investigation represents a first effort that demonstrates the ways in which interfaces can be systematically manipulated to uncover what properties produce what effects. More definitive conclusions must await larger sample sizes, which might generate more significant results. Additionally, to further parse the effects of modalities and anthropomorphism in HCI, two additional features should be included in future experimental designs: speech input and "intelligent" responses. These features correspond directly to prototypical human attributes, namely comprehension of spoken words and responses based on multiple prior inputs. The reasons for not including speech recognition in this study were purely practical. The fact that no "intelligence" was employed for the dialogue script, however, was due to the goal of creating conditions that would be directly applicable to current software and interface design of, for example, search agents and information filtering devices. Future variants of this study could include manipulating the information presented, such as the partner giving faulty information or demonstrating expertise through specialized knowledge, altering persuasive strategies used by the partner, altering the clarity of presentation, or alternating partner-versus-subject initiation of the conversation. Also, synthetic faces of increasing quality could be evaluated, ranging from low resolution texture mapping, through very high detailed animations, to pre-recorded videos of real humans. Subsequent studies should, based on the experiences gained, aim at identifying which dimensions are crucial to perceived realism. Ultimately, studies should also be designed to address the question to what degree realism is important, and which other dimensions are equally or even more important.

8. Conclusion

Overall, it would appear that FtF interaction is best for generating positive social judgments and interpersonal relationships, but that HCIs are more influential in decision-making tasks. Whether these differential benefits generalize to other kinds of tasks and interactions remains to be seen. Perhaps this is but a transient phenomenon — much like the panic that ensued at the first public presentations of cinematography at the turn of the century, where people ran from the screen on which a speeding train was approaching. If so, what we call virtual communication could just be the temporary byproduct of the transition from older types of information technology to the new computerized, networked, mobile information technology. Of interest will be whether such effects remain after a period of initial adjustment to such innovations.

Acknowledgments

A previous version of this paper was presented at the annual meeting of the Hawaiian International Conference on Systems Sciences, 1999. Portions of this research were supported by funding from the US Army Research Institute (Contract #DASW01-98-K-009). The views, opinions, and/or findings in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision.

References

Alperstein, N. M. (1991). Imaginary social relationships with celebrities in television commercials. *Journal of Broadcasting and Media, 35*, 43–58.

Beskow, J. (1995). *Rule-based audiovisual speech*. Unpublished doctoral dissertation, Royal Institute of Technology, Stockholm, Sweden.

Bonito, J. A., Burgoon, J. K. & Bengtsson, B. (1999). *The role of expectations in human–computer interaction*. Paper presented at GROUP '99: International Conference on Supporting Group Work, Phoenix, AZ.

Brazier, F. M. T., & Treur, J. (1999). Compositional modeling of reflective agents. *International Journal of Human-Computer Studies, 50*, 407–431.

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Ramirez, A., & Dunbar, N. (2000). Testing the interactivity model: communication processes, partner assessments, and the quality of collaborative work. *Journal of Management and Information Systems, 16*, 33–56.

Burgoon, J. K., Johnson, M. L., & Koch, P. T. (1998). The nature and measurement of interpersonal dominance. *Communication Monographs, 65*, 308–335.

Burgoon, J. K., Pfau, M., & Birk, T. (1990). Nonverbal behaviors, persuasion, and credibility. *Human Communication Research, 17*, 140–169.

Carlson, R., Granstrom, B., & Hunnicut, S. (1982). A multi-language text-to-speech module. *Proceedings of ICASSP, 3*, 1604–1607.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Erlbaum.

Dehn, D. M., & van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies, 52*, 1–22.

Dertouzos, M. L. (1997). *What will be: How the new world of information will change our lives*. San Francisco, CA: HarperEdge.

Ekman, P. (1982). *Emotion in the human face (2nd ed.)*. Cambridge, UK: Cambridge University Press.

Goffman, E. (1967). *Interaction ritual*. New York: Pantheon Books.

Harmon, S. W. (1996). Novice use of a dimensional scale for the evaluation of the hypermedia user interface: caveat emptor. *Computers in Human Behavior, 11*, 429–437.

Jiang, J. J., Klein, G., & Vedder, R. G. (2000). Persuasive expert systems: the influence of confidence and discrepancy. *Computers in Human Behavior, 16*, 99–109.

Keppel, G. (1991). *Design and analysis: A researcher's handbook (3rd ed.)*. Englewood Cliffs, NJ: Prentice-Hall.

Lipsey, M. W. (1990). *Design sensitivity*. Newbury Park, CA: Sage.

Lundeberg, M. (1997). *Multi-modal speech communication development of a test-environment*. Unpublished doctoral dissertation, Royal Institute of Technology, Stockholm, Sweden.

Marakas, G. M., Johnson, R. D., & Palmer, J. W. (2000). A theoretical model of differential social contributions toward computing technology: when the metaphor becomes the model. *International Journal of Human-Computer Studies, 52*, 719–750.

McCroskey, J. C., Hamilton, P. R., & Weiner, A. M. (1974). The effect of interaction behavior on source credibility, homophily, and interpersonal attraction. *Human Communication Research, 1*, 42–52.

McCroskey, J. C., & McCain, T. A. (1974). The measurement of interpersonal attraction. *Speech Monographs, 41*, 261–266.

McCroskey, J. C., & Young, T. J. (1981). Ethos and credibility: the construct and its measurement after three decades. *Speech Monographs, 41*, 261–266.

McGuire, W. J. (1985). Attitudes and attitude change. In G. Lindzey & E. Aronson, *The handbook of social psychology* (3rd ed., pp. 233–346). New York: Random House.

McKevitt, P., Partridge, D., & Wilks, Y. (1999). Why machines should analyze intention in natural language dialogue. *International Journal of Human-Computer Studies, 51*, 947–989.

Moon, Y., & Nass, C. (1996). How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research, 23*, 651–674.

Morkes, J., Kernal, H. K., & Nass, C. (1999). Effects of humor in task-oriented human-computer interactions and computer-mediated communication: a direct test of SRCT theory. *Human-Computer Interaction, 14*, 395–435.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45*, 669–678.

Nass, C., Steuer, J., Henriksen, L., & Dryer, C. (1994a). Machines, social, and ethopoeia: performance assessments of computers subsequent to self- or other-evaluations. *International Journal of Human-Computer Studies, 40*, 543–559.

Nass, C., Steuer, J. & Tauber, E. R. (1994b). *Computers are social actors*. Paper presented at the annual meeting of the Conference on Human Factors in Computing Systems, Boston, MA.

Nongjian, Z. (1999). A core of ego and a new system of artificial intelligence. *Computers in Human Behavior, 15*, 625–652.

Perrolle, J. (1987). *Computers and social exchange: Information, property, and power*. Belmont, CA: Wadsworth.

Rubin, B., & McHugh, M. P. (1987). Development of parasocial interaction relationships. *Journal of Broadcasting and Electronic Media, 31*, 279–292.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn-taking for conversation. In J. Schenkein, *Studies in the organization of turn-taking for conversation* (pp. 7–55). New York: Academic Press.

Sanders, R. E. (1987). *Cognitive foundations of calculated speech*. Albany: SUNY Press.

Sanders, R. E. (1997). The production of symbolic objects as components of larger wholes. In J. O. Greene, *Message production: Advances in communication theory* (pp. 245–278). Mahwah, NJ: Erlbaum.

Shapiro, M., & McDonald, C. (1992). I'm not a real doctor, but I play one in virtual reality: implications of virtual reality for judgments about reality. *Journal of Communication, 42*, 92–114.

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When interface is a face. *Human-Computer Interaction, 11*, 97–124.

St. Amant, R. (1999). User interface affordances in a planning representation. *Human-Computer Interaction, 14*, 317–354.

Turner, J. R. (1993). Interpersonal and psychological predictors of parasocial interaction with different television performers. *Communication Quarterly, 41*, 443–453.

Walther, J. B. (1996). Computer-mediated communication: impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*, 3–43.

Warner, R. M., & Sugarman, D. B. (1986). Attributions of personality based on appearance, speech, and handwriting. *Journal of Personality and Social Psychology, 50*, 792–799.

Wheeless, L. R., & Grotz, J. (1977). The measurement of trust and its relationship to self-disclosure. *Human Communication Research, 3*, 250–257.