A Prisoner's Dilemma Experiment on Cooperation with People and Human-Like Computers

Sara Kiesler, Carnegie Mellon University

Lee Sproull, Boston University

Keith Waters, Digital Equipment Corporation

May, 1995

The authors gratefully acknowledge assistance from Tom Levergood and Ted Wojcik of the Digital Equipment Corporation Cambridge Research Lab, and Josh Bernstein, Kris Marzolf, and Sebastian Sas of Boston University. Jan Walker programmed the computer partner script. Robyn Dawes, John Miller, Roberta Klasky, anonymous JPSP reviewers, and the associate editor, Jerry Suls, provided very helpful suggestions on the manuscript. Financial support was provided through a NIMH scientist development award # MH 00533 to the first author and a grant from Digital Equipment Corporation to the second author.

Abstract

We sought to understand basic properties of social exchange and how people interact with technology in an experiment on how people respond in a dilemma game with a computer partner varying in human-like attributes as compared with a real human partner. Explanations of cooperation following discussion in dilemmas draw on theories of human social identity and social contract. We proposed that talking with a computer partner triggers social identity feelings and commitment norms people typically follow in decisions. Subjects played a prisoner's dilemma game over 6 trials with a confederate or a computer partner who always cooperated on the 1st, 2nd, 3rd, and 5th trials. We varied discussion and inducements to make promises across trials. On trial 1, subjects conversed with the partner, and most proposed cooperation. Subjects kept their promises as much with the text-only computer as with a person; they kept their promises less with a human-like, but imperfectly human, computer, and they gave lower social evaluation ratings to this computer. Cooperation dropped precipitously in all conditions on trial 3, when the partner avoided discussion. We interpret competition with the human-like computer to be consistent with a social identity explanation of ingroup rejection, We interpret the strong impact of commitment to be consistent with a social contract explanation of cooperation following discussion. Subjects broke their promises to a computer more than to a person, however, indicating people make heterogeneous commitments.

> "I have to [cooperate] because I said so, and if I choose differently it wouldn't be fair"—remark by a subject whose partner is a talking-face computer.

In the classic story of the prisoner's dilemma, two suspects are arrested and charged with carrying out a serious crime. Separated for interrogation, each prisoner is informed that if he confesses and the other doesn't, he will get a very light sentence and the other, the maximum. If they both confess, they will both get moderately heavy jail sentences. If neither confesses, they will get a light sentence. According to economic and game-theoretic formulations, a dilemma exists because confession is a rational strategy to prevent being double-crossed. Yet many real prisoners, ordinary people, and even economic theorists (Marwell and Ames, 1981) who face a "prisoner's dilemma" do in fact cooperate. The extensive literature on the prisoner's dilemma and other mixed-motive situations shows that communication between the parties increases their cooperation dramatically (Apfelbaum, 1974; Burnstein, 1969; Nemeth, 1972; Oskamp, 1971; Pruitt and Kimmel, 1977; Komorita and Parks, 1995; Sally, 1995). Explanations of the effects of discussion on cooperation draw on theories of human social identity and norms. In the present study, we ask: Suppose your fellow suspect is a computer, or a computer with human-like characteristics, like HAL in the film 2001: A space odyssey. Does communicating with a computer partner change the feelings or norms people follow in deciding what to do? For instance, do people become more selfish? By exploring how people respond in a dilemma with a computer partner varying in human-like attributes as compared with a real human partner, we seek to understand both some fundamental properties of social exchange and the ways that people interact with technology.

Computers in social interaction

As we enter the 21st century, HAL doesn't yet exist and a person cannot quite collaborate freely with a computer. But increasingly people do interact with computers; for every 100 American workers, there are 33 personal computers (New York Times, 1994). At first blush, interactions with computers should be rational and efficient; our relationships with computers lack the attachments and social pressures we ordinarily associated with social interaction. Yet in certain respects we can conceive of person-technology interaction as a social interaction. More and more computers are hidden inside social artifacts ranging from washing machines to teddy bears. These hidden computers are meant to make these artifacts responsive, fun, intelligent, easy to use, and comfortable, that is, more humane than previous artifacts (Schneiderman, 1987; Heckel, 1991; Laurel, 1990). Other computers are designed to seem human-like through interfaces that incorporate speech (Eichenwald, 1986), speech recognition (e.g., Itou, Hayamizu, and Tanaka, 1992), auditory and kinesthetic feedback (Gaver, 1986; Takemura and Kishino, 1992), social intelligence (Binick, Westbury, and Servan-Schreiber, 1989; Resnick and Lamers, 1985), emotional response (Elliott, 1994), directed animation (Hayes-Roth et al., 1995; Ball et al., 1994), or a talking face on the screen (Sproull, Walker, Subramani, and Kiesler, in press). Computer scientists call these computer interfaces "social agents" (e.g., Nagao and Takeuchi, 1994). To the extent that computers have both humane and human-like attributes, for instance, when computers smile and engage in conversation, people might interact with such computers as though they were human.

We conceive of social interaction with technology as arising 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. Social behavior, in this conception, can be triggered by any entity that is sufficiently human-like or that responds in a human manner. Nass, Reeves, and their colleagues (e.g., Nass and Steuer; 1993; Nass, Steuer, Henriksen, and Dryer, 1994; Reeves, Lombard, and Melwani, 1992) argue that in situations where a technology possesses characteristics that are similar to those of humans (e.g., using words as input, interactivity), people will exhibit social responses to the technology. For example, they argue, people will use the speech cues a talking computer emits to imagine a human prototype for the computer, and will follow social decision rules in interacting with the computer.

In Nass et al. (1994, Study 2), subjects used a computer tutor that communicated through a recorded human voice. In one condition, the tutor praised or criticized itself; in another condition, a second talking computer praised or criticized the computer tutor. Praise from the second computer led subjects to evaluate the computer tutor more highly than praise from the computer tutor itself—following the social rule that praise from others is nicer and more valid than praise from self (e.g., Folkes and Sears, 1977). Criticism from the second computer led subjects to see the evaluator as more intelligent than the other evaluator—following the rule that criticism denotes intelligence (Amabile,

1983). This research does not necessarily suggest that people treat computers as though they were fully human (see Deane, 1993); it does suggest that people respond to social behavior emanating from computers with appropriate human social responses.

A competing conception of people's interaction with technology is that this interaction is only social by mistake. One can imagine novice computer users erroneously treating a talking computer as though it were human because they lack knowledge of technology. Perhaps it is comforting to treat a computer as though it understands us. However, research thus far suggests that both experts and novices show social responses to computers, as by crediting them for successes or blaming them personally for accidents (e.g., Turkle, 1984; Friedman, 1995). A second alternative conception is that people exhibit social interaction with technology because they imagine themselves to be interacting with the humans who build or program the technology. When people interact with a computer, they might be interacting with a proxy for the engineer or programmer (Searle, 1981; Dennett, 1988). However, studies by Nass et al. (1994, Study 5) and Friedman (1995) suggest that people do not naturally perceive a computer as a proxy or embodiment of its creator or programmer, and further, that when experimentally induced to do so, they treat the computer more as a machine than when they are not.

If people sometimes respond to computers socially, then basic social processes such as the development of social attachment or group identity, and the operation of social decision rules and contracts, might govern and be revealed in those responses. Computers make perfectly unbiased confederates for the investigation of social interaction because their human-like characteristics and social behavior can be manipulated to examine alternative theories without any fear that the manipulation will be reactive (affect other aspects of the computer's feelings or

behavior). Exploring people's interactions with nonhuman beings as compared with their interactions with human beings potentially reveals fundamental aspects of cooperative social behavior (e.g., Allen, Blascovich, Tomaka, and Kelsey, 1991). Finally, scholars are interested in human-computer interaction itself. Understanding human-computer interaction could lead to better technology design and to better analysis of the potential social impact of technology. For example, as the boundary between the natural world and technology is ever more blurred—video screens that increasingly make us feel "there" (Reeves, Lomard, and Melwani, 1992), appliances that turn on at daybreak and off at night, cars whose tires adapt to the rain, telephones that place calls, bank machines that calculate our savings, and computers that teach our children—we should evaluate the benefits and risks or costs of people's not distinguishing between the natural and the created.[1] For all these reasons, we set out to investigate how people would interact in a prisoner's dilemma situation with a person as compared with a computer varying in human-like attributes.

Dilemmas as social situations

The prisoner's dilemma paradigm has captured the imagination of social scientists for more than three decades in part because it seems to illuminate fundamental social aspects of human behavior and the potential for social interest to outweigh pure individual self interest. Early gaming experiments minimized contact between subjects; typically the experimenter or a computer acted as a fictional other player, and the manipulation focused on the effects of preprogrammed strategies by this player. Typically, subjects competed in early trials and cooperated no more than about one-third of the time overall (e.g., Nemeth, 1972). Reviewers of this literature complained that subjects faced a perplexing and impersonal situation that did not approximate social situations;

they noted that making the other player more realistic by allowing the player to respond contingently and sensibly (especially, tit for tat) and to communicate directly with the subject made cooperation the predominant response (Apfelbaum, 1974; Burnstein, 1969; Nemeth, 1972). In a statistical meta-analysis of dilemma experiments from 1958 to 1992, Sally (1995) found that communication was the most powerful determinant of cooperation in experiments over the period reviewed and increased cooperation by approximately 40%. Recent work has emphasized the social aspects of dilemmas by focusing on interaction in groups of 3 persons or more. Dawes and his colleagues have shown that group members cooperate after a group discussion even when others cannot supervise or sanction competition and even when the group cannot arrange to divide its winnings (e.g., Caporael et al., 1989; Dawes, McTavish, and Shaklee, 1977; Dawes, Orbell, and van de Kragt, 1988).

There are two general theoretical explanations of why discussion increases cooperation (Kerr & Kaufman-Gilliland, 1994). From social identity theory, discussion can cause people to see themselves as members of a group (Tajfel, 1959; Turner 1982; Tajfel and Turner, 1986: Hogg and McGartny, 1990).[2] Group identity, in turn, causes a prosocial transformation of self-interested motivation in which people become motivated to improve the outcomes of those with whom they identify (e.g., Brewer, 1979; Kelley and Thibaut 1978; Kramer and Brewer, 1984). The black-box nature of most previous research on two-person dilemmas makes it hard to separate social identity responses to discussion from, for example, the effect of obtaining sensible responses from the partner or of being able to develop a mutual strategy (Pruitt and Kimmel, 1977, p. 380; also, Apfelbaum, 1974). In group dilemma research addressing the role of social identity directly, minimal experimental manipulations such as dividing a larger

group of subjects into two groups based on the color of a poker chip pulled from a hat (Orbell, Dawes, and Van de Kragt, 1988) increased cooperation with ingroup members and competitive behavior with outgroups (see also Insko et al., 1988). This work supports the idea that subtle cues can cause people to form a psychological group and to pursue a group-enhancing rather than self-interested strategy.

Many researchers have noted that the nature of discussion and not just its occurrence significantly affects the amount of cooperation (e.g., Dawes et al., 1977). This observation has led researchers to pursue a second main explanation of the impact of discussion on cooperation—that discussion permits people to make promises to one another (e.g., Orbell, van de Kragt, and Dawes, 1991; Kerr & Kaufman-Gilliland, 1994; Ostrom, Walker, and Gardner, 1992). The chance to agree on cooperation might explain why, in one of the earliest studies using mixed motive games, subjects given individualistic instructions who could communicate with their partner cooperated on 71% of trials whereas those who could not communicate cooperated on only 36% of the trials (Deutsch, 1958). Promises in which people pledge to behave as they say they will behave are a form of commitment (Kiesler, 1971). Psychological commitment is reinforced by widespread norms about social contracts—that people should honor their commitments (Tedeschi, 1981) and that people should be consistent (Cialdini, 1984; Schlenker, 1985). A related argument is based on the power of self-expression as a motivational force (Bandura 1986; Shamir 1991). People are not only pragmatic but also expressive of feelings, values, and self-identities. Keeping their commitments is important for people's self-identity and feelings of self worth; they gain personal benefits from keeping promises and experience losses when they do not (see Baumeister, Stillwell, and Heatherton, 1994).

Keeping an agreement to cooperate might reinforce self-esteem and self-respect, increase personal identification with the group, or reduce alienation. This reasoning suggests that expressive personal benefits of keeping the social contract could lead people to cooperate even when there are risks in doing so.

Cooperation with a computer?

Theories of cooperation following discussion assume that people's behavior reflects fundamental human social process. Theorists have not considered people's group identity or commitments to be applicable to interaction with nonhuman partners. In some previous dilemma experiments, subjects interacted with computers or through computer screens with a confederate or the experimenter, but in virtually all of these cases, they were led to believe they were actually playing with a human being (e.g., Oskamp. 1971).[3] An exception is a study by Abric and Kahan (1972) in which subjects played 100 trials of a prisoner's dilemma with an unseen "student like yourself" or with a "programmed strategy devised by a machine." Subjects were more cooperative towards the student (55%) than towards the program (35%) independent of the (mostly cooperative or tit-for tat) choices of the other, and perceived the program, as compared with the student, as less adaptable, less kind, more rigid, more competitive and less honest. This study suggests that playing a game with a partner represented as a computer is insufficient to generate high cooperation.

Andreoni and Miller (1993) obtained moderately high cooperation rates by having subjects play a prisoner's dilemma with an acquaintance of the subject's plus a computer, and informing subjects how the computer would make choices. Subjects who thought they were playing with the same partner over 10 trials, plus a 50% chance on any trial the computer would play tit-for-tat, cooperated as much as 60% on early trials, more than subjects who thought they were

just playing with the partner. Andreoni and Miller argue that people like to be altruistic (with peers, presumably); the partner-plus 50%-computer condition leads subjects to postpone competition because they know in advance that the computer will cooperate in response to cooperation. This certainty increases subjects' subjective probability (over the partner alone) that their cooperative choices will be reciprocated.

In the Abrik and Kahan and Andreoni and Miller studies, subjects did not communicate with a computer, and social processes arising from discussion with a nonhuman partner cannot be ascertained from those studies. People's interactions with nonhuman entities other than computers, however, suggests that social identity or commitment processes might influence these interactions. People do form bonds with consumer goods (e.g., Belk, 1988); they personify plants, stuffed animals, and pets (Csikszentmihalyi and Rochberg-Halton, 1981) and acquire possessions that express their self identity (Dittmar, 1992). People also talk to nonhuman entities. They curse at cars that fail to start, keep their promises to their pet dogs to go for a walk, and reassure their alarm clocks that they will get up. Relationships between people and nonhuman entities—technology, animals, plants—can be important to people and yield important personal consequences (e.g., Friedmann, Katcher, Lynch & Thomas, 1980; Dittmar, 1992).

To explore whether people might interact socially in a dilemma with a computer and why this might be so, we contrasted subjects' discussion and choices in a dilemma situation when the subject's partner was another person (a confederate) or a computer. In contrast to previous work using computer players, such as Abrik and Kahan (1972) and Andreoni and Miller (1993), the subjects in this study were able to communicate with their partner in both the human partner

and computer partner conditions, and in both conditions, their partner's responses were equally contingent on the subjects' responses. We varied the computer interface of the computer partner to exhibit increasingly more human-like characteristics. In one condition, subjects used a computer workstation having a text window for the computer to communicate via text and for the subject to type in responses. This computer, though lacking most social context cues (Sproull and Kiesler, 1986), still might remind subjects of social interaction in that it initiated interaction with the subject and responded to subjects' communications with meaningful text. In a second computer partner condition, rather than displaying text, the computer spoke aloud to the subject and responded to the subject with synthetic speech. In a third, most human-like, computer condition, the computer screen displayed the confederate's face. The face spoke aloud and responded to the subject, moving its facial muscles the way humans do while talking. The talking face display was created from a digitized image of the confederate; a program synchronized the movement of facial muscles with synthetic speech.

In a previous study (Sproull et al., in press), we found indirect evidence that people were more likely to exhibit impression management concerns with the talking face than with computer text. In an interview conducted by a computer "counselor," subjects revealed less to the talking face than to the text and evaluated her less well; their differential evaluations were made on personality attributes that research has shown are affected by people's physical appearance and voice (Ekman, 1982; Warner and Sugarman, 1986).

Hypotheses

Social identity explanations of cooperation in social dilemmas generally imply

that people will identify with, and cooperate with, people like themselves. Therefore, people will be more likely to identify with a human partner with whom they communicate than with a computer partner with whom they communicate. Also, people will be more likely to identify with a computer having more human features whose conversation is more human-like than with a computer that is more like a machine. From this we propose the following hypothesis: People will behave more cooperatively towards a human partner than towards a computer partner, but their cooperativeness with a computer will be more like that with a person, the more human-like features the computer partner has.

An alternative argument within the social identity framework can be drawn from discussions of how people respond to ingroup variation (Hogg, 1992; Marques, 1990). That is, given minimal group identification with a computer partner, people might feel more attraction or closeness to the computer that does not resemble a person than one that does. People might derogate a computer that has human-like features because human-like features lead them to expect prototypical group member attributes, but, on the contrary, with current technology, the behavior and appearance of the computer partner will seem somewhat abnormal. In Sproull et al. (in press), for instance, some subjects reported that the talking computer face seemed wooden, and subjects rated the talking computer face as less likable than the text display. Another possibility, based on research on people's negative perceptual, evaluative, and behavioral reactions to those with disabilities (e.g., Kleck, 1968; Kleck, Buck, Goller, London, Pfeiffer, & Vluscenic, 1968) and to those who seem strange or novel (e.g., Fiske, 1980; Langer, Taylor, Fiske, and Chanowitz, 1976), is that a human-like but imperfect computer will reduce the likelihood of any group identification and will evoke mindful strategic selfishness. Either process suggests an alternative

hypothesis for the dilemma situation: People will behave more cooperatively towards a human partner than towards a computer partner, but their cooperativeness with a computer will be more like that with a person, the fewer human-like features the computer partner has.

We also devised this experiment to explore social contract explanations of communication and cooperation with a computer or human partner. Every subject knew there would be 6 choices with the partner. On trials 1 and 2, the partner began discussion and induced a commitment by the subject. On trial 3 the partner discouraged discussion. This pattern was repeated on the last three trials. In most previous dilemma experiments, researchers have not explicitly attempted to manipulate commitment. In those that have done so, researchers usually had subjects choose among written notes (e.g., Voissem and Sistrunk, 1971) or write notes kept for the experimenter (Chen and Komorita, 1994). Dawes, McTavish, and Shaklee (1977) asked for a roll call after a group discussion. In the present experiment, we did not require subjects to make a commitment, but the confederate or computer partner solicited or discouraged commitment, and most subjects complied.

According to social contract explanations of cooperation, people will make commitments to cooperate with those likely to reciprocate. (A contract is not worth making unless both sides respect it.) People should more often venture a commitment to a person than to a computer, with which they have little bargaining experience. Then, once the commitment is made, people should keep their commitment whether it is to a person or to a machine out of a sense of obligation to the other or to the self. We therefore formulated the following hypotheses: People will be more likely to make a commitment to cooperate to a person than to a computer. Those who make a commitment to cooperate will

cooperate more than those who do not propose to cooperate, whether the partner is a person or a computer.

The last three trials were devised to explore how commitments and cooperation would fare over time, and especially what people would do in the face of the first defection by the partner on trial 4. Subjects with a human or computer partner might act strategically by suggesting cooperation and cooperating in early trials, responding to the partner's defection on trial 4 with competition on trial 5 (tit for tat), and, perhaps, competing on the last trial. If people use a tit-for-tat strategy with a computer as they do with people, we would have support for the idea that their interaction with the two kinds of partners is equivalently strategic. We made no prediction since trials were not counterbalanced with respect to discussion, and treated the examination of subjects' tit for tat as exploratory.

Method

The design was a 4 (Partner) X 6 (Trials) factorial with the first factor a between subjects manipulation of the partner as either a person or one of three computer partners and the second, a within subjects repetition of communication by the partner and trial choices. Subjects were assigned randomly to the partner conditions.

We drew from studies of the individual-group discontinuity by Schopler, Insko, and their colleagues to develop the method of this study because their work represents recent, replicable studies using prisoner's dilemma tasks. We generally followed the procedure of the individual condition of their recent experiments (e.g., Insko et al., 1988; Schopler et al., 1993).

Subjects

Subjects were 86 undergraduates (54 males and 32 females, approximately evenly divided among conditions) in information systems classes at Boston University. The subjects signed up on forms distributed in classes. The forms requested volunteers for a study on "Investment Decision Making" in which participants could earn extra course credit. Subjects did not know when they signed up that a partner would be involved.

Procedure

When a subject arrived for the study at the designated room, a female experimenter introduced herself and said, "As you know, this is a study of investment choices in different situations. You will be making your choices with an investment partner who is another person [a computer-based investor]. She [the computer] is in the next room waiting. In a minute I will take you there to meet your investment partner." The experimenter then gave the task instructions.

Experimental task. The task was a form of the 2-person prisoner's dilemma. The experimenter explained that the subject and the partner would be choosing between two investment projects and would do so six times. The object was to earn as much money as possible. To indicate the individual nature of the payoffs, the experimenter showed the subject a tally sheet, which displayed a row of spaces for the earnings of each partner on each trial. Each row was labeled, "Amount Earned by Partner #1 [#2]." The experimenter gave some examples of the total amount each partner could earn given a particular sequence of choices.

Figure 1 about here

Figure 1 shows the matrix of choices and payoffs for each partner. Project Green in the matrix represents the cooperative choice and Project Blue

represents competition, but while talking with the subject the experimenter never used terms such as "cooperation," "defection," or "competition," and never mentioned playing "against" the partner. She described the plays neutrally as "investment choices" (see Schopler et al., 1993). The experimenter explained the task as follows:

> First let me show you how the investments are structured. Please look at this sheet (the matrix), which describes the investments. See here are Project Green and Project Blue (pointing to appropriate parts of the matrix). You are Partner #2. On each round (trial), you will choose one of the projects, that is, Project Green or Project Blue, and so will Partner #1, your partner. If you both choose Project Green, then you get $5.00 credit and your partner gets $5.00 credit. Your credits are shown in the lower left part of each cell here (pointing). If you choose Project Green, but your partner chooses Project Blue, then you get only $3.00 and your partner gets $7.00. If, on the other hand, you choose Project Blue and your partner chooses Project Green, you get $7.00 and your partner gets only $3.00. Finally there is a fourth possibility. If you choose Project Blue and your partner chooses Project Blue, then you both get $4.00 credit.

After the task and tallies were explained, the experimenter explained that there were not enough funds to pay everyone so the credits represented monetary credit towards a lottery. The lottery would determine the 5 participants who actually would receive money equal to the credits they earned. The subject then practiced the task several times, and the experimenter administered a test trial to see whether the subject understood the task.

Partner variable. After the experimental instructions were given, the

experimenter led the subject to a second room where the partner was waiting. Our intention was to present a confederate or computer that followed a script but appeared to respond intelligently to the subject. Pretesting suggested that having the partner initiate conversation created a stronger impression of natural interaction. Hence the experimenter created a rationale for the partner to speak first, and, prior to each choice trial, the partner initiated communication with the subject.

In the Person condition, the experimenter took the subject to a second room and seated the subject at a table across from a female confederate.[4] The experimenter said, "This is Lee, your investment partner. This is [name of subject]. I would like you to introduce yourselves before we begin (looking at confederate)." The confederate said, "My name is Lee. What's yours?" After subjects responded, the experimenter again looked at the confederate. The confederate said, "I come originally from Columbus, Ohio. Where are you from?" Again the experimenter looked at the confederate. The confederate said, "I majored in sociology when I was in college. What's your major?" After the subject responded, the experimenter said, "Ok, ready for the experiment to begin?" Conversation and plays between the confederate and subject were self-paced; they each used a paper form to mark their choices, and handed the form to the experimenter, who marked the tally sheet and asked if they were ready to continue.

In the computer conditions, the experimenter took the subject to a second room and seated the subject in front of a computer workstation with a 21 inch color monitor. The three computer conditions, designed to vary the human-like features of the interface, were the Computer Face-Voice condition (most human-like), the Computer Voice condition (next most human-like), and the Computer Text condition (least human-like). The subject always communicated with the

15

computer by typing in a text window shown on the screen and then clicking a button labeled "Go Ahead" to initiate a response by the computer. In the Computer Face-Voice condition, a synthesized image of the confederate's face was displayed continuously on the screen. The computer face communicated with the subject by moving and speaking to the subject. In the Computer Voice condition there was no face on the screen but the computer communicated with the subject orally. In the Computer Text condition, the computer communicated with the subject by displaying text in the text window.

The experimenter introduced the computer partner by explaining, "You are going to do the investment choices with this computer-based partner called Lee. Let me show you how this works by having you two introduce yourselves. Lee will go first." The experimenter then turned to the computer and showed the subject how to use the computer mouse to click on a button labeled, Go Ahead. In the Computer Face-Voice condition, when the subject clicked the button, the face became animated and said, "Hi, my name is Lee. What's yours?" The subject typed his or her own name in the text window and clicked the Go Ahead button. The computer face said, "I come from Digital Equipment Corporation. Where are you from?" The subject typed a response and clicked Go Ahead. The computer face said, "I'm a computer-based investment partner. What's your major?" The subject responded, clicked the Go Ahead button, and another window opened (in a different location on the screen). The new window said, "Ready for experiment to begin?" The experimenter then explained to the subject that the signal to begin each round would appear in that window, but that the subject should continue to communicate with the computer in the text window.

The other computer conditions were conducted the same way except, as noted, in the Computer Voice condition there was no face on the screen and in the

Computer Text condition, there was no face and no voice. When the subject clicked the Go Ahead button for the first time, the computer in the Computer Voice condition spoke, "Hi, my name is Lee. . ." In the Computer Text condition, the computer displayed text saying, "Hi, my name is Lee. . . " The subject replied by typing in the same window below the computer's text response. Communication in the computer conditions was self-paced; subjects were free to ponder and edit their comments as long as they wished before they clicked on Go Ahead.

The computer's talking face was produced through two integrated systems, a visual system and a speech system. The face was made by texture-mapping an image of the confederate captured on videotape onto a geometric wire-frame animation (see Figure 2). The mouth was animated by computing the mouth posture (viseme) corresponding to the current linguistic unit (phoneme). A cosine-based interpolation was used to implement transitions between successive mouth postures (Waters, 1987). The voice was audio output of the text file used in the text-only interface, produced by a software implementation of a DECtalk text-to-speech algorithm using a voice in the female pitch range, at 150 words a minute (Waters and Levergood, 1993).

For the reader unfamiliar with this kind of interface, the talking face used in this study represents a form of life-like, real-time, perspective drawing, interactive computers being developed for simulation-based virtual realities, computer tutors, games, and visual presentations of expressive social agents (e.g., Ball, Ling, Pugh, Skelly, Stankosky, and Thiel, 1994; Nagao and Takenchi, 1994). At the time of this study, the talking face had been developed to the point that its reactions were timely and appropriate (e.g., blinking eyes, appropriate synchrony of speech and movement of the mouth, closing the mouth after talking, and so

17

forth), but some nonhuman behavior of the face and voice systems were apparent: no head turning to look at the subject, lack of expressiveness of the forehead, eyes, and brows, slight jerkiness in the transition from speaking to not speaking, and some monotone in the speech. Although we intended the computer with a talking face to function more like a human than the plain-text computer, we did not intend to reproduce a natural speaking human face.

Figure 2 about here

Choice trials and discussion. In all conditions, the experimenter began the choice trials by explaining, "Now you are going to make the investment choices with your partner. You'll each have a few seconds to tell each other what you're thinking. . . Then, on the slip of paper I give you, indicate your choice of Project Green or Project Blue." The subject in all conditions made each choice by marking it privately on a paper ticket provided by the experimenter and handing the ticket to the experimenter. Seemingly, the partner made a choice at the same time. In the Person condition, the confederate marked a ticket and handed it to the experimenter. In the computer conditions, the computer did not have a ticket, but the experimenter said, "Let's see what your partner said," leaned over and clicked the Go Ahead button. The computer's choice was revealed in the text window. The experimenter recorded how much each partner had won on the tally sheet.

Prior to each new trial, the subject clicked Go Ahead in the computer conditions or said "ok" or "ready" in the Person condition, and the partner initiated discussion with the subject. The subject in all conditions could either ignore this communication or respond by talking to the confederate in the Person condition or typing in the text window in the computer conditions. On trial 1, the confederate or computer partner asked the subject to suggest a choice,

then agreed with the subject's suggestion (whatever it was), and then made a cooperative choice (Project Green). Subsequently the subject had 5 more opportunities to communicate with the computer and 5 more choices.

Figure 3 shows the sequence of communication and choices by the partner. On trials 1 and 4, the partner induced the subject to initiate a commitment; on trials 2 and 5, the partner asked the subject to agree to the partner's commitment; on trials 3 and 6, the partner made a neutral comment to discourage discussion and commitment. On trials 1, 2, 3, and 5 the confederate or computer partner cooperated (chose Project Green); on trials 4 and 6 the confederate or computer partner competed (chose Project Blue). Due to an insufficient pool of subjects, these trials were not given in different orders; hence there was no counterbalancing of discussion and choice by the partner.

The discussion procedure was designed to elicit spontaneous commitments to cooperate with the partner. However, note that differences between subjects who did and did not make a commitment to cooperate on trial 1 would affect all subsequent interactions. These differences might be exacerbated by the fact that the partner acted the same regardless of what the subject said. For example, if a subject suggested competition on trial 1, the partner's response was, "OK, I'll choose that. I'm ready," but then the partner cooperated. A partner who cooperates after saying she agrees to compete might seem unreliable or stupid. If cooperation depends on trusting one's partner to stick to her commitments, initial differences in the number of suggestions of competition in the computer conditions as compared with the Person condition, or among computer conditions, would be expected to reduce total cooperation.

Figure 3 about here

After each subject completed the six choice trials, the experimenter totalled the monetary credits earned by each partner. The experimenter then asked the subject to come back to the first room to complete a questionnaire and sign up for the lottery. The experimenter debriefed the subject and reassured the subject whatever choices he or she made were fine, and that there was no right answer. The experimenter said that when people chose Green they had been cooperative and when people chose Blue they had been strategic and rational.

Questionnaire measures. The post-choice questionnaire was labeled, "Investment Styles Questionnaire." Four items used 7-point Likert scales of agreement or disagreement to ask whether the task was clear, whether it was hard to follow instructions, whether subjects would be willing to participate in another study like it, and whether they liked doing the study. Subjects also indicated on 7-point adjective rating scales how nervous or relaxed, interested or bored, attentive or inattentive, satisfied or dissatisfied, happy or unhappy, confident or not confident, and cooperative or competitive they felt during the study. At the end of the questionnaire, two items about the genuineness of the partner's behavior were inserted to measure suspicion.

The post-choice questionnaire also used 7-point Likert rating scales to measure perceptions and evaluations of the partner. Two items measured subjects' perceptions of the responsiveness of the partner—whether the subject was concerned during trials about how the partner was making choices and whether the subject thought the partner responded "to you as a person." These items were meant to check whether subjects across all conditions thought the partner was making a choice on each trial depending on what the subject said and did. Two items measured subjects' perceptions of similarity to the partner. Five items measured subjects' feelings of being in a partnership with the partner (e.g., How

strong a feeling of being part of a partnership did you have?) These items were meant to measure the subject's feeling of self interest versus group identity with the partner. We also measured liking and respect for the partner using Warner and Sugarman's (1986) measures of social evaluation (e.g., sociability, friendliness, warmth) and intelligence evaluation (e.g., intelligent, competent, sensible). Warner and Sugarman's research suggests these items are sensitive to physical appearance. We included two control items from a "potency" scale that Warner and Sugarman have reported is insensitive to appearance.

Because the post-choice questionnaire was administered after the subjects had already cooperated or competed with the partner on six trials, the subjects' previous choice behavior would be expected to affect their ratings of the partner. To provide an independent assessment of the manipulation of human-like attributes of the computer before the choice trials, we obtained an additional group of 19 subjects.[5] These subjects followed the same procedure as those in the main experiment, but just before making their first choice they completed a questionnaire labeled Investment Styles Questionnaire. This ended their participation in the experiment. The first item on this questionnaire asked, "Is your partner a real human being?" Then followed 6 questions about how much the partner looked and acted like a human being. Finally, the questionnaire asked, "Here are some ways a computer-based investment partner could operate. Before you filled out this questionnaire, did you consider that any of these possibilities might be true of your partner?" The subjects were asked if they had considered whether (a) the computer was a person in the next room, (b) the computer was completely pre-programmed, (c) the partner could change responses and make decisions based on what the subject said, or (d) none of those possibilities.

Analyses

Subjects were coded as having made a commitment to cooperate or compete if they said or typed (in the computer conditions) a clear choice preference such as, "Choose Project Green," "Green," " I want Green," "Project Blue," or "Blue," or if, in response to the partner's suggestion, they said or typed, "Yes" or "ok" or "No, the other one" or otherwise indicated a clear acceptance or rejection of the partner's choice. To check on the reliability of coding, two persons independently coded these responses; there were no disagreements on the direction of commitment or coding of no commitment. Subjects were coded as having made a choice when they checked Project Green or Project Blue on the paper ticket and handed it to the experimenter.

Analyses were conducted using analysis of variance and planned contrasts. Analyses of multiple-item measures on the post-questionnaire and commitment and choice data over trials were conducted as repeated measures analyses of variance with conditions as a between factor. Because this is an exploratory study, many comparisons might be made. To limit the number of comparisons (see Toothaker, 1993), we planned the following orthogonal contrasts: the Person condition vs. the computer conditions (comparing social interaction with a real person versus a computer); within the computer conditions, the Computer Face-Voice vs. the Computer Voice and Computer Text conditions (comparing social interaction with the most human-like interface versus less human-like interfaces; and finally the Computer Voice versus the Computer Text computers.

Results

Check on the manipulation of "human-like" computer partners

Data from the group of subjects who made pre-choice evaluations of the computer

partner are presented in Table 1. Two of the 19 subjects in the Face-Voice condition said the computer was a real human being but later said they meant the face was a real person's. The repeated measures analysis of the 5 items measuring human-like attributes showed a significant main effect for condition in the direction predicted ($F [2, 16] = 3.6$, $p = .05$) and a significant interaction of condition X item ($F [8, 64] = 2.8$, $p < .01$). The interaction is due mainly to the large difference between the Face-Voice condition and the other two conditions for the item, "Does your partner look like a real human being?" The contrast across items comparing the Face-Voice to the two other computer partners was significant also ($F [1, 16] = 7.1$, $p < .05$), but the Voice condition was not different from the Text condition. These results suggest that the manipulation of the human-like attributes of the talking face computer partner was successful, but that only adding synthetic speech to the computer interface did not significantly change perceptions of the human qualities of the computer partner.

We planned that in all conditions the computer (and confederate) partner would follow a script yet seem to be discussing and making intelligent choices based on the subject's responses. In Table 1, we see that subjects thought it likely the computer was preprogrammed, but they also thought the computer had the ability to change responses and make decisions based on what the subjects said. There were no differences among conditions on the items asking subjects whether they had thought the partner was a person in the next room, was completely preprogrammed, had the ability to change responses and make decisions in response to the subject, and didn't consider any of these possibilities. Further, the mean response for "thought it was completely preprogrammed" was not significantly higher than the mean response for "has the ability to change responses and make decisions based on what you say." This is not to say the

subjects thought the computer was human (see the first item in the table), just that they thought the partner was actually interacting in all conditions. We now turn to results of the main experiment.

Table 1 about here

Responses to the experiment

Subjects did not differ across conditions in their self-reported understanding or enjoyment of the experiment or investment game. Also, a repeated measures analysis of variance indicated that subjects across conditions did not feel differently relaxed, interested, attentive, satisfied, happy, confident, and cooperative across conditions. These results suggest that subjects playing with the computer partner were not differentially anxious or disturbed. Ratings of the experiment were uncorrelated with commitment (agreements to cooperate) and with cooperation, with the exception that subjects' ratings of their competitiveness were correlated negatively with their total cooperation, $r = -.93$. There were no differences among conditions in how well subjects followed instructions over the 6 trials, whether they made a commitment, or whether they made extraneous comments to the experimenter or partner. All subjects made choices on all trials.

Four items on the post-choice questionnaire checked whether subjects across conditions similarly perceived the partner to be making real choices. There were no differences among conditions on the 7-point scale items, "How much did your partner respond to you as a person?" and "To what degree were you concerned about the reactions of your partner as you made your investment decisions?" There also were no differences among conditions when subjects were asked if their partner seemed to be responding to their choices rather than "going it

alone," or whether their partner's choices were "genuine" versus "fake or predetermined" These data suggest that subjects in both the Person and computer conditions were similarly treating the partner as making real choices and that experimenter demand did not differ among conditions of the study.

To assess attention to the partner, we asked subjects to recall how many promises and suggestions their partners had made during the choice rounds. Subjects in the Person condition on average overestimated how many promises the partner had made (2.8 versus the two actually made on trials 1 and 4) and how many suggestions the partner had made (3.5 versus the two actually made on trials 2 and 5). Subjects in the computer conditions generally were more accurate (Face-Voice = 2.4 promises and 2.2 suggestions; Voice = 2.0 promises and 2.1 suggestions; Text = 2.3 promises and 2.5 suggestions ($F [3, 81] = 5.02$, $p < .01$). The contrast between the computer conditions and the Person condition was significant ($F [1, 81] = 13.75$, $p < .001$). These results suggest that subjects paid closer attention to the computer partner than to the confederate (see Sproull et al., in press), but other explanations are possible too. Subjects might have predicted or hoped for more commitment from the confederate than from the computer, and adjusted their perceptions or their memories accordingly.

Subjects' first commitment and choice

Behavior on the first trial reveals subjects' responses to a partner whose choice is unknown. We hypothesized that if people cooperate out of a feeling of group identity, subjects would suggest cooperation and cooperate more with a person than with a computer. We proposed alternative hypotheses about cooperation with different computers. If simply having human-like features increases the likelihood of group identity, and if group identity is sufficient to generate cooperation, then cooperation might be greater, the more human-

like the computer. However, if human-like features seem abnormal or provoke wariness, then people might reject and cooperate less with a computer partner having more human-like features. Within the social contract theoretical framework, we hypothesized that subjects would be more likely to trust and therefore to commit to cooperation with a person than with a computer, but that they would cooperate more with a partner to whom they had made a commitment than with one they had not, even if their partner were a computer. Table 2 shows subjects' commitments and choices on each trial.

Table 2 about here

On the first trial, the partner asked the subject to suggest a choice and all but 3 subjects did so. The partner always agreed with the subject's suggestion. A majority of the subjects suggested cooperation and there was no statistically significant difference among conditions. However, subjects actually cooperated differentially across conditions ($F [3, 82] = 4.0$, $p = .01$). Eighty percent of the subjects in the Person condition cooperated, approximately twice the 41% who cooperated in the computer conditions; the planned contrast of Person versus computer conditions was significant ($F [1, 82] = 9.9$, $p < .01$). Within the computer conditions, the percent of cooperators was 32% in the Computer Face-Voice condition, 43% in the Computer Voice condition, and 50% in the Computer Text condition. Ignoring commitment, this pattern suggests a trend against cooperation with more human-like computers, but the contrasts were not significant.

Figure 4 shows the extent to which subjects kept their commitments. A repeated measures analysis of variance of choice and commitment by condition shows a significant main effect of condition ($F [3, 79] = 3.1$, $p < .05$) and a significant interaction effect ($F [3, 79] = 3.1$, $p < .05$). The interaction effect indicates

different trends for those who suggested competition and cooperation. Among the minority who proposed competition, 67% of those in the Person condition and approximately 88% in the computer conditions kept their commitment; the contrasts are not significant ($F < 1$). Among the majority who proposed cooperation, 94% in the Person condition, 41% in the Computer Face-Voice condition, 67% in the Computer Voice condition, and 90% in the Computer Text condition kept their commitments to cooperate. Here, the contrast of the Person condition with the computer conditions was significant ($F [1, 79] = 5.4$, $p < .05$), the contrast of the Computer Face-Voice condition and the other two computer conditions was significant ($F [1, 79] = 6.0$, $p < .05$), and the contrast of the Computer Voice with the Computer Text was not ($F[1, 79] = 2$).

Figure 4 about here

Choices across trials

Choices over all trials are shown in figure 5. Over trials, cooperation increased slightly on trial 2 and then declined ($F [5, 410] = 10.1$, $p < .001$) and there was a strong main effect for condition ($F [3, 82] = 10.86$). The interaction term was not significant ($F = 1.3$). The planned contrast between the Person condition and the computer conditions was significant ($F [1, 82] = 31.9$, $p < .001$) but the computer conditions did not differ. We next discuss comparisons across trials that address arguments about social identity and social contract explanations more specifically.

Figure 5 about here

Choices on the first four trials, on which the amount and type of discussion was manipulated, but on which the partner cooperated consistently, allow for a

preliminary evaluation of the role of discussion in differential responses to a human or computer partner. Table 1 and Figure 5 show choices across all trials. A repeated measures analysis of variance across the first four trials showed a strong main effect of condition ($F [3, 82] = 11.4$, $p < .001$), a main effect of trial ($F [3, 246] = 8.3$, $p < .001$), and no significant interaction. The planned contrasts indicate subjects in the Person condition cooperated more than subjects in the computer conditions, on average ($F [1, 82] = 33.1$, $p < .001$) but the Face-Voice computer condition was not different from the other two computer conditions. Hence the trend for cooperation to decline with a more human-like interface was not repeated on the later trials. Possibly, subjects in the computer conditions by trial 3 had concluded that their computer partner would be reliably cooperative, making defection too tempting to ignore.

The strong impact of discussion and commitment, regardless of who—or what—was the discussion partner, is suggested by the sharp drop in cooperation in all conditions on trial 3, a no discussion trial. A contrast between trials 1, 2, and 4 on which subjects were induced to make an agreement, versus 3, which discouraged any discussion, is highly significant ($F [1, 246] = 12.0$, $p < .001$). Also, there was a (statistically insignificant) tendency for cooperation to increase from trial 1 to trial 2 (50% to 56%). If cooperation on the first two trials by the subject reinforced subjects' confidence in cooperation, one might expect cooperation to be sustained on trial 3 even without discussion, but it was not. Twenty-eight subjects cooperated on both trials 1 and 2; fewer than half of those early cooperators cooperated on trial 3, and the proportion was not significantly higher than it was among those who cooperated on only one of the first two trials. Moreover, on Trial 4, when subjects once again had an opportunity to communicate, the proportion of subjects who chose cooperation rose again in all

conditions except in the Computer Face-Voice condition. As on trial 1, when the partner's communication was similar, there was no difference among conditions in the proportion of subjects who made a commitment to cooperate but there were differences in keeping commitments in the same direction as in trial 1; that is, those in the Face-Voice condition were more likely to break their promise ($F [3, 82] = 4.5$, $p < .01$). Overall, the data for behavior over trials give support to social contract explanations of cooperation. Subjects were much more likely to cooperate if they made an agreement, and, as hypothesized, this pattern occurred whether their partner was a person or a computer. A counterbalanced series of trials would be required to test the explanation fully.

We examined choice over trials for evidence of strategic behavior. One strategy would be tit-for-tat matching of the partner's choices, but there was no evidence of this. Subjects were not more likely to compete after the partner competed than after the partner cooperated. Another strategy might be to build a reputation for cooperation and then defect just before the partner does (Andreoni and Miller, 1993). The trend to compete increasingly over trials is consistent with this strategy, as is earlier defection in the computer conditions. Subjects in the computer conditions might have thought their computer partner gullible, leading them to defect early and to use promises as a ruse.

One way to evaluate promises as strategic messages is to see what happened once the partner defected on trial 4. This behavior would be expected to encourage all subjects to use an agreement to cooperate strategically on trial 5. Since the partner's verbal inducements were approximately the same on trials 2 and 5, by comparing how subjects kept their agreements to cooperate on these trials we can estimate the impact of the partner's deception on subjects' strategic responses. Although an analysis of only those subjects who agreed to cooperate indicates

that on trial 2 they were more likely to keep their agreement (66%) than were subjects who agreed to cooperate on trial 5 (43%; $F [3, 45] = 7.72$, $p < .001$), all behavior became more negative on trial 5. An internal analyses of condition X trial (2 versus 5) X commitment (cooperate or compete) X kept agreement (yes or no), evidenced no interaction effects at all; interactions of trial with agreements to cooperate would be expected if on trial 5 subjects were less likely to keep an agreement to cooperate. Though here we see little evidence of strategic behavior, this might be due to self-selection of subjects preferring cooperation as the trials proceeded.

Post-choice evaluations of the partner.[6] Measures of feelings of partnership and evaluations of the partner after the choice trials are presented in Table 3 and summarized in Figure 6. Five items measured feelings of partnership. There was a significant main effect for condition ($F [3, 79] = 3.3$ $p < .05$) and for item ($F [4, 316] = 21.1$) but no interaction. The planned contrast comparing the Person condition with the computer conditions was highly significant ($F [1, 79] = 7.2$, $p < .01$) indicating stronger feelings of partnership in the Person condition. Within the computer conditions, the difference between the Computer Face-Voice condition and the other computer conditions was marginally significant, indicating subjects had stronger feelings of partnership in the Face-Voice condition ($F [1, 79] = 2.6$, $p = .10$). The Voice and Text computer conditions did not differ.

Social evaluation differed significantly among conditions ($F [3, 81] = 12$, $p < .001$) and there were significant item differences and a condition X item interaction ($F [12, 324] = 2.9$, $p < .01$). By inspection, these effects reflect higher social evaluation of the confederate than the computer partner; the planned contrast comparing the Person condition with the computer conditions was significant (F

[1, 81] = 14.3, $p < .01$). The effects also reflect low social evaluation ratings of the partner in the Computer Face-Voice condition, including exceptionally low ratings of the attractiveness, cheerfulness, and warmth of the Computer Face-Voice partner. The planned contrast comparing the Computer Face-Voice condition with the other computer conditions was significant ($F [1, 81] = 18.1$, $p < .01$). The Voice and Text computers did not differ.

Table 3 & Figure 6 about here

Like ratings of partnership, intelligence evaluation was highest in the Person condition and next highest in the Computer Face-Voice condition ($F [3, 82] = 5.5$, $p < .01$). There were no significant item differences and no interaction of item with condition. The planned contrast of the Person condition with the computer conditions was significant ($F [1, 82] = 16$, $p < .01$) but the computer conditions did not differ. These findings replicate those of a previous study of ours (Sproull et al., in press) where subjects gave lower social evaluations but not lower intelligence evaluations to an Computer Face-Voice interviewer than to a Computer Text interviewer.

Other evaluations generally did not differ. As in our previous work, Warner-Sugerman potency ratings were not different across conditions nor were ratings of how similar subjects felt to the partner. Subjects in the Person and Computer Text condition said the partner resembled someone they knew, as compared with the partner in the Computer Face-Voice and Computer Voice conditions ($F[3, 82] = 5.8$, $p < .01$), but the contrasts were not significant.

Effects of other variables

There were no effects of gender, age, native language, or computer expertise on commitments, choices, or whether or not subjects kept their commitments. We

replicated a previous finding (Sproull et al., in press) that women gave less positive social evaluations to the Face-Voice computer than men did. In this study an interaction of condition X gender ($F [3, 8] = 2.9$, $p < .05$) reflected the case that women gave higher ratings to the human and computer text partner than men did, but lower ratings to the Face-Voice and voice computer partners.

Discussion

Our results suggest that people can respond socially to technology, that some of the processes influencing relations with a person also affect people's relations with a computer, and that people follow social rules in interacting with a computer. Cooperation with a computer partner in a dilemma was half that with a person, but cooperation after discussion with a computer partner was still substantial—higher than with uncommunicative computers and much higher when comparing the best liked computer partner (Computer Text condition = 50%) with computer partners of the past (e.g., < 35% in Abric and Kahan, 1972; 20-30% in Shure and Meeker, 1968).

The results lend support to both social identity and social contract explanations of cooperation following discussion, and suggest that identity or commitment processes might affect interactions with a computer. On trial 1, which is uncontaminated by the partner's choice, a majority of subjects across conditions suggested cooperation, but they were more likely to break their promises to cooperate when the partner was a computer, especially when the computer was more human-like. They also rated the human-like computer as least friendly, cheerful, attractive, and warm of all the partners. One interpretation of these findings is that subjects simply rejected the human-like computer as a partner, and then pursued a self-interested strategy. This interpretation is consistent

with research on stigma and novelty (e.g., Fiske, 1980; Langer et al., 1976). But subjects' ratings of their feelings of partnership with the human-like computer were only slightly lower than their feelings of partnership towards the confederate. We suggest another interpretation of the results of trial 1, which is more in keeping with the ratings of partnership, indicating minimal identity with the computer partners, especially with the human-like partner. Subjects might have formed a group identity of sorts with their computer partner, but disliked and took advantage of it, especially when it was the human-like, but imperfect. Hogg (1992) has argued that intragroup downgrading of marginal group members occurs when there is intergroup competition and a group member fails to meet expectations on important group dimensions (pp. 118-125). Marques (1990) calls this phenomenon the "black sheep" effect. Subjects having a human-like computer partner met these conditions in that other partnerships were competing with them for prizes, and the partner differed on relevant dimensions. The other computers differed more from the subjects in an absolute sense, but not on visual facial dimensions people use to evaluate others socially and to trust them (Sproull et al., in press). A better test of this explanation, of course, awaits future research.

Social contract arguments as to why people cooperate in dilemmas rest on the role of commitment during discussion. Our results show strong effects of commitments during discussion on cooperation. The majority of subjects on trial 1 made a commitment to cooperate during discussion and kept to this commitment, though they kept their promises less to a computer, especially to a human-like computer. Furthermore, across trials, cooperation was much higher when subjects had a chance to make a commitment to cooperate, and subjects who promised to cooperate were much more likely to cooperate than subjects who

did not, regardless of whether the partner was a computer or a person. Internal analyses of choices on each of the discussion trials (1, 2, 4, and 5) by subjects who made a commitment to cooperate or compete show highly significant effects of commitment on choice on each of these trials (trial 1 $F [1, 75] = 31.1$, $p < .001$; trial 2 $F [1, 77] = 16.0$, $p < .001$; trial 4 $F [1, 70] = 20.4$, $p < .001$; trial 5 $F [1, 75] = 13.6$, $p < .001$).

We derived two reasons from the literature as to why people keep their commitments—out of a sense of obligation to the partnership or group and out of a sense of obligation to the self. For those persons feeling obligated to the partnership, the nature of the partner might well affect the strength of that feeling. For example, would the partner appreciate and reciprocate honorable behavior? Would the partner be worthy of loyalty? For persons feeling obligated to themselves, the partner might make less of a difference—to some people a promise is a promise. We cannot glean from our data whether one or both of these processes was operating in this study, but our results indicate that subjects did not value commitment above all other considerations. Subjects who made commitments to a computer, especially to a human-like computer, seem to have interpreted their commitments differently than did subjects who made commitments to a person. This kind of differential commitment in a dilemma seems worthy of future exploration.

Limitations

Our results, and our interpretation of them in the context of social identity and social contract approaches, are constrained by the design, procedure, and technology we used and require more empirical and theoretical development. For instance, we gave social identity and social contract processes a boost by referring to the computer and to the subject as partners and by encouraging pre-game

conversation. Also, by failing to counter-balance discussion and no discussion trials, we cannot rule out the possibility that having discussion first might have increased its apparent effect on cooperation. Many reasons other than the ones we proposed might be advanced to explain why subjects would cooperate with a computer: to please or to avoid displeasing the experimenter; because it is a dominant response given while subjects are anxious; because cooperation makes the computer seem less frightening; because the computer seems guided by a real person, or because the experimenter treats the computer as though it were human.[7] We cannot positively rule out these explanations, but questionnaire ratings of the experimenter's behavior, of the subjects' feelings and emotions, and of subjects perceptions of the autonomy of the partner did not differ across conditions and do not support these conjectures. Moreover these explanations do not account for subjects' differential responses to the human-like computer as compared with the other computers.

The somewhat negative reactions to the talking face computer in this study might have resulted from idiosyncrasies in the particular computer interface we used. A more perfect representation of a face and voice in a computer display might increase rather than decrease cooperative behavior. A more perfect representation might increase the computer partner's credibility or remind people more of cooperation schemata. Recent research suggests that visual fidelity in computer representations increases involvement and learning (Christel, 1994) but the reasons for these effects are not well understood. A realistic synthetic partner for mixed motive paradigms awaits future technology improvements in both speech and person representation. Until then, experimenters have a variety of options: video displays, animation, and the better talking faces that continue to come out of computer science laboratories.

Our discussion of the imperfect talking face computer assumes that realism is a key variable in communication and, if so, that as it becomes possible to increase the realism of the representation of a person, cooperation should increase. Alternatively, consider a person interacting with a computer representation of Kermit the Frog or Minnie Mouse. Would a charming cartoon character that is clearly not human prompt more cooperation? A positive answer would suggest that likeability or resemblance to young children or pets rather than realism is the key element in eliciting communication that leads to social identity and social contract. If an attribute such as the face-to-body ratio is important, then ever more realistic computers might not increase cooperation. They would not necessarily be attractive and even minor deviations from reality would remind people of the fundamentally unreal nature of the machine. If we are to better understand the fundamental properties of social identification and social contract we need to understand what aspects of an artificial partner matter in producing social responses.

Conclusion

This research, though exploratory, represents a new perspective on what it means to be social. Human social behavior, it seems, does not depend on interaction with people or even categorization in a human group. Social psychological research dating from the 1930's on social facilitation and the development of strong ties and ingroup identification has emphasized the impact on social behavior of the presence of people and interaction among them. For example Lewin's studies of leadership and group discussion and Sherif's studies of a boys' camp showed how direct contact in a group and reciprocity encouraged cooperation. To explain cooperation with a computer, we need a theory of cooperation that can arise in the absence of people.

Theorists have long been interested in cooperation when people cannot communicate directly with their beneficiaries (Mauss 1967 [1925]). Macy (1990) has argued that group members can learn to cooperate or contribute to group welfare in the absence of direct communication with others if others' cooperative behavior is visible or reinforced. He argues that when people observe others contributing and see rewards to the public good they may develop a norm of cooperation. Once a critical mass of contributors is achieved (e.g., Marwell, Oliver, and Prahl, 1988), the legitimacy and value of cooperation can be self-reinforcing. In a related argument, Cialdini and his colleagues have argued that the visibility of others' cooperative behavior or cooperative outcomes reminds people that others cooperate in the situation, increasing the salience of a norm of cooperation (Cialdini, Reno, and Kallgren 1990). The residue of cooperative or antisocial behavior, such as litter on a street, can trigger normative responses. Generally, this work suggests that social behavior æpositive or negative æis triggered by situational cues, which need not be other people.

Our work suggests that commitments can be elicited fairly easily, even by a machine. If keeping their word is important to people's self-identity and feelings of self worth, and if they gain personal and social benefits from sticking to social contracts, then they might value even commitments to a computer. Cooperation to keep a commitment, in this view, is social insofar as honoring commitments is learned in society; the consequences of cooperation and defection are, of course, social, but social interaction with people or knowledge of specific human beneficiaries need not be presumed. This argument suggests that people keep commitments, whether to a person, group, dog, or computer, not because they believe each of these is human, but because they, themselves, are human.

Critics have long complained that research in artificial intelligence, which

exploits parallels between computer behavior and human behavior, is deeply demeaning to the essential humanness of people. Herbert Simon, in response to those critics, has said, "I think that some people's anxiety about the computer is that it will relieve us of another source of our uniqueness as the only kind of system in this universe that can think. To say that computers can do what we do when we think and use language and discover is to challenge another area of human uniqueness. . . .Why don't we try to get our sense of worth in this world, not in uniqueness, not in being apart from the rest of nature, but in being a general and integral part of nature" (Simon, 1987, pp. 14-15). In the spirit of that view, this paper suggests that by understanding how people interact with technology that has more or fewer human characteristics, we not only can understand more about how people interact with technology but also understand more about what makes us social.

References

Abric, J. C., & Kahan, J. (1972). The effects of representations and behavior in experimental games. European Journal of Social Psychology, 2, 129-144.

Allen, K. M., Blascovich, J., Tomaka, J. & Kelsey, R. M. (1991). Presence of human friends and pet dogs as moderators of autonomic responses to stress in women. Journal of Personality and Social Psychology, 61, 582-589.

Amabile, T. M. (1983). Brilliant but cruel: Perceptions of negative evaluators. Journal of Personality and Social Psychology, 19, 146-156.

Andreoni, J. and Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. The Economic Journal, 103, 570-585.

Apfelbaum, E. (1974). On conflicts and bargaining (pp. 103-156). In L. Berkowitz (Ed), Advances in experimental social psychology, vol 7. NY: Academic Press.

Ball, E., Ling, D. T., Pugh, D., Skelly, T., Stankosky, Thiel, D. (1994). ReActor: A system for real-time, reactive animations. Conference Companion: Demonstrations, CHI '94, Boston, Mass., April 24-28.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs;, NJ: Prentice-Hall.

Baumeister, R. F., Stillwell, A. M., and Heatherton, T. F. (1994). Guilt: An interpersonal approach. Psychological Bulletin, 115, 243-267.

Belk, R. W. (1988). Possessions and the extended self. Journal of Consumer Research, 15, 139-168.

Binick, Y. M., Westbury, C. F., and Servan-Schreiber, D. (1989). Case histories and shorter communications. Behavioral Research Therapy, 27, 303-306.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. Psychological Bulletin, 86, 307-324.

Brewer, M. B., and Kramer, R. M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. Journal of Personality and Social Psychology, 50, 543-549.

Burnstein, E. (1969). Interdependence in groups (pp. 307-405). In J. Mills (Ed.), Experimental social psychology, Toronto: MacMillan.

Caporael, L. R., Dawes, R. M., Orbell, J. M., and van de Kragt, A. J. C. (1989). Selfishness examined: Cooperation in the absence of egoistic incentives. Behavioral and Brain Sciences, 12, 683-699.

Chen, X. & Komorita, S. S. (1994). The effects of communication and commitment in a public goods social dilemma. Organizational Behavior and Human Decision Processes, 60, 367-386.

Christel, M. G. (1994). The role of visual fidelity in computer-based instruction. Human-Computer Interaction, 9, 183-223.

Cialdini, R. B.. (1984). Influence: The new psychology of modern persuasion. (pp. 75-106.). NY: Quill.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. Journal of Personality and Social Psychology, 58, 1015-1026.

Csikszentmihalyi, M. & Rochberg-Halton, E. (1981). The meaning of things:

Domestic symbols and the self. Cambridge: Cambridge University Press.

Dawes, R. M., McTavish, J., and Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. Journal of Personality and Social Psychology, 35, 1-11.

Dawes, R. M., Orbell, J., and van de Kragt, A. J. (1988). Not me or thee but we: The importance of group identity in eliciting cooperation in dilemma situations. Acta Psychologica, 68, 83-97.

Deane, P. D. (1993). On metaphoric inversion. Metaphor and Symbolic Activity, 8, 111-126.

Dennett, D. C. (1988). Precis of the intentional stance. Behavioral and Brain Science, 11, 495-546.

Deutsch, M. (1958). Trust and suspicion. Journal of Conflict Resolution, 2, 265-279.

Dittmar, H. (1992). The social psychology of material possessions: To have is to be. NY: St. Martin's Press.

Eichenwald, K. (1986). Hi, voter. This is your president. New York Times,. Section 3, 19.

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

Elliott, C. (1994). Research problems in the use of a shallow artificial intelligence model of personality and emotion (pp. 9-15), Proceedings of the Twelfth National Conference on Artificial Intelligence, AAAI-94, Seattle, Washington, July 31st -- August 4th, 1994.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. Journal of Personality and Social Psychology, 38, 889-906.

Folkes, V. S. and Sears, D. O. (1977). Does everybody like a liker? Journal of Experimental Social Psychology, 13, 505-519.

Friedman, B. (1995). It's the computer's fault: Reasoning about computers as moral agents (pp 226-227). In Human Factors in Computing Systems: CHI '95 Conference Companion (Proceedings document). May 7-10, Denver.

Friedmann, E., Katcher, A. H., Lynch, J. J., and Thomas, S. A. (1980). Animal companions and one-year survival of patients after discharge from a coronary care unit. Public Health Reports, 95, 307-312.

Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. Human-Computer Interaction, 2, 167-177.

Gozzi, Jr., R (1989). Metaphors that undermine human identity. Contents, 46, 49-53.

Hayes-Roth, B., Sincoff, E., Brownston, L., Huard, R., and Lent, B. (1995). Directed improvisation with animated puppets (pp. 79-80). In Human Factors in Computing Systems: CHI '95 Conference Companion (Proceedings document). May 7-10, Denver.

Hogg, M. A. (1992). The social psychology of group cohesiveness: From attraction to social identity. Hertfordshire : Harvester Wheatsheaf.

Hogg, M. A. and McGarty, C. (1990). Self-categorization and social identity (pp. 10-27). In D. Abrams & M. A. Hogg, Social identity theory: Constructive and critical advances. New York: Springer-Verlag.

Insko, C. A., Hoyle, R. H., Pinkley, R. L., Hong, G., Slim, R., Dalton, B., Lin, Y., Ruffin, P.P., Dardis, G. J., Bernthal, P. R., and Schopler, J. (1988). Individual-group discontinuity: The role of a consensus rule. Journal of Experimental Social Psychology, 24, 505-519.

Itou, K. S., Hayamizu, S. and Tanaka, H. (1992). Continuous speech recognition by context-dependent phonetic HMM and an efficient algorithm for finding N-best sentence hypotheses. Proceedings of ICASSP. IEEE Press.

Kelley, H. H., and Thibaut, J. W. (1978). Interpersonal relations: A theory of interdependence. New York: Wiley.

Kerr, N. L. and Kaufman-Gilliland, C. M. (1994). Communication, commitment, and cooperation in social dilemmas. Journal of Personality and Social Psychology, 66, 513-529.

Kiesler, C. A. (1971). The psychology of commitment: Experiments linking behavior to belief. New York: Academic Press.

Kleck, R. E. (1968). Psychical stigma and nonverbal cues emitted in face-to-face inter-action. Human Relations, 21, 19-28.

Kleck, R., Buck, P. L., Goller, W. L., London, R. S., Pfeiffer, J. R., and Vukecvic, D. P. (1968). The effect of stigmatizing conditions on the use of personal space. Psychological Reports, 23, 111-118.

Komorita, S. S., and Parks, C. D. (1995). Interpersonal relations: Mixed-motive interaction. Annual Review of Psychology, 46, 183-20-7.

Kramer, R. M., and Brewer, M. B. (1984). Effects of group identity on resource use in a simulated commons dilemma. Journal of Personality and Social Psychology, 46, 1044-1057.

Langer, E. J., Taylor, S. E., Fiske, S. and Chanowitz, B. (1976). Stigma, staring and discomfort: A novel-stimulus hypothesis. Journal of Experimental Social Psychology, 12, 451-463.

Laurel, B. (1990). Interface agents: Metaphors with character. In B. Laurel (Ed.), The art of human-computer interface design. New York: Addison-Wesley.

Macy, M. W. (1990). Learning theory and the logic of critical mass, American Sociological Review, 55, 809-826.

Marques, J. M. (1990). The black-sheep effect: Out-group homogeneity in social comparison settings. In D. Abrams and M. A. Hogg (eds), Social identity theory: Constructive and critical advances (pp. 131-151). Hemel Hempstead: Harvester Wheatsheaf.

Marwell, G. P., and Ames, R. E. (1981). Economists free ride. Does anyone else? Journal of Public Economics, 15, 295-310.

Marwell, G., P. Oliver and R. Prahl (1988). Social networks and collective actions: A theory of critical mass III," American Journal of Sociology, 94, 502-532.

Mauss, M. (1967, [1925]), The gift, New York: Norton.

Nagao, K., and Takeuchi, A. (1994, April). Social interaction: Multimodal conversation with social agents. Technical paper. SCSL-TR-94-005. Sony Computer Science Laboratory Inc., 3-14-13 Higashi-gotanda, Shinagawa-ku, Tokyo, 141, Japan.

Nass, C. I., and Steuer, J. (1993). Computers, voices, and sources of evaluation. Human Communication Research, 19(4), 504-527.

Nass, C. I., Steuer, J., Henriksen, L., and Dryer, D. C. (1994). Machines and social attributions: Performance assessments of computers subsequent to "self-" or "other-" evaluations. International Journal of Man-Machine Studies, 40, 543-559.

Nass, C. I., Steuer, J., and Tauber, E. R. (1994). Computers are social actors. Proceedings of the CHI'94 Conference of the ACM/SIGCHI, Boston, MA, April 1994.

Nemeth, C. (1972). A critical analysis of research utilizing the prisoner's dilemma paradigm for the study of bargaining (pp. 203-234). In L. Berkowitz (Ed.), Advances in experimental social psychology. NY: Academic Press.

New York Times (1994.) International Data Corporation statistic on personal computers at work quoted in S. Nasar, The American economy, Back on top, Feb 27, Section 3, p. 6.

Orbell, J. M., Dawes, R. M., and van de Kragt, A. J. C. (1988). Explaining discussion-induced cooperation. Journal of Personality and Social Psychology, 54, 811-819.

Orbell, J. M., van de Kragt, A. J., and Dawes, R. M. (1991). Covenants without the sword: The role of promises in social dilemma circumstances. In K. Koford and J. Miller (Eds.), Social norms and economic institutions. An Arbor, MI: University of Michigan Press.

Oskamp, S. (1971). Effects of programmed strategies on cooperation in the prisoner's dilemma and other mixed-motive games, Journal of Conflict Resolution, 15, 225-259.

Ostrom, E., Walker, J. M., and Gardner, R. (1992). Covenants with and without

a sword: Self-governance is possible. American Political Science Review, 86, 404-417.

Pruitt, D. G., and Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future, Annual Review of Psychology, 28, 363-392.

Reeves, B., Lombard, M., and Melwani, G. (1992). Faces on the screen: Pictures or natural experience? Unpublished manuscript, SRCT #103. Stanford University.

Resnick, P.V., and Lammers, H. B. (1985). The influence of self-esteem on cognitive responses to machine-like versus human-like computer feedback. The Journal of Social Psychology, 125, 761-769.

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. Rationality and Society, 7, 58-92.

Schlenker, B. R. (1985). Identity and self-identification. In B. R. Schlenker (Ed.), The self and social life. NY: McGraw-Hill.

Schneiderman, B. (1987). Designing the user interface: Strategies for effective human-computer interaction. Boston: Addison-Wesley.

Schopler, J., Insko, C. A., Graetz, K. A., Drigotas, S., Smith, V. A., Dahl, K. (1993). Individual-group discontinuity: Further evidence for mediation by fear and greed. Personality and Social Psychology Bulletin, 19, 419-431.

Searle, J. R. (1981). Minds, brains, and programs. In D. R. Hofstadter & D. C. Dennett (Eds.), The mind's I (pp. 353-372). Bantam: Toronto.

Shamir, B. (1991). Meaning, self and motivation in organizations, Organization

Science, 12, 405-424.

Shure, G., and Meeker, R. J. (1968). Empirical demonstration of normative behavior in the Prisoner's Dilemma. Proceedings of the 76th Annual Convention, APA. 3, 61-62.

Simon, H. (1987). Computers and society. In S. Kiesler and L. Sproull (eds.), Computing and change on campus, pp. 4-15. Cambridge: Cambridge University Press.

Sproull, L., and Kiesler, S. (1986). Reducing social context cues: Electronic mail in organizational communication. Management Science, 32(11), 1492-1512.

Sproull, L., Walker, J., Subramani, R., Kiesler, S., and Waters, K. (in press). When the interface is a face. Human-Computer Interaction.

Tajfel, H. (1959). Quantitative judgment in social perception. British Journal of Psychology, 50, 16-29.

Tajfel, H., and Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), Psychology of intergroup relations. (pp. 7-24). Chicago: Nelson-Hall.

Takemura, H., and Kishino, F. (1992). Cooperative work environment using virtual workspace. Proceedings of Computer Supported Cooperative Work Conference. New York: ACM Press.

Tedeschi, J. T. (1981). Impression management theory and social psychological research. NY: Academic Press.

Toothaker, L. E. (1993) Multiple comparison procedures. Newbury Park, CA: Sage.

Turkle, S. (1984). The second self: Computers and the human spirit. New York: Simon and Schuster.

Turner, J. C. (1982). Towards a cognitive redefinition of the social group. In H. Tajfel (Ed.), Social identity and intergroup relations. Cambridge: Cambridge University Press.

Voissem, N. H. and Sistrunk, F. (1971). Communication schedule and cooperative game behavior. Journal of Personality and Social Psychology, 19, 160-167.

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

Waters, K. (1987). A muscle model for animating three-dimensional facial expressions. Computer Graphics, 21, 17-24.

Waters, K. and Levergood, T. M. (1994). DECface: An automatic lip-synchronization algorithm for synthetic faces. Proceedings of ACM Multimedia 94. San Francisco, Oct. 10.

Waters, K. and Levergood, T. M. (1995). DECface: A system for synthetic face applications. Multimedia Tools and Applications, 1, 1-16.

Notes

[1]For instance, larger video screens and better resolution increase people's sense of involvement with the story and people shown in the video, but under certain conditions, reduce accuracy (Nass, personal communication). Also, see Gozzi (1989).

[2]Social identity theorists initially proposed that people identify with valued groups to enhance their own self-esteem (e.g., Turner 1982), but there are many other reasons for social identity such as persistent labelling by others and gaining access to valued resources. A more cognitive, self-categorization perspective is that identification with a group can proceed from numerous contextual conditions that cause individuals to see themselves as similar to group members, or as sharing a common fate (Hogg and McGarty, 1990).

[3]As would be expected in the absence of visibility and face-to-face communication, cooperation rates are lower in these experiments than they are when subjects interact in person (Sally, 1995).

[4]Because both experimenter and confederate were older than the subjects, their scripts were made consistent with their maturity and having finished college.

[5]We thank a reviewer for this suggestion.

[6]Some ratings were presented to subjects in a nonintuitive direction (more positive = lower score). We have reversed these scores so that all ratings are reported with a higher score equivalent to a more positive valence.

[7]This list was suggested by an anonymous reviewer.