Abstract
This paper describes a pilot study on decision influence, understanding, and subjective measures of credibility in human interaction with a computerized partner for a fictional decision-making task. Subjects (N=70) were randomly assigned to one of 5 different computer partners or to a human partner. Subjects completed the Desert Survival Task, and engaged in a dialogue with their partner. Pre- and post-interaction rankings were used to measure decision quality and influence. Results revealed that face-to-face interaction generated the most positive social judgments, and more positive social judgments were associated with greater understanding, but the computer conditions were more influential, especially when the computer was the least anthropomorphic.

1. Introduction

Modern information technology allows more and more people to communicate and exchange information in ever new ways. The most prominent feature of the so-called "information society" is an increasingly intense and mobile communication, based on a multitude of media and various forms of expression—speech, text, pictures, hypermedia, multimedia, and so on. These new technologies have the potential to affect collaborative work in important ways [1].

One exciting but currently little understood innovation is the augmentation of human sources of information with intelligent computer agents. Such agents may "converse" with group members and even function as a team member (see [2]), creating what has been called virtual communication [3], [4], [5]—communication in which one or more parties is non-real, fictive, or simulated, or deviates so much from the perceived interlocutor that the interchange becomes untrue to reality in one or more respects. Team members may encounter information in shapes that, in reality, have no well defined, simple or clearly delimited transmitter or source. For instance, messages may be generated by computers or from compilations of information from different databases, i.e., they may represent collective cut-and-paste creations that no longer originate from a single and identifiable human source. At the same time, virtual agents may be fitted with distinctly human-like qualities through the addition of such anthropomorphic features as synthetic faces and voices and apparent reasoning ability.

The introduction of these forms of communication and information exchange into collaborative activities may have profound effects on peoples’ attitudes toward each other, toward the information they exchange, and the work they produce. While the potential benefits of such computer tools may seem self-evident, the possible risks may be less so. One particular concern surrounds issues of authenticity, trust, and credibility. Users must be able to discriminate between that which is valid and valuable and that which is fallacious, feigned, or fabricated. An important consideration in designing computer interfaces is the impact that the interface has on users’ abilities to negotiate these claims successfully in the course of interaction [20]. An important consideration in managing information systems is selecting the appropriate means of communication and information exchange for the task at hand. Uncritical reliance on computer-generated information or selective attention to certain information sources, for example, might bias information processing and adversely affect the quality of decision-making.

Such an issue lends itself to an interdisciplinary examination, which is the purpose of the current investigation. Although much relevant research has been conducted,[1] some of which is reviewed below, few experiments have systematically and comprehensively compared multiple HCI (human-computer interaction) interfaces to FfF (face-to-face) interaction to determine which features affect the process and products of collaborative work. This report presents results from a first study in a planned program of research testing how features of the interface—specifically, its degree of anthropomorphism and inter-activity—affect users’ abilities to utilize and judge the messages and information they receive. Our focus here is on how different HCI

---

1 For a recent survey of such research, see [6]. For a more detailed definition of virtual communication, see [3].
interfaces compare to each other and to FTF interaction on resultant credibility, understanding, and influence during a decision-making task. Specifically, we examine how much credibility people assign to a human or computer partner, how accurately people process and recall the information they receive, and how much influence the partner is able to exert on the final decision. A companion paper presents user assessments of the communication format, the interaction process, and the partner, and how those judgments relate to these same outcomes of influence, understanding, and credibility (see [7]).

We begin by clarifying the interrelationships among credibility, understanding, and influence. We then turn to the properties of human communication that might affect the “realism” of HCI interfaces and review relevant research and theorizing that informed our predictions as to how these properties might affect influence, understanding, and credibility in a collaborative task. Results of the first experiment are then presented, and implications for computer-aided collaboration are discussed.

2. The Nature of Influence, Understanding, and Credibility

Psychologists and communication scholars have long held that credibility, understanding, and influence are closely interrelated. Credibility refers to the recipient’s judgment that a message and/or its source are believable and convincing. Assigning credibility to a message or information source is usually a prerequisite to message or information acceptance. Recipients who believe a piece of information is of dubious value or its source is untrustworthy, for example, are unlikely to incorporate such information and arguments into their own decision-making. Credibility is not just a single judgment. 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 [8], 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. It may gain its believability because the source is seen as truthful, trustworthy, and reliable, in which case the character dimension is implicated. It may also gain credibility 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).

Influence may also depend on understanding, i.e., the degree to which the contents of a message are comprehended and recalled accurately. Poorly understood or remembered information, for example, will not be utilized properly in making a decision.

Influence itself can also be judged in several ways. Here we consider the extent to which participants shift their decisions in the direction of a given team member’s recommendations and whether final decisions reflect that member’s input. Because we utilize a task previously evaluated by experts, we can judge the actual quality of the decisions our participants reach by determining how closely they correspond to those recommended by experts.

3. Properties of Real and Virtual Interpersonal Interaction

Three different lines of research converge to offer insights into what constitute the “defining” properties of interpersonal interaction, properties that that might be required to make interaction between people and machines convincingly “real.” The first approaches the issue from the perspective of interpersonal communication and its inherent qualities that many computer technologies seek to emulate. The second approaches the issue from the perspective of mediated communication between humans, whether in the form of broadcast media such as television or narrowcast media such as computer-mediated interaction. The third approaches the issue from the perspective of human-computer interaction and viewing computers as social actors. We consider each of these bodies of literature in turn.

3.1. The Principle of Inter-Activity

The first framework that informs this issue originated from Buller and Burgoon’s [9] interpersonal deception theory (IDT), a theory designed to apply principles of interpersonal communication to the phenomenon of deception. Because that theory’s development was motivated by the need to distinguish interactive from noninteractive deception, its originators enumerated intrinsic properties that they believed differentiated highly interactive contexts from less interactive or noninteractive ones. These properties collectively define the principle of inter-activity (shorthand for interpersonal interactivity), a term intentionally hyphenated to underscore the active nature of interpersonal communication.

One way to conceptualize these properties is according to the structural affordances that inter-activity offers communicators. These are features that are “built into” or intrinsic to a given communication modality or format. As compared to formats that are lacking or low in
inter-activity, highly inter-active formats are posited to be:

1. **unmediated** (i.e., interactants are physically co-present and messages are exchanged orally rather than transmitted via some electronic or mechanical medium).
2. **propinquitous** (i.e., they are same-place rather than geographically dispersed).
3. **participative** (i.e., 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), and evolutionary).
4. **modality- and information-rich** (i.e., participants have full access to a wide array of environmental, visual, audio, and verbal context cues).
5. **synchronous** (i.e., occurs in real time rather than being delayed or asynchronous).
6. **contingent** (i.e., each turn at talk is dependent on prior turns at talk, so that the flow of discourse is somewhat unpredictable and meanings are managed “locally”; see also [10], [11]).
7. **identified** (rather than anonymous).

If communication modalities are arrayed along an inter-activity continuum, then traditional FtF dialogue is the current prototype for the high inter-activity pole because it is unmediated, same-place, same-time, participative, informationally rich and complex, and identified. Moving down the continuum would be various mediated and distributed formats, with asynchronous electronic correspondence among unacquainted members of distributed work teams located near the noninteractive pole. (It should be noted that although FtF interaction is thought of as the prototype of high inter-activity, technologies of the future, such as virtual reality, may eventually exceed the capacity of FtF interaction to maximize certain inter-active properties; see [12]).

Inter-activity may also vary qualitatively within a given communication format. Televised broadcasts, for instance, are usually noninteractive because there is little dependency between a subsequent message and its prior, but the call-in talk show is an exception. Another example might be computer technologies that create time lags between messages and thus inhibit the kind of smooth, synchronized interaction or message modification necessary for the creation of common meanings [13], [14].

An alternative way to conceptualize and operationalize inter-activity is according to how it **looks and feels**, i.e., according to the indicators by which we “know” that a situation is inter-active. These are the qualities of the interaction process itself. From this perspective, a communication environment or format is viewed as more inter-active 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 “together,” “connection, and receptivity to one another).
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 [15]. Identity creates 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, immediacy, mutuality, and individuation.

Two questions present themselves. One is whether computers might match or exceed humans in their ability to create understanding, earn favorable credibility judgments, and influence decision-making. The second is, if computers are able to match or exceed humans on these measures, are differences due to features of inter-activity per se or the specific anthropomorphic qualities that are responsible. A simplistic and design-driven response might be that increased inter-activity should enable better understanding as well as the social identification and mutuality that lead to viewing the partner as credible which in turn should enable the partner (human or computer) to exert more influence. If this reasoning is correct, then computer agents may suffer relative to FtF interaction because, despite technological advances that are constantly expanding the frontiers of what is feasible, at present computers still interact awkwardly. They are unable to supply the kind of contingent and fully synchronous interaction that is present in FtF conversation. Moreover, the sheer interjection of an electronic medium may “distance” interactants relative to FtF interaction. And computer agents, even in multimedia form, do not supply the richness of social and sensory information that is available FtF.

On the other hand, some of the shortcomings of computers might be offset by the addition of anthropomorphic features which confer the appearance if not the reality of interacting with another human. Computers can already be made to appear virtually human—they can be given a face and a voice—which creates a sense of identity. This in turn can foster perceptions of interpersonal connection. In this sense, computers can be made more or less anthropomorphic in appearance and speech far more easily than they can be made to appear human in terms of other interactional
capabilities. What remains to be seen is whether qualitative additions to interfaces in the form of greater anthropomorphism can create sufficient realism to cause users to overlook imperfections in other interactional properties such as contingency, perhaps to the point of virtual agents equaling or exceeding human partners in their credibility and influence on users. If we focus specifically on search engines, web-agents, filtering devices, etc.—applications that are rapidly becoming ubiquitous—an affirmative answer may seem more likely, because users probably do not anticipate sophisticated communication abilities for such uses, when it comes to managing a (prolonged) conversation—simply because that is not the nature of the encounter. Like many situations in real life, these situations do not necessarily involve a lot of contingencies. This would imply that when comparing human and computerized interactional partners in relatively deterministic situations, humans do not automatically get the upper hand, as far as contingency management is concerned, since there is no room to display such abilities.

IDT also alerts us to the role that ascriptions of truth play in predicting how team members will respond to information introduced by a computer or human partner. As a cooperative venture, a key feature of interperson conversation is that participants must negotiate with one another to establish that what is being said is meaningful and true, and that the speakers are sincere [16]. Most often, the presumption is that interlocutors are truthful, which leads people to be fairly uncritical of the messages and information emanating from others, i.e., they exhibit positivity and truth biases when judging such messages and information.

It is possible that these tendencies to judge others leniently and as truthful may extend to computer-generated messages as well as to human-generated ones, especially if the computer has human-like qualities. Another principle from IDT would support this supposition. According to IDT, communicators seek to manage the informational content, accompanying nonverbal behaviors, and general image conveyed by their messages [17]. In FtF interaction, humans vary substantially in their successful achievement of each of these. Comparatively, a computerized agent, endowed with human-like features, should be totally unconcerned with behavior and image management, if not with information management. Facial animation and speech synthesis techniques could possibly render a highly “managed” agent, in terms of behavior and image. Faced with such an agent, a human conversational partner might well find it hard, or even ridiculous, to question the validity of information presented by the agent during the course of interaction, even though the receiver is aware that the sender is non-human. Thus, one might expect receivers to be more influenced by information when the sender is a computerized agent, interacting by means of an animated face and synthetic voice, than in face-to-face conversations.

On the other hand, it is plausible that humans have more behavioral resources at their disposal to achieve an appealing and credible demeanor and that they are better able to adapt their conversations if there are indications that their image is suffering. This would advantage humans over computers. This leads to an alternative prediction, namely that receivers will be less influenced by information exchanged during a discourse when the sender is a computerized agent than in FtF conversations.

A third possibility is that humans may gain on some outcomes while computers gain on others. The body of literature reviewed next partially speaks to this issue.

### 3.2. Mediated Communication

A second body of research that has some bearing on these possibilities comes from the domain of mass media. One line of research on parasocial relationships has found 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 [18], [19]. 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 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 blend what they actually see with prior media exposure and gossip to form their construction of reality [21]. From this we can surmise that computer agents may be the beneficiaries of similar tendencies.

Other research has documented the suggestive powers of today’s media and tomorrow’s VR techniques in projecting credibility and exerting influence [21]. 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. 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, appropriateness, and so forth. Therefore, its very availability in mediated form is taken as prima facie evidence of its authenticity. From these findings, we
might expect mediated communication forms in general to be the beneficiary of a positivity bias whereby that which is delivered through a mechanical or electronic medium is regarded de facto as credible.

Yet other media research implies a gradient of increasing credibility as a medium becomes “richer” in sensory channels involved and amount of social information it supplies, i.e., as in increases along one of the properties of inter-activity. People trust pictures more than the printed word and they are more likely to trust a television image than a newspaper article [22]. Thus we might predict that the addition of computer graphics will increase trust in computer-generated and computer-presented information.

3.3. Computers as Social Actors in Human-Computer Interaction

One reason for these attributions of credibility is that people respond to media, computer artifacts, and other communication technologies in fundamentally social ways. This has been demonstrated in a very clever series of studies by Nass and colleagues (e.g., [2], [23], [24]) in which basic social science findings related to humans are replicated but with computers substituted for humans in the experimental procedures. Among the findings generated by this program of research are that users apply politeness norms, notions of ‘self’ and ‘other’, and gender stereotypes when interacting with computers. Also, subjects who are told they are interdependent with a computer form a connection with the computer as a team. These responses are not the result of the mistaken belief that computers are human-like, or that they act as proxies for human programmers. Furthermore, a limited set of characteristics normally associated with humans provides sufficient cues to encourage users to exhibit such behavior. For example, people are more likely to exhibit impression management concerns with a talking face than with computer text [1]. In an interview conducted by a computer “counselor,” participants revealed less to the talking face than to the text and evaluated it less well. These evaluations were made on personality attributes that research has shown are affected by people’s physical appearance and voice [25], [26].

Recent research begins to illuminate what properties of human-computer interfaces are particularly significant and why. Kiesler, Sproull and Waters [27] conducted an experiment that contrasted subjects’ discussion and choices while playing a prisoner’s dilemma game during which the partner was either another person or a computer. The participants in this study were able to communicate with their partner in both the human partner and computer partner condition. The researchers varied the interface to the computer partner to exhibit increasingly human-like characteristics. In one condition, the computer communicated solely by means of text. In a second condition, the computer spoke using synthetic speech. In a third, the computer screen displayed a human face that spoke and moved its facial skin accordingly.

The hypothesis under test was that people will behave more cooperatively toward a human partner than toward a computer partner, but the more human-like features the computer has, the more people’s cooperativeness with a computer will be like that with a person. However, Kiesler and colleagues also entertained the alternative that imperfections in current technology might lead people to discredit a computer that has human-like features that fall short of their prototypical expectations or they might fail to identify with the human-like but imperfect computer altogether and respond instead with strategic selfishness.

Results revealed that the manipulation of the human-like attributes of the talking face computer partner were successful in causing participants to view the computer as making real choices, and participants paid closer attention to the computer partner than to a human partner. However, adding only synthetic speech to the computer interface did not significantly change perceptions of the human qualities of the computer partner. Furthermore, there was a (non-significant) trend against cooperation with more human-like computers, though the results included some anomalous contradictions. The researchers speculated that negative reactions to the talking face computer might have resulted from idiosyncrasies in the particular computer interface used and that a more perfect representation of a face and voice might increase the credibility of, and cooperation with, the computer partner. Such speculation is bolstered by recent studies [28] that have shown that visual fidelity in computer representations increases involvement and learning.

4. Hypotheses and Research Questions

We have proposed that in certain aspects, we can conceive of interaction between humans and technology as social interaction [29]. Social interaction with technology seems to arise from the general psychological tendency of people to respond socially in situations in which they are reminded of their own humanity or social selves, or in which they form an attachment to another. We hypothesize that social behavior, then, can be triggered by any entity that is sufficiently human-like or that responds in a human manner. Nass and colleagues have argued that in situations where a technology possesses characteristics that are similar to those of humans (e.g., words for input, responses based on multiple prior inputs, and human sounding voices) people
will exhibit social responses to the technology. People will, for example, use the speech cues emitted by a talking computer to imagine a human prototype for the computer, and will follow social decision rules in interacting with the computer.

Framing human-computer interaction in terms of social identity and the properties associated with interactivity implies that people will feel most connected and involved with entities that are most like themselves. This argues for the greatest credibility and understanding being evident in F2F interaction, where the partner is a human, followed by those conditions in which a computer agent is the most anthropomorphic. This emphasis on the social aspects of interaction would seem to elevate the credibility dimensions of sociability and character (which includes trust) to the fore. Additionally, humans have greater ability to be nonverbally expressive and energetic, which may gain them benefits in terms of dynamism. These gains in credibility and understanding should translate into greater influence in decision-making.

However, a countervailing perspective comes from the finding that people routinely assign high credibility to media and computers. If this positivity bias reflects an unconscious belief that whatever messages and information emanating from the media and computers have already been screened, authenticated, and validated and/or originate from highly knowledgeable, skilled professionals, then people may be more inclined to believe computer agents than humans, at least in terms of judgments of competence. We thus anticipated that humans would be rated higher on sociability, character, and dynamism but that computers might be rated higher on competence.

Given competing credibility benefits to humans and computers, it is unclear which dimension(s) of credibility are most likely to prevail in producing influence and understanding. In light of this, and under the assumption that interaction with computers is still in some degree artificial, and that such artificiality is consequential for how interaction with human and computer partners is perceived and assessed, we hypothesized that face-to-face interaction differs from human-computer interaction on influence, credibility, and comprehension, but left open to follow-up testing the direction of the difference.

As for the comparisons among forms of human-computer interaction, all three bodies of literature reviewed suggest that people will be more likely to identify with a computer that is richer in social information and incorporates more human-like qualities. (Note that this, in itself, does not imply that people treat computers as humans. However, people do respond socially to computers employing/manifesting/emulating social behavior.) From this it follows that assessments of human-computer interaction should depend on the degree to which computers are perceived as social actors, and such perceptions should be affected by how inter-active and anthropomorphic the machine is made to appear and sound. However, it is important to keep in mind that it is not just appearance and sound that induces social responses and judgments. For instance, output based on multiple prior inputs; the ability to accept spoken words as input; and openness to others’ suggestions; are factors that determine how we perceive agents (and people). Also, given that these criteria have been met, the question of what to say, and how to say it comes into play, for agents as well as for people. Thus, we hypothesized that increases in inter-activity and anthropomorphism of the computer lead to corresponding increases in credibility and influence. We chose to further explore whether it is inter-activity per se of anthropomorphism that makes a difference by conducted further focused contrast tests among conditions, as described in the results section.

Finally, the literature hints at increases in inter-activity and anthropomorphism creating greater understanding through the mechanisms of involvement, mutuality, and credibility, but this research has not expressly tested this in the context of HCI. Thus we posed as a final research question whether increases in inter-activity and anthropomorphism of the computer produce corresponding increases in the accuracy of recalled information delivered by the computer. Again, focused contrast tests further explored differences among these features.

5. Method

5.1. Subjects

Subjects (N = 70) were male undergraduate students in the social science faculty at Umeå University who were solicited via posters on campus to engage in a study of alternative problem-solving methods. They were paid approximately $10 (100 Swedish crowns) for their participation.

5.2. Procedure

Upon arrival at the research site, participants completed consent forms and were thanked for agreeing to be recorded, after which they heard the following introduction to the study:

You and your partner will be discussing the Desert Survival Problem after which you will answer some questions about the discussion. In today’s world, much decision-making no longer takes place face-to-face. Sometimes people interact with
others by computer, telephone, or videophone; and sometimes they interact with a computer. In this study, some people will be conducting the decision task electronically and some will be conducting it face-to-face.

Subjects 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; scripted face-to-face interaction; and unscripted face-to-face interaction), so that for each condition there were 10 subjects. In the face-to-face conditions, the partner was a male confederate2. In the computer conditions, the partner was a computer named Holger (described in more detail below; see Figure 1).

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’ x 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 entitled “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 they would be evaluated on the quality of their final rankings and those of their partner, so 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.”

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 HP workstation placed on a table in the middle of the room. In the control 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 face-to-face 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 [23], [24]. 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.

Figure 1. The animated face of the computer partner.

5.3. Experimental Conditions and Apparatus

The experimental conditions were as follows: Text-only. In the text-only condition, subjects were seated in front of a computer presenting its ranking and reasons according to a fixed interaction script. The presentation was given by means of text only in a window on the screen (see Fig. 2). After each communication presented by the computer, subjects typed in their response in another window and pressed a continue button. Text and voice. In this condition, the text was accompanied by a synthesized voice emanating from two loudspeakers on either side of the computer, vocalizing the text on the screen. The speech synthesis software used in this study was a development of the KTH text-to-speech synthesis, described in [30]. Text, voice, and image. This condition was the same as the previous, except that a facial (still) image (see Fig. 1)

---

2 This was due to the fact that the facial images and animations in the computer conditions were those of a male character.
was also shown on the computer screen. The size of the image was adjusted to correspond to the size of a face of a person sitting on the opposite side of the table.

*Text, voice, and lip-synched animation.* In this condition, the facial image was replaced by an animated face, moving its lips and facial skin synchronized with the speech, occasionally flashing its eyelashes. 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 [31] and [32].

*Voice and lip-synched animation (no text).* This condition was the same as the previous, except that there was no text output. The subject still typed in his answers in a text window, however. This condition was included based on the assumption that some subjects might focus on the text-output window and miss the facial animations while reading.

*Scripted face-to-face interaction.* In this control condition, a male confederate was seated on the opposite side of the table, interacting with the subjects according to the same script used for the computer conditions. The confederate was instructed to adhere very closely to this script. This sometimes meant disregarding questions and comments by the subjects, causing the dialogue to become rather unnatural. With regards to confederate behavior, the aim here was to minimize idiosyncratic gestures, but at the same time not to have confederates behave rigidly or have their speech devoid of the normal back channel cues like nods, smiling, etc. Above all, we wanted to maintain some continuity across face-to-face conditions on these features.

*Unscripted face-to-face interaction.* This condition was the same as the previous, except that the confederate was not instructed to follow the script. Instead, he was allowed to use his own words and adapt to subjects’ questions, while still trying to convey all the information contained in the script but without introducing any new information.

### 5.4. Dependent Measures

Following the interaction, subjects completed the following dependent measures, among others.\(^3\)

*Influence:* This was measured in three ways; as decision quality, absolute influence, and relative influence. Subjects’ rankings of the salvaged items prior to interaction (“pre-rankings”) and following the interaction

---

\(^3\) Only those measures pertaining to this paper are reported here. Other measures—such as measures of homophily; attraction; expectedness and valence of communication; perceived mutuality, closeness, and social presence; interaction ease; behavioral mutuality; and sense-making—will be reported elsewhere.

---

**Figure 2. Sketch of the computer screen in the computer partner conditions. Certain screen elements were only visible during some conditions.**
Interim reliabilities, calculated with Cronbach’s alpha, were .81 for character, .68 for truthfulness, .75 for competence, .59 for dominance, .68 for dynamism, and .86 for sociability.

Understanding: 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 correct or incorrect matches, and also as absolute distance from actual partner ranking. In the latter case, a small distance indicated high understanding.

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 interjudge reliability test (r=.92). The individual ratings were averaged to produce a single understanding score.

6. Results

Due to the exploratory nature of this investigation, which led us to use two-tailed t-tests, and to the somewhat small sample size per cell necessitated by so many different versions, we set alpha at .10 to offset the lower statistical power. (However, it should be noted that most findings were significant at or below the .05 level). The hypotheses were tested with planned contrasts (see [37] and Table 1). Contrast 1 compares face-to-face (FtF) to the human-computer interaction (HCI) conditions. Contrasts 2 through 4 are a series of Helmert contrasts that collectively examine whether ordinal increases in the inter-activity property of modality richness make a difference. To do this, we arrayed the media conditions from least to most amount of social presence and social information to form the richness variable as follows (in ascending order):

1. Text-only
2. Text and voice
3. Text, voice, and image
4. Voice and animation
5. Text, voice, and animation
6. Unscripted face-to-face

The first three conditions mirror hypothesized increases in available modalities. We had some uncertainty about the rank-ordering of conditions 4 and 5. Although the latter condition clearly contains more information, as mentioned above, the voice and animation condition (no text) was included since we suspected that
subjects might focus on the text output instead of the facial animations. It was therefore possible that animations, in isolation, would actually be higher in social presence than animations going partly unnoticed due to a more obtrusive text output. We therefore combined these two conditions for the primary hypothesis test. Additionally, the observations from experimental assistants that the instructions in the scripted face-to-face condition were producing highly awkward, unnatural interactions led us to omit this condition for purposes of contrasting FtF with HCI. (See the companion paper by Burgoon et al. [7] for analyses pertaining to this noncontingent FtF condition.)

To further assess the relative impacts of anthropomorphism and inter-activity, we conducted an additional contrast. Contrast 5 compares the voice and animation condition to the text, voice, and animation condition. Arguably, the latter is less anthropomorphic than the former but also richer in available modalities. If the former condition produces better outcomes than the latter, it suggests that qualitative differences in anthropomorphism are at work rather than inter-activity per se. All codes were orthonormalized in the data analysis, which was conducted using SPSS version 8.0 for Windows.

Table 1. Contrast coefficients

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Richness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Text-only</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

6.1. Comparison of FtF to HCI

Influence and Decision Quality. Figure 3 displays the pre- and post-rankings, Table II displays the absolute and relative influence scores and decision quality means by condition. The planned contrasts, conducted with separate variance estimates due to heterogeneity of variance, were significant for both absolute and relative influence, absolute \( t(19.68) = -1.80, p = .087 \), two-tailed; relative \( t(20.77) = -2.10, p = .048 \), two-tailed. Influence was significantly higher in the computer conditions than in the unscripted face-to-face condition. A strong correlation was found between absolute and relative influence (Pearson’s \( r = .97 \)). The planned contrasts revealed no significant differences in decision quality.

Understanding. The planned contrasts on the accuracy of recall and understanding measures failed to produce significant results and were uncorrelated with the influence measures (all \( r’s \leq .13 \)). However, in addition to correlating with each other (\( r = .27, p < .05 \)), the deviation from accurate recall and understanding measures were correlated with the credibility measures. Those subjects who saw their partner as more credible in terms of competence, character, truthfulness, and dynamism also were more accurate (i.e., closer to the...
confederate’s answers) and better able to report the content of the partner’s arguments (see Table III). This suggests that perceptions of credibility affected subjects’ attention and information recall.

Table 3. Correlations between credibility dimensions and measures of understanding

<table>
<thead>
<tr>
<th></th>
<th>Deviation from accurate answers</th>
<th>Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Character</td>
<td>-.231*</td>
<td>.210</td>
</tr>
<tr>
<td>Truthfulness</td>
<td>-.373**</td>
<td>.215*</td>
</tr>
<tr>
<td>Competence</td>
<td>-.339**</td>
<td>.446**</td>
</tr>
<tr>
<td>Dynamism</td>
<td>-.298*</td>
<td>-.019</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (1-tailed)
** Correlation is significant at the 0.01 level (1-tailed)

Credibility: Analyses were conducted on six measures: competence, character, truthfulness, dynamism, dominance, and sociability. The means appear in Table IV. The planned contrasts produced significant results for three measures: truthfulness, \( t(25.30) = 2.52, p = .019 \), two-tailed; dynamism, \( t(54) = 2.09, p = .042 \), two-tailed; and sociability, \( t(54) = 2.75, p = .008 \) two-tailed. In contrast to the results for influence, partners were seen as much more credible in the FtF than the HCI conditions. Thus, the conditions had opposite effects on credibility and influence, which were uncorrelated with one another.

6.2. Comparisons among HCI Conditions

The various planned contrasts among the HCI conditions produced no significant effects on influence or understanding, although the pattern of means in Table II reveals that influence was highest in the text and voice condition. With a larger sample size, this condition might emerge as more influential than other combinations. The credibility analyses yielded a significant effect on dominance such that the text condition differed from the text and voice condition, \( t(16.68) = -1.76, p = .098 \). Unexpectedly, the text-only condition was seen as the more dominant. There was a similar and near-significant pattern for dynamism (\( p = .15 \)). Additionally, nonsignificant but suggestive patterns on dynamism (\( p = .11 \)) and sociability (\( p = .15 \)) revealed that the addition of the still image might reduce credibility relative to text alone or text and voice. These results fail to support the hypothesized gain due to increased anthropomorphism of the computer.

Table 4. Subject's ratings of partner credibility (Means)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Character</th>
<th>Truthfulness</th>
<th>Competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-only</td>
<td>4.89</td>
<td>4.40</td>
<td>4.74</td>
</tr>
<tr>
<td>Text and voice</td>
<td>4.71</td>
<td>4.50</td>
<td>4.49</td>
</tr>
<tr>
<td>Text, voice, and image</td>
<td>4.84</td>
<td>4.70</td>
<td>4.78</td>
</tr>
<tr>
<td>Voice and animation</td>
<td>4.79</td>
<td>4.20</td>
<td>5.02</td>
</tr>
<tr>
<td>Text, voice, and animation</td>
<td>5.04</td>
<td>4.30</td>
<td>4.88</td>
</tr>
<tr>
<td>Unscripted face-to-face</td>
<td>5.09</td>
<td>5.30</td>
<td>5.06</td>
</tr>
<tr>
<td>Total</td>
<td>4.89</td>
<td>4.57</td>
<td>4.83</td>
</tr>
</tbody>
</table>

Table 4 (cont.). Subject’s ratings of partner credibility (Means)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Dominance</th>
<th>Dynamism</th>
<th>Sociability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-only</td>
<td>5.35</td>
<td>4.63</td>
<td>4.60</td>
</tr>
<tr>
<td>Text and voice</td>
<td>4.60</td>
<td>4.03</td>
<td>4.05</td>
</tr>
<tr>
<td>Text, voice, and image</td>
<td>4.95</td>
<td>3.77</td>
<td>3.75</td>
</tr>
<tr>
<td>Voice and animation</td>
<td>5.40</td>
<td>3.97</td>
<td>4.05</td>
</tr>
<tr>
<td>Text, voice, and animation</td>
<td>5.35</td>
<td>4.13</td>
<td>4.10</td>
</tr>
<tr>
<td>Unscripted face-to-face</td>
<td>4.80</td>
<td>4.77</td>
<td>5.10</td>
</tr>
<tr>
<td>Total</td>
<td>5.08</td>
<td>4.22</td>
<td>4.23</td>
</tr>
</tbody>
</table>

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 this first in a series of investigations was undertaken.

The results produced to date are informative in some senses and mystifying in others. Consistent with the argument that social identification is a key consideration in assessing communication formats, partners were seen as more sociable, likable, dynamic, and truthful when participants engaged in face-to-face than human-computer interaction. Gains in credibility were also associated with greater understanding, and yet, this did not translate into greater influence. Computerized agents were more influential on the whole than were human partners. Contrary to the hypothesized benefits of greater anthropomorphism, there was a (nonsignificant) tendency for greatest influence and high-quality decisions to materialize 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. It is bound to prompt deeper exploration of the interrelationships among these variables and properties of computer agents. In light of the pilot nature of this experiment, we are unable to offer definitive explanations for these findings yet, but we submit for consideration the speculations that follow.

7.1. Influence, Decision Quality, and Understanding

The planned contrast between FtF and the combined HCI conditions revealed that the least absolute and relative influence was exerted in the unscripted face-to-face condition. This same condition achieved the lowest decision quality, on average. The finding that the collective computer versions were 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 credibility ratings related to competence. 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 tenuous opinions.

Such a conjecture warrants further exploration of the basis for participants’ judgments, some of which appear in the companion paper. However, two immediate implications merit comment. First, anthropomorphic qualities and high degrees of interactivity, at least in terms of modality richness, may be unnecessary to achieve influence. Second, and paradoxically, those qualities may need to be included when the desire is to mitigate what might otherwise be undue influence by a computer agent. If, for example, the aura of knowledgableness and validity conferred by the computer medium leads users to passively and mindlessly accept the information they receive, in essence abdicating their critical assessment responsibilities, then the addition of more human-like qualities may actually be necessary to make computers less influential.

Although the HCI conditions did not differ among themselves, we noted a tendency for the text and voice condition to generate somewhat higher influence and decision quality scores. If this finding were to replicate in further studies with larger sample sizes, it might lend credence to what Kiesler and colleagues have dubbed “the black sheep effect,” namely, that imperfect representations of human-like qualities detract from the effectiveness of computerized agents. This possibility is developed further when we consider the credibility results. However, it should be noted that if idiosyncracies in the computer interfaces were distracting, they were insufficiently so to impair understanding, as the extent to which participants could recall their partner’s rankings and could paraphrase their partner’s reasons were unrelated to how much they were influenced by the partner. It is possible, though, that the limited variability in understanding was more a function of the task than the nature of the communication interface. 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.

7.2. Credibility: Judgments of Character, Competence, and Sociability

As predicted, unscripted face-to-face interaction generated higher perceptions of credibility than did human-computer interaction. FtF partners were seen as more sociable, truthful, and dynamic than were HCI partners. However, contrary to predictions that credibility ratings would increase with corresponding increases in anthropomorphism, the text-only condition earned the highest ratings on dominance, and the other credibility dimensions showed highly variable patterns across conditions. Also though not meeting conventional standards of statistical significance, the condition combining text, voice, and still image tended to garner lower ratings on sociability and dynamism.

These findings are puzzling. However, two explanations can be offered. 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 or belittlement by the subjects. The synthesized voice, however, as indeed any voice, is unique and particular thus opening up for all kinds of attributions on the part of the subject. Even more so with a facial image, not to mention the facial animation!

A second possibility—or perhaps a more detailed account of the black sheep effect—is the following conjecture:

In any communicative setting involving two parties, previously unknown to each other, the communication partners remain the ideal partners to each other in every aspect that is not explicitly revealed during the discourse.

This conjecture is reminiscent of what has been called “hyper-personal communication” [38] in the computer-mediated communication literature. These are interactions that create such a high sense of personalism that they actually exceed face-to-face 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; to construct and reciprocate representations of their partner and relationships without the interference of environmental reality; perhaps more so when this communication is asynchronous and when the computer-mediated communication link is the only one there is.

The correlations between the influence and credibility measures were near zero which means that respondents’ influence by the interaction partner was not due to seeing him as more or less credible.

7.3. Future Work

This investigation represents a first pilot effort and is presented more to demonstrate the ways in which features of interfaces could and should be systematically compared than to provide any definitive conclusions about what properties produce what effects. More definitive conclusions must await larger sample sizes, which might generate more significant results.

With regard to the computer conditions, two additional features should be added to future experiments to explore anthropomorphism in human-computer interaction: 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 pilot study were purely practical. The fact that no “intelligence” was employed for the dialogue script, however, has to do with the fact that we wanted to adopt an application perspective, allowing for implications of our findings to 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; alternating partner versus subject initiation of the conversation, and so forth, to test additional hypotheses.

The scripted face-to-face condition tended to become fairly unnatural and subjects seemed to experience it as rather awkward. In light of this, a scripted FtF control condition is needed that is more realistic but maintains control over the verbal component.

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 face-to-face interaction is best for generating positive social judgments and interpersonal relationships, but that
human-computer interactions are more influential. However, this cannot be attributed to computers uniformly being seen as more competent since there was no significant difference between the computer and face-to-face conditions in this respect. Given a larger sample, though, it is plausible that such a difference may be demonstrated.

Maybe this is but a transient phenomenon—much like the panic that ensued at the first public presentations of the cinematograph 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.

References


