Affective Feedback from Computers and its Effect on Perceived Ability and Affect: A Test of the Computers as Social Actor Hypothesis

PUNYA MISHRA
Michigan State University
USA
punya@msu.edu

We report an experimental study that examined two questions: (a) The effect of affective feedback from computers on participants' motivation and self-perception of ability; and (b) whether people respond similarly to computer feedback as they do to human feedback. This study, framed within the Computers As Social Actors (CASA) framework, essentially replicated a prior study on human-human interaction (Meyer, Mittag, & Engler, 1986) except that human evaluators were replaced with computer evaluators. The Meyer et al. study showed that there was a paradoxical relationship between praise and blame feedback and students perception of ability and motivation to engage in a task. Results of our study indicate that, consistent with the CASA hypothesis, people do respond to praise and blame feedback when provided by a computer. However, there are important differences between the results of our study and the Meyer et al. study. The participants in our study took the feedback from the computer at "face value" and seemed unwilling to commit to the same level of "deep psychological processing" about intentionality as they appeared to do with human respondents. We believe that this research combining existing theory and research on motivation and human computer interaction offers significant implications for the design of educational technology and also points to directions for future research. Praise, like penicillin, must not be administered haphazardly. There are rules and cautions that govern the handling of potent medicines—rules

> about timing and dosage, cautions about possible allergic reac-
> tions. There are similar regulations about the administration of
> emotional medicine (Ginott, 1965, p. 39).

Does praise by a teacher always have a positive impact on student achievement and motivation? Correspondingly, does criticism always have a negative effect on students' feelings of self-efficacy? Educational researchers have built an impressive body of empirical and theoretical knowledge in answering questions such as these (Graham, 1991; Stipek, 1993, 1996; Weary, Stanley, & Harvey, 1989, see Henderlong & Lepper, 2002 for a good review of research). Research exploring achievement motivation has generated a great deal of practical knowledge about how instructional practices affect student motivation. Though the findings of attribution research are of no surprise to researchers and practicing educators, they have not received a similar level of attention from designers and scholars of educational technology (Pridemore & Klein, 1991; Schurick, Williges, & Maynard, 1985).

COMPUTERS AS SOCIAL ACTORS

There are many reasons why designers of educational technology systems have not paid much attention to the nature and effects of affective feedback on students motivation and affect. The design of feedback of educational technology systems is often based on the simplistic (and as we indicated earlier, erroneous) framework where praise is assumed to affect behavior in a specific way when contingent upon performance. This naïve belief that praise strengthens the probability of a particular behavior has been called the "perspective of reinforcement" (Henderlong & Lepper, 2002). In pragmatic terms this means that criticism is rarely if ever considered as being of any value in the design of educational technology systems. Designers are often worried that blame or criticism may have a negative effect on students' self-perception of ability or motivation. Finally, from a pragmatic point of view, offering praise (in the form of audio, textual, or other forms of feedback) is easily designed into the system.

Part of the reason for this may also be that most research in the area of educational technology has emphasized the cognitive and information processing aspects of learning where computers are viewed as being neutral cognitive tools that sidestep issues of attitudes and stereotypes typical of human interactions (Lajoie & Derry, 1993). Moreover, certain researchers believe that users or learners could find computer responses such as praise,

criticism or helping behavior as being implausible and unacceptable (Lepper, Woolverton, Mumme, & Gurtner, 1993).

However, over the past few years there has been some intriguing research in the area of human computer interaction (HCI) that undermine the "computers as neutral cognitive actors" perspective. This research most often described as functioning within the Computers as Social Actors (CASA) paradigm, shows that people may be unconsciously perceiving interactive media as being "intentional social agents" and reading personality, beliefs, and attitudes into them, and more importantly, may be acting on these beliefs. There is a growing body of empirical evidence to support this position: People are polite to machines (Nass, Moon, & Carney, 1999), read gender and personalities into machines (Alvarez-Torres, Mishra, & Zhao, 2001; Nass, Moon, & Green, 1997), are flattered by machines (Fogg & Nass, 1997), treat machines as teammates (Nass, Fogg, & Moon, 1996), and get angry and punish them (Ferdig, & Mishra, 2004). Responding socially to a computer (such as talking to it) has often been viewed as being atypical or abnormal. It was believed that only people who lack knowledge about how computers function (such as children or novices) would engage in these behaviors. However, it appears that such social responses towards interactive technologies are quite common. These responses occur even when users explicitly deny believing that computers have feelings or intentionality (Reeves & Nass, 1996). This "intentional stance" (Dennet, 1987) appears to be unconscious, instinctual, and independent of age, experience, and expertise (Turkle, 1984; Reeves & Nass, 1996; Weizenbaum, 1976). The automatic nature of this response indicates that this response may not be available to conscious introspection (Mishra, Wojcikiewicz, & Nicholson, 2002). As Reeves and Nass (1996, p. 12) said, these responses are "easy to generate, commonplace, and incurable." Moreover, this social response towards media is also easy to manipulate and is triggered not just by interacting with some fancy graphic voice-driven interface but even by interacting with the simplest of text-based command-line interfaces (chatterbots which attempt to mimic human conversation are a good examples of these). Nass, Moon, and Carney (1999) provided evidence for the fact that the intentional stance is triggered equally strongly by text-based or voice-based interfaces.

WHY IS THE COMPUTER TREATED AS A SOCIAL ACTOR?

Drawing on evidence from cognitive science, developmental psychology, and evolutionary psychology, we have argued (Mishra, Nicholson, & Wojciekiewicz, 2001/2004) that this intentional stance (Dennett, 1987) is a

"cognitive illusion," an artifact of the way our minds were designed by natural selection (Barkow, Cosmides, & Tooby, 1992). Our experience of other minds is a form of "naïve psychology" (Gopnik & Meltzoff, 1997; Wellman, 1990) that is the product of highly sophisticated and deeply entrenched inferential principles that operate at a level of our brain that is quite inaccessible to conscious introspection or voluntary control—which is why people are surprised by the results of the CASA research (Pinker, 1997; Baron-Cohen, 1997). This is a form of what Hermann von Helmholtz, called "perception as unconscious inference" (Shepard, 1990). Our ability to create these interactive artifacts, which emerged only recently on an evolutionary time scale, enables us to present stimuli to our minds that could only have been from other "intentional actors," such as animals or other humans (Mishra et al.).

Thus, because computers use natural language, interact in real time, and fill traditionally social roles, even experienced computer users tend to respond to them as social entities (Reeves & Nass, 1996). Computers today are used to control other machines; search and find patterns in large datasets (Kaufmann & Smarr, 1993); to record, understand, and speak in human voice (Cawsey, 1989; Maes & Kozierok, 1993; Schmandt, 1994); to recognize handwritten scripts (Wang, 1991); to interact with users based on complex contingencies (Johnson, Rickel, Lester, 2000); and to learn from experience (Selker, 1994). Moreover, computers have also begun to fill many social roles that have traditionally been filled by people such as bank tellers (ATM machines), receptionists (electronic voice mail systems), teachers/tutors (computer assisted instruction), interviewers (automatic interview systems), and opponents and/or partners in games (video games).

One of the key research topics in the area of HCI has been to instill into computers, human-like qualities in terms of both intelligent functionalities and communicative capabilities. Researchers in this area believe that tapping into the social aspects of computing provides the potential for designing more natural interaction. (Maes, 1997; Laurel, 1997) Research on anthropomorphic and believable agents (Bates, 1994; Isbister & Nass, 2000) has received a great deal of attention in the recent past, and a number of technologies in artificial intelligence, affective computing (Picard, 1997), natural-language processing (Cassel & Thorisson, 1999), and multi-modal interfaces have been devoted to create personalities in computer software (Hershey, Mishra, & Altermatt, 2005).

Computers have also become easier to work with. Computer users today need less technical training and can rely more on their natural language abilities and culturally determined scripts, such as interacting with a bank

teller (Gaines, 1981; Ferrari, 1986; Cawsey, 1989; Nielsen, 1990). Other scholars have argued that such social responses to technology are due to the over-application of social scripts that we have been taught and learned over time (Nass & Moon, 2000).

Thus the findings of the CASA approach are part of a larger set of findings that indicate that people often approach the world mindlessly (Gilbert, 1991; Langer 1989, 1992) and prematurely commit to overly simplistic scripts without necessarily basing their responses on all the relevant features of the situation. Computers due to a wide variety of reasons, some of which have been outlined previously, offer situations where social responses appear "natural." If the mindlessness argument is right (see Nass & Moon, 2000 for a good review) then one could argue that these social responses would *actually increase* as computers and interactive media enter all aspects of our lives.

AFFECTIVE FEEDBACK: A BRIEF REVIEW OF THE RESEARCH

So what does the research on human-human interaction say about affective feedback? There have been numerous studies, both experimental and classroom-based, that provide strong evidence that teachers' instructional decisions play a very significant role in student motivation (Stipek, 1996). Research has shown that teacher expectations can unintentionally lead to negative or positive student attributions about their ability and motivation. Research also shows that attributional statements—statements about the cause of performance outcomes—made by teachers play a very significant role in student motivation and achievement-related beliefs (Weiner, 1986; Weiner, Graham, Stern, & Lawson, 1982; Graham, 1990, 1991). This research indicates that sympathizing, praising, criticizing, and offering help has implications for students' perceptions of their own competence.

This research also highlights the importance of the context within which praise, criticism, and help are offered. Teachers' emotional reactions to success and failure also have been shown to affect children's causal attributions and expectations of success. This is due to the fact that people believe emotional reactions reflect a person's perception of the cause of the behavior (Weiner, 1986). Children understand that anger is aroused when another's failure is attributed to controllable factors, such as lack of effort, but it is much later that they understand that pity is aroused when another's failure is perceived to be caused by uncontrollable causes (Weiner et al., 1982; Graham, 1990, 1991). Teachers' emotional responses can also affect stu-

dents' self-perceptions. Graham (1984) showed that children who received sympathy from the experimenter after failure at a task, tended to attribute their failure to a lack of ability, while those who received mild anger tended to attribute their failure to lack of effort. Moreover, children who received pity had lower expectations for success in the future compared to those who received the angry response. This finding shows that even well-intentioned teacher behaviors can have a negative impact on students' beliefs related to achievement.

One consistent finding of the research on teacher feedback has been that praise is not universally good, neither is blame automatically bad, for the student. Research has undermined the naïve idea that praise is always beneficial to the learner and blame necessarily harmful (Henderlong & Lepper, 2002). This is not to argue that praise cannot have positive effects, but rather that these positive effects are dependent on the *context* within which this feedback is offered. In particular, researchers who study the effect of praise or criticism on students' achievement related beliefs have found that praise and blame may have effects on perceived ability that appear paradoxical when viewed from a framework that assumes a direct correlation between positive feedback and positive emotional, behavioral, and motivational consequences. For instance, depending on the context (such as the nature of the task), praise can even have negative effects on students' self-confidence, while criticism can actually have positive effects on students' self-confidence (Meyer, 1982; Parsons, Kaczala, & Meece, 1982). For instance, praise given for success on an easy task can have a negative effect, that is, it lowers, rather than raises, self-confidence (Meyer, 1982). Similarly, criticism of a poor performance can, under certain circumstances, be interpreted as an indication of the teachers' high perception of students' ability. This is because praise and criticism are assumed to be associated with the level of perceived effort. Additionally, people perceive an inverse relationship between effort and ability (Nicholls & Miller, 1984). Thus, if somebody is criticized for failing, this failure can be interpreted as being due to lack of effort (rather than lack of ability). And similarly, being praised for success on an easy task can be interpreted as an indication of low ability (else why would the person be praised!).

We must emphasize that we are not arguing that praise cannot have positive effects and that criticism can often have negative consequences. However, praise (and criticism) can be interpreted in may different ways and these interpretations can influence how the recipient responds to the feedback (Brophy, 1981; Kanouse, Gumpert, & Canavan-Gumpert, 1981; Meyer et al.,1979; Meyer, 1982). As Meyer (1992) said, "The effects of praise and

criticism are therefore not straightforward and invariant, but are mediated by the recipient's processing of these events and thus can be manifold. The analysis of praise and criticism solely from the perspective of reinforcement is far too simplistic" (p. 260).

BRINGING ATTRIBUTION THEORY TO HCI

The findings of the CASA approach have the potential to reconfigure the domain of educational technology by forcing us to reevaluate the conception of the computer and other forms of media as being mere tools. Failure to recognize the existence of people's social responses towards media can thwart the pedagogical goals of designers of educational software. Research in educational technology can focus on harnessing this natural reaction to interactive media to our advantage. These findings emphasize the importance of the social relationship that can develop between a computer and the learner. This relationship requires the thoughtful design such that the software (or agent) is perceived as being trustworthy and competent (Maes, 1997), empathetic (Lepper & Chabay, 1985), responsive (Laurel, 1997), demonstrating emotion (Bates, 1994), honest, and cooperative while providing feedback. A fundamental grounding in the conceptual underpinnings of people's social responses to interactive media, informed by the findings of cognitive and social psychology, may well provide the potential for the design of better learning environments.

However, given the complexity inherent in educational situations we must be careful not to blindly apply the findings of the CASA research. It is important that our research be sensitive to the larger educational context. For instance, research shows that indiscriminate flattery gives users a better feeling towards the computer program (Fogg & Nass, 1997). Flattering users, irrespective of context, may make sense for commercial software, but its application to educational technology is problematic. It may give learners a false sense of accomplishment, which may be more harmful than beneficial in the long run.

If the CASA approach is right, it is plausible that people could make attributions about their abilities and self-efficacy based on feedback from computers much as they make such attributions in response to feedback from people. In other words, if people react to media just as they respond to people, the results of the human-computer interaction would be the same as that found in the case of human-human interaction. This has important implications for the design of feedback from computer-based learning systems.

The study reported here attempts to combine these two research streams (educational psychology and HCI) by investigating the effect of affective feedback on the motivation and affect of students as they work with a computer program.

A note on the methodology: Previous research on the psychological responses to interactive media has followed a straightforward methodology (Reeves & Nass, 1996). Researchers select an existing social science research finding on human-human interaction and replicate it after replacing one of the human respondents with a computer. If people react to media just as they respond to people, the results of the human-computer interaction would be the same as that found in the case of human-human interaction.

For instance, Alvarez-Torres, Mishra & Zhao, (2001) replicated a finding in the second language acquisition (SLA) research. Research shows that a native speaker is regarded as being more credible and intelligent than a non-native speaker (Delamere, 1996; Raisler, 1976). This study examined whether people made similar attributions to computer software by replicating the previous studies after replacing the instructor by a native or non-native computer tutor. The native computer read the instructions for using the tutorial in an American accent, while a non-native computer read the instructions in a fluent and flawless Hispanic accent. The information that the participants had to learn was identical for both conditions: it was just text on screen. Scores on a test of recall indicate that the participants who worked with the native computer recalled significantly more information than those working with the non-native computer—a finding similar to that found in the human-human interaction literature. Similarly the current study replicates an existing study on human-human interaction by replacing one of the respondents by a computer.

We base our experimental design on the design of an earlier experiment (Meyer, Mittag, & Engler, 1986) conducted entirely with human participants. We essentially replicated the Meyer et al. study by replacing the human evaluators with a computer evaluator. We begin with a description of the original Meyer et al. (1986) study.

The Meyer, Mittag, and Engler (1986) Experiment

Meyer, Mittag, and Engler (1986) in an experimental study investigated the effects of praise and blame on evaluations of one's own performance and on affect. Of the four participants in each experimental session, two acted as evaluators and two acted as students. The two evaluators were actually ex-

perimental confederates. First, a test of logical ability was administered to the students by the experimenter. After the test was completed, both evaluators allegedly scored the tests in one condition (the test-scored condition) to induce the students to believe that the evaluators knew their ability. In another condition (the test-not-scored condition) the evaluators did not score the tests. In both conditions, the students then worked at two tasks, one easy and one difficult, and both were told that they had succeeded on the first easy task (i.e., received success feedback) and failed on the second more difficult task (i.e., received failure feedback). What differed was the affective feedback they received on their performance (success on easy task and failure on difficult task). While one of the two participants was praised by the evaluators for success on the easy task and not blamed for failure on the difficult task, the other participant received no praise for success on the easy task but was blamed for failure on the difficult task. At the end of the study the participants were asked to complete a set of Lickert scale items that measured feelings and perceptions of their performance on the tasks.

Results showed that when the participants could assume that the evaluators knew their ability (test-scored condition), receiving no praise for success and blame for failure, it led the participants to a more positive evaluation of their own test performance than receiving praise for success and no blame for failure. Within the test-scored condition, receiving no praise for success/no blame for failure also led to more positive affect than receiving praise for success/no blame for failure. Within the test-not-scored condition, the performance and affect ratings between students receiving no praise for success/blame for failure and those receiving praise for success/no blame for failure did not differ. The results of this study indicate that the effects of praise and blame are context dependent. In brief, this study showed that in certain contexts praise may actually be counterproductive (such as when offered for success on an easy task) and in other contexts ascribing blame may actually increase motivation (such as when offered for failure at a difficult task). When the evaluators had no knowledge of the participant's ability (non-test-scored condition), the feedback had virtually no effect on any of the performance evaluations or affective reactions.

The Current Study

In the current study we essentially replicated the earlier study replacing the "human evaluators" by "computer evaluators." The methodology, structure of tasks, survey instruments, nature of feedback, was identical in almost

all respects to the Meyer et al. (1986) study. (A few minor changes had to be made to the experimental protocol to compensate for the fact that the "human evaluators" had to be replaced by "computer evaluators.") Our study investigated two key questions:

(a) What is the effect of affective feedback (praise/blame) on individual motivation and affect as they work on a computer-based testing and evaluation system? In other words, this is similar to the questions answered by the Meyer et al. experiment, except that feedback this time is provided by computers rather than by humans.

(b) What are the similarities and differences between the Human-Human interaction case and the HCI case, that is, do people respond the same way to computer feedback as they do to feedback from humans? In other words, we seek to compare the findings of our research study to those of the findings of the Meyer et al. study, a test of the CASA hypothesis.

Participants: A total of 114 students enrolled in an introductory communication course at a large Midwestern university served as participants. While participation in this particular study was voluntary, the students did receive points that were to be applied toward their research participation grade, a required component of the course.

Procedures: The laboratory was furnished with two laptop computers, which rested on a table, facing one another, in such a way that the participants could not see one another's screens. Off to one side, easily visible to both participants, sat the Evaluation Computer (EC), a desktop computer complete with a large, color monitor visible to both participants.

The participants were then assigned randomly to one of four experimental conditions: Scored Test versus Non-Scored Test, crossed with praise for success on easy task, and no blame for failure on difficult task; versus no praise for success on easy task, and blame for failure on difficult task. After being instructed on the use of the laptop mouse and a handful of obligatory keyboard keys, the participants were told that they would be working with a new computer program for testing students. "For instance," they were informed, "if you were to take a test on the computer, the software could determine if your answers were right or wrong and would give you feedback on your responses."

The participants were then asked to complete a 20-item logical ability test. Those in the Scored Test condition were told that their response would

be uploaded to an online database where it would be scored for correctness and then their performance would be compared to the responses of thousands of previous test takers. Participants in the Non-Scored Test condition were told simply that their responses would be saved onto their laptop. Each item in the test consisted of a row of seven numbers that were associated according to a certain rule (e.g., 1, 2, 4, 7, 11, 16, 22, ___). The subject had to detect the rule and find the next number (29). The test consisted of 20 such tasks, of which a few were unsolvable. Thus, a certain amount of ambiguity, concerning a participant's own performance, was created by preventing each individual from eventually completing or correctly solving all the problems. The participants were allotted five minutes to work on the test.

After time had expired, on-screen graphics appeared, making it appear to the participants in the Scored Test conditions as though their responses were being transmitted to, and scored and evaluated by an online assessment program. Conversely, participants in the Non-Scored Test condition were lead to believe that the responses were being saved to their laptop. In this (non-scored condition) the possibility of evaluation was neither stated nor implied (similar to the case in the previous Meyer et al. [1986] study).

The subsequent procedure was identical for both experimental conditions. The experimenter explained that both students would now receive some tasks to work on. The participants were told that these tasks were similar to the tasks of the test. The participants were also told that the EC would react to their performance (once their data had been sent to the EC over the network) and provide feedback to them. The participants were then given two additional logic problems, one at a time, and were given 45 seconds to solve each problem. The first question was identified by the researcher as "easy," and defined aloud as "a question that at least 80% of previous test takers answered correctly." Following the first time interval, another set of on-screen graphics appeared, making it appear to the participants in all conditions as though their responses were being transmitted to, and scored and evaluated by the assessment program (running on the EC).

The EC then offered feedback to each of the participants. One of the participants was offered praise on success on the easy task ("Answer correct. Very impressive") while the other participant received neutral feedback ("Answer correct. Task completed"). Once the participants had read the feedback they were offered the last problem.

The last problem was described as being difficult and defined as "a question that at least 80% of previous test takers answered incorrectly." Actually the problem was unsolvable. Once the participants completed the problem, the data was "sent" to the evaluation computer. Both participants

were told that they had gotten the problem wrong. However, the participant who has received praise on getting the easy question right was now provided neutral feedback ("Wrong answer. Task completed"). In contrast, the participant who had received the neutral feedback on solving the easy problem correctly was now offered critical feedback ("Wrong answer. Should have done better"). Once the participants had read the feedback they were administered the posttest measure.

It is worthwhile to clarify that at no time were any of the participants' responses actually transmitted, scored, evaluated, or saved. On the contrary, the feedback, provided by the evaluation computer, was programmed in advance.

In essence, in each of the conditions (scored or not-scored), one of the participants was praised for success on an easy task and neither praised nor blamed for failure on the difficult task. Furthermore, the computer indicated to these participants that "it" (the computer) was very impressed with their ability to solve an easy problem and neutral to the participant's inability to solve the difficult problem. In contrast, the second participant was not praised for success on the easy task, but was blamed for failure on the difficult task. Additionally, the computer also informed the participant that they should have done a better job answering the difficult question, indicating that the second participant was capable of solving even the most difficult of problems.

Whether or not participants believed the feedback from the evaluation computer, was manipulated by whether or not the participants were in the Scored Test of Non-Scored Test condition. Participants in the Scored Test conditions were told that the computer's assessment of them, be it positive or negative, was based on a comparison of their responses with a large database of responses of prior test takers. In short, the computer had much evidence upon which to base its feedback. In contrast, those in the Non-Scored Test conditions were not told that the feedback was based on any prior knowledge (apart from the solution they had provided of the two final problems).

In summary, the experimental design was a 2 x 2 factorial with the following indpendent variables (a) evaluation of participant's ability (from the participants' perspective), manipulated by the alleged scoring (vs. not scoring) of the tests; and (b) type of feedback, (praise for success at easy task/ no blame for failure on difficult task vs. no praise for success on easy task/ blame for failure on difficult task).

DEPENDENT MEASURES

The questionnaire distributed following the feedback assessed the following variables. For the most part this was identical to the questionnaire used in the Meyer et al. (1986) study. Each of the ratings were made on a 7-point scale (exact descriptions given below):

Comparative evaluation of own test performance: "In your opinion, how well did you perform on the initial test as compared to the other student?" These ratings ranged from 1 ("much worse") to 7 ("much better").

Evaluation of own test performance: "How would you evaluate your performance on the initial test?" This evaluation ranged from 1 ("much below average") to 7 ("much above average").

Evaluation of other's test performance: "How would you evaluate the performance of the other pupil on the initial test?" This evaluation ranged from 1 ("much below average") to 7 ("much above average").

Affects: "What feelings did the reactions of the evaluation computer release in you?" A list of six affects (anger, joy, dejection, surprise, disappointment, and confidence) was presented, with separate scales for each affect. Each scale ranged from 1 ("not at all") to 7 ("very strong").

Interest: "How interesting do you find the tasks on which you worked?" This rating ranged from 1 ("completely uninteresting") to 7 ("very interesting").

Fairness: "In your opinion, how fair was the evaluation computer's assessment of your performance?" This evaluation ranged from 1 ("extremely unfair") to 10 ("extremely fair").

The only change in variables from the Meyer et al. (1986) study was replacing an item probing participant's perception of how much the human evaluator "liked" them. In this study, where the human evaluators were replaced by computer evaluators, a question of this sort was considered as being too leading, possibly biasing the participant towards considering the computer as being an affective agent. Since this was one of the research questions being investigated we chose to replace this item with the question on "fairness," a variable that was relatively neutral in its connotation.

ASSUMPTIONS AND PREDICTIONS

We can make two distinct predictions about the results of the replication of the Myer et al. (1986) study in the HCI situation.

1. *The common sense prediction*: The participants in the study would not consider the praise or blame feedback from the computer to be of any value. They would not believe or trust such affective comments when made by a software program running on a desktop PC. If this were true, we would see no effect of the praise or neutral feedback from the computer that is, there would be no difference between the means for the different conditions. Individuals in the different groups would discount all feedback from the computers, considering them to be programmed responses and not based on any "cognition" by the machines. If this were the case, we would have to reject the CASA hypothesis—at least in this context.
2. *The CASA prediction*: If on the other hand, the CASA paradigm is correct (and there is considerable support for this in the HCI research literature) we would find exactly the same results in the HCI case as was found in the HHI case. In other words, the results of this study would match those of Meyer et al. (1986) study in which the participants differentially responded to the feedback depending on the experimental manipulations (scored versus not-scored).

We must add that there is a possible third alternative to the two predictions made. It is possible that the CASA paradigm is weakly supported, in which case the participants would accept some of the feedback from the computer but not all of it. We currently do not have any specific theoretical or experimental evidence to indicate which would be the case.

RESULTS

Each dependent variable was subjected to a between-participants 2 x 2 (feedback x test scoring) analysis of variance.

Evaluation of Test Performance

A participant's performance on the test was assessed by three different items. Concerning a participant's evaluation of his own performance, rela-

tive to the other student, the analysis of variance revealed a significant main effect for feedback, $F(1, 110) = 9.47$, $p=.003$. The main effect for scoring was not significant ($F(1, 110) = .002$, $p=.962$); neither was the interaction of feedback x scoring significant $F(1, 110) = .083$, $p=.773$. Participants who received praise for success and no blame for failure (irrespective of whether they believed their ability level had been measured) estimated their performance to be higher, relative to the students who received blame for failure and no praise for success.

Each participant also evaluated their own test performance. The structure of results concerning the ratings of one's own performance is very similar to that of the participant's own performance ratings relative to the other students. The analysis of variance again revealed a significant main effect for feedback, $F(1, 110) = 3.901$, $p=.05$. The main effect for scoring was not significant ($F(1, 110) =1.468$, $p=.228$) and no interaction effect, $F(1, 110)=.271$, $p=.603$). In other words, the participants who were praised for success on an easy task (and not blamed for failure on a difficult task) believed their performance to be better than those who had received no praise for success on the easy task (and blamed for failure on the difficult task).

A somewhat different picture emerges when we look at participant's ratings of test performance of the other student in each session. Analysis of variance revealed a significant main effect for test scoring, $F(1, 110) = 7.767$, $p=.006$. But there was no significant effect for the main effect for feedback ($F(1, 110)=.089$, $p=.766$) and no interaction effect, $F(1, 110)=1.351$, $p=.248$. This means that, when participants believed that ability level had been taken into account, they rated the other participant's performance as being worse than when ability level was not considered. In other words, when ability was known, the other participant was considered as being of lower ability than when ability was not known.

Affective Reaction

Of the six affects assessed, anger, dejection, and disappointment were considered to be negative emotions, while joy and confidence were considered to be positive emotions. Surprise, as reaction to unexpected events, was considered neither positive nor negative. Because there were substantial correlations between the two positive emotions (0.68) and the three negative emotions (0.47, 0.57, 0.64), the respective ratings were summed to form a score for positive affect and negative affect. With respect to surprise, an analysis of variance showed no main effects and no interaction, $F's < 1$.

The means for positive and for negative affect are shown in Table 1. With respect to positive affect, analysis of variance revealed a significant main effect for feedback, $F(1, 110) = 10.061$, $p=.002$. The main effect for scoring was not significant ($F(1, 110) = .013$, $p=.911$); neither was the interaction of feedback x scoring significant $F(1, 110) = .022$, $p=.881$.

Table 1
Means and Standard Deviations of Dependent Variables

		Experimental condition			
		Test scored		Test not-scored	
		M	SD	M	SD
Evaluation of one's own performance	P/nB[a]	3.89	1.25	4.28	0.99
	nP/B[a]	3.57	1.06	3.72	1.36
Evaluation of the other participant's performance	P/nB	4.21	1.13	4.48	0.73
	nP/B	4.07	0.76	4.72	0.84
Evaluation of own performance relative to other participant's performance	P/nB	4.57	1.34	4.62	1.01
	nP/B	4.00	0.98	3.93	0.99
Positive affects	P/nB	3.62	1.49	3.55	1.80
	nP/B	2.69	1.31	2.70	1.28
Negative affects	P/nB	2.08	1.04	1.77	0.97
	nP/B	2.98	1.38	2.80	1.37
Surprise	P/nB	3.71	1.80	3.52	2.02
	nP/B	3.54	1.75	3.62	1.89
Fairness	P/nB	5.07	1.41	5.90	1.11
	nP/B	4.71	1.60	4.59	1.57

[a] P/nB denotes the experimental condition in which participants were praised for success on an easy task and not blamed for failure on a difficult task. nP/B denotes the condition in which participants received no praise for success on an easy task but blame for failure on a difficult task.

Similarly with respect to negative affect, analysis of variance revealed a significant main effect for feedback, $F(1, 110) = 18.372$, $p=.000$. The main effect for scoring was not significant ($F(1, 110) = 1.205$, $p=.275$); neither was the interaction of feedback x scoring significant $F(1, 110) = .082$, $p=.775$.

In the next step, we analyzed the difference between the scores of positive and the scores of negative affect. These scores were uncorrelated (.019). An analysis of variance of these data revealed a significant main effect for feedback, $F(1, 110) = 31.00$, $p=.000$. The main effect for scoring was not significant ($F(1, 110) = .423$, $p=.517$); neither was the interaction of feedback x scoring significant $F(1, 110) = .005$, $p=.945$.

Interest and Fairness

An analysis of variance on the interest ratings showed no main effects and no interaction, $F's < 1$. However, a pattern similar to what we have already found was also seen in the analysis of the participant's perception of fairness of the evaluation computer's assessment of their performance. Analysis of variance revealed a significant main effect for feedback, $F(1, 110) = 9.585$, $p=.002$. The main effect for scoring was not significant ($F(1, 110) = 1.675$, $p=.198$); neither was the interaction of feedback x scoring significant $F(1, 110) = 3.132$, $p=.080$.

RESULTS

We began with two research questions. The first was regarding the effect of affective feedback on individual motivation and affect while the second was regarding the differences and similarities between the two experiments (once when feedback was provided by humans and the other when feedback was provided by the computer). We consider each of these questions in turn:

Question 1: What is the effect of affective feedback (praise/blame) on individual motivation and affect as they work on a computer-based testing and evaluation system?

The analysis of the data from the current study indicates that affective feedback made a significant difference to the participants motivation and affect. In general, the participants who received positive feedback for success (even on an easy task) and neutral feedback on failure (on a difficult task) rated their performance as being better than those who received neutral feedback for success on the easy task and blame feedback for failure on the difficult task. Overall, participants who received praise for success on the easy task and neutral feedback for failure on the difficult task were: (a) more likely to report that their performance was better than the other participant's;

(b) more likely to evaluate their performance on the initial test as being better; (c) have more positive affect and less negative affect about their performance than the participants who received neutral feedback for success on the easy task and blame feedback for failure on the difficult task. In other words, praise had a uniform positive impact on participants' motivation and affect. Also the participants who received the positive feedback (praise for success and neutral feedback for failure) also perceived the evaluation computer as being more fair than those who received negative feedback (neutral feedback for success and blame for failure).

Interestingly, for the most part (the exception will be discussed later), there was no effect of scoring, that is, the participants' response to the affective feedback did not depend on whether or not their ability level had been measured. There was also no interaction effect (except for one variable, discussed next).

The only variable where the determination of ability level appeared to matter was when the participants were asked to evaluate the performance of the other participant. In this case, the participants who believed that their ability level had been measured were more likely to view their performance as being better than the participants whose ability level had not been measured.

In summary, these findings are somewhat consistent with the CASA paradigm in that affective feedback from computers does make a difference in how the participants perceive the overall task as well as their self-perception of ability and confidence. This is further support for the CASA paradigm and undermines the argument that people would ignore or not accept affective feedback from computers.

However, matters are not that simple. The participant's responses to the affective feedback did not completely match the findings of the Meyer et al. (1986) study. This is discussed in the discussion of the second research question.

Question 2: What are the similarities and differences between the Human-Human interaction case and the HCI case, that is, do people respond the same way to computer feedback as they do to feedback from humans?

The strong CASA perspective would predict no differences between the results of Meyer et al. (1986) and the current experiment. As previously discussed, the results of the current study indicate that affective feedback from computers does make a difference in how participants view themselves and their ability and motivation. However, the results do not exactly match what was found in the Meyer et al. study. The Meyer et al. experiment (where feedback was provided by human evaluators) had found that when the par-

ticipants believed that the evaluators knew their "ability" they tended to use the criticism they received for failure on a difficult task as indicating that the evaluator expected them to do better. Correspondingly those who were praised for success on an easy task inferred that they were of lower ability. Essentially this can be clearly seen in the interaction effect between the feedback and the scoring experimental conditions. This interaction effect indicates that (in the human feedback case) participants responded differentially to the feedback. Being praised for success on an easy task (and receiving neutral feedback for failure on a difficult task) was considered negatively when the participants believed their ability level was known. Correspondingly, participants responded favorably to receiving neutral feedback for success on an easy task and receiving blame feedback for failure on a difficult task.

The current study found no significant interaction effects and for the most part significant effects were found only for being praised for success on an easy task and not being blamed for failure on a difficult task. Thus, though the participants in the current study did respond to the feedback it was at a somewhat superficial level. In other words, when receiving feedback from the computer, the participants seemed to take it at "face value." Praise was seen as being better and more motivating irrespective of whether the task was easy or difficult and irrespective of whether their ability had been measured!

DISCUSSION

If the "computers as neutral tools" school of thought is right we ought to have found no difference between the groups based on the experimental manipulations (differential feedback crossed with test scored vs. not scored). The fact that we do find these differences indicates that people do respond to affective feedback from computers and that it does make a difference to their self-perception and motivation. In that respect the CASA paradigm is supported.

However, if the CASA paradigm is correct across the board, there ought to be no difference between the results of the human-human study and the results of the current HCI experiment. The results of this study do not completely support this position either. It appears that people accept affective feedback from the computer but do not necessarily respond to it the same way as they do if they receive the same feedback from humans. One way of interpreting this is that people accept feedback from the computer at face

value. In the case of receiving feedback from humans, people are more interpretive, and seek to understand the context of the feedback and this is not something they do when working with computers.

The results of this study are interesting primarily because they indicate that the psychological aspects of human computer interaction are complex and are hard to explain using simplistic frameworks such as "computers are neutral tools" or "interacting with computers is just the same as interacting with humans." The results of our study indicate that there is partial validity to the CASA paradigm, because, participants in the study *did* respond to affective feedback from the computer. However, they did so somewhat indiscriminately, disregarding the context within which the feedback was offered. The Meyer et al. (1986) study showed that when people receive feedback from humans they factor in the context within which this feedback was offered. Praise for success on an easy task is discounted, and may even reduce motivation, particularly when they believe that their ability level is known. Blame for failure on a difficult task leads to positive affect, when ability level is known, because the participants infer that they would not be blamed unless the evaluator did not think they could have solved the problem, that is, had a high opinion of their ability. This is a complicated and sophisticated chain of inferences that the participants clearly made in the Meyer et al. study. It is as clear that the participants in the HCI case did not make this chain of inferences. They merely saw praise from the computer as being positive, irrespective of whether or not they believed their ability level was known, or whether or not the task was easy or difficult.

We hypothesize that this difference in results between the Meyer et al. (1986) study and the current experiment could be due to the "level of processing" participants were willing to do. In the human evaluator case the participants clearly were able and willing to engage in a series of deep psychological chain of inferences about intentionality. The results of the current study indicate that, despite the findings of the CASA paradigm, people may be unwilling to engage in a similar level of deep processing about the intentionality and reasons behind the computer's responses. Clearly this study was not designed to explain why this happens though it is credible that we take feedback from the computer at face value but do not engage in the kind of "deep psychological processing" about intentionality that we do with human respondents. Nass and Moon (2000) have argued that people approach interaction with computers mindlessly (Gilbert, 1991; Langer, 1992) seeing them as instantiations of "social scripts" based on our prior interactions with social agents (namely humans). They argue that mindlessly approaching interaction with computers is what leads to the CASA effect. In fact most

of the research in the CASA paradigm have been based on standard social scripts (politeness, reciprocity, stereotyping, etc.).

The difference between the results of the previous human-human inter-action study and our current study could be the result of the fact that there was no existing social script that our participants could fall back upon. They had to actively process the information, and in a situation such as this, they clearly distinguished between human and computer respondents. This indi-cates that the participants in the study do not expect computers to be intelli-gent or make sophisticated inferences (i.e., that computers are incapable of a deeper level cognitive processing). So in the Meyer et al. (1986) experiment, the participants acceptance of feedback from the human evaluator was based on a complex chain of inferences about the intentions and purposes of the human evaluator. For instance, when the participants were criticized by the human evaluator for failing on a task, it led them to infer that the evaluator had respect for the participant's intelligence and "knew" that the participant could do "better." This is a complex chain of inferences to be made by the participant about the cognitive processes of the evaluator. It could be that the participants were not willing to make the same level of inferences about the "computer evaluator."

It could be that it needs more than text-based feedback for the social scripts to get instantiated. The use of newer interface technologies, such as anthropomorphic agents, or the use of human (or computer generated) voice interfaces may lead to different results (more in keeping with the Meyer et al. [1986] study). Clearly this is an area worthy of further investigation.

As sometimes occurs in research, this experiment raises more questions than answers. That said, the results of this study indicate is that it is impor-tant for designers of educational technology to move beyond an emphasis on merely cognitive or information processing aspects of using and learn-ing from computers. Failure to recognize the existence of people's social responses towards media could thwart the pedagogical goals of designers of educational software. We believe this is an extremely fascinating area of work. This research can contribute to our knowledge in multiple domains: including educational psychology, social psychology, artificial intelligence, philosophy, HCI, cognitive science, and the theory and practice of software development.

References

Alvarez-Torres, M., Mishra, P., & Zhao, Y. (2001). Judging a book by its cover. Cultural Stereotyping of interactive media and its effect on the recall of text information. *Journal of Educational Multimedia and Hypermedia, 10*(2), 161-183.

Barkow, J., Cosmides, L., & Tooby, J. (1992). *The adapted mind: Evolutionary psychology and the generation of culture.* New York: Oxford University Press.

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: MIT Press.

Bates, J. , (1994). The role of emotion in believable agents. *Communications of the ACM*, *37*(7), 123-125.

Brophy, J.E. (1981). Teacher praise: A functional analysis. *Review of Educational Research*, *51*(1) 5-32.

Cassell, J., & Thorisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence, 13*(4-5), 519-538.

Cawsey, A. (1989). Explanatory dialogues. *Interacting with Computers*, *1*, 69-92.

Delamere, T. (1996). The importance of interlanguage errors with respect to stereotyping by native speakers in their judgments of second language learners' performance. *System*, *24*(3), 279-297.

Dennett, D. (1987). *The intentional stance.* Cambridge, MA: MIT Press.

Ferdig, R. E., Mishra, P. (2004). Emotional responses to computers: Experiences in unfairness, anger and spite. Journal of Educational Multimedia and Hypertext. 13. 2, p. 143-161.

Ferrari, G. (1986). Man machine interaction in natural language: Computational models for dialogue. *Current Psychological Research and Reviews*, *5*(2), 163-174.

Fogg, B. J., & Nass, C. (1997). Silicon sycophants: Effects of computers that flatter. *International Journal of Human-Computer Studies*, *46*(5), 551-561.

Gaines, B. R. (1981). The technology of integration—dialogue programming rules. *International Journal of Man Machine Studies*, *14*, 133-150.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*(2), 107-119.

Ginott, H. (1965). *Between parent and child*. New York: Macmillan.

Gopnik, A., & Meltzoff, A. N. (1997). Words, thoughts and theories. Cambridge Mass.: Bradford, MIT Press.

Graham, S. (1984). Communicating sympathy and anger to black and white children: The cognitive (attributional) consequences of affective cues. *Journal of Personalitv and Social Psvchology*, *47*(1), 40-54.

Graham, S. (1990). Communicating low ability in the classroom: Bad things good teachers sometimes do. In S. Graham & V. Folkes (Eds.), *Attribution theory: Applications to achievement, mental health, and interpersonal conflict*, (pp. 17-36). Hillsdale, NJ: Lawrence Erlbaum.

Graham, S. (1991). A review of attribution theory in achievement contexts. *Educational Psychology Review, 3*, 5-39.

Henderlong, J., & Lepper, M. R. (2002). The effects of praise on children's intrinsic motivation: A review and synthesis. *Psychological Bulletin, 128*, 774-795.

Hershey, K., Mishra, P., & Altermatt, E. (2005). All or nothing: Levels of sociability of a pedagogical software agent and its impact on student perceptions and learning. *Journal Educational Multimedia and Hypermedia*, *14*(2), 113-127.

Isbister, K., & Nass, C. (2000). Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies, 53*(1), 251-267.

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, *11*, 47-78.

Kanouse, D.E., Gumpert, P., & Canavan-Gumpert, D. (1981). The semantics of praise. In J.H.Harvey, W.Ickes, & R.F.Kidd (Eds.), *New directions in attribution research* (97-115). Hillsdale, NJ: Lawrence Erlbaum.

Kaufmann, W.J., & Smarr, L. L. (1993). *Supercomputing and the transformation of science*. New York: W. H. Freeman.

Lajoie, S., & Derry, S. (1993) *Computers as cognitive tools*. Hillsdale, NJ: Lawrence Erlbaum.

Langer, E. J. (1989). *Mindfulness.* Reading, MA: Addison-Wesley.

Langer, E. J. (1992). Matters of mind: Mindfulness/mindlessness in perspective. *Consciousness and Cognition*, *1*, 289-305.

Laurel, B. (1997). Interface agents: Metaphors with character. In J.M. Bradshaw (Ed.), *Software agents* (pp. 67-77). Cambridge, MA: MIT Press.

Lepper, M. R., & Chabay, R. W. (1985). Intrinsic motivation and instruction: Conflicting views on the role of motivational processes in computer-based education. *Educational Psychologist*, *20*, 217-231.

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S.P. Lajoie & S.J. Derry (Eds.), *Computers as cognitive tools* (pp. 75-106)*. Hillsdale, NJ: Lawrence Earlbaum.

Maes P. (1997). Humanizing the global computer. Interview in: *IEEE Internet Computing, 1*(4).

Maes, P., & Kozierok, R. (1993). Learning interface agents. In *Proceedings of the Eleventh National Conference on Artificial Intelligence, AAAI-93*,(pp. 459-464), Washington, DC.

Meyer, W.U. (1982). Indirect communications about perceived ability estimates. *Journal of Educational Psychology, 74*, 888-897.

Meyer, W.U. (1992). Paradoxical effects of praise and criticism in perceived ability. In W. Strobe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 3, pp. 259-283). Chichester, UK: Wiley.

Meyer, W.U., Mittag, W., & Engler, U. (1986). Some effects of praise and blame on perceived ability and affect. *Social Cognition*, *4*(3), 293-308.

Mishra, P., Nicholson, M., & Wojcikiewicz, S. (2001/2004). Does my wordprocessor have a personality? Topffer's law and educational technology. *Journal of Adolescent and Adult Literacy*, *44*(7), 634-641. Reprinted in B. C. Bruce (Ed.), *Literacy in the information age: Inquiries into meaning making with new technologies* (pp. 116-127). Newark, DE: International Reading Association.

Mishra, P., Wojcikiewicz, S., & Nicholson, M. (2002). Taking things at face value: A review of the media equation. *Journal of Educational Computing Research, 26*(2), 219-226.

Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81-103.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45*(6), 669-678.

Nass, C., Moon, Y., & Carney, P. (1999). Are respondents polite to computers? Social desirability and direct responses to computers. *Journal of Applied Social Psychology, 29*(5), 1093-1110.

Nass, C., Moon, Y., & Green, N. (1997). Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology, 27*(10), 864-876.

Nicholls, J. G., & Miller, A. T. (1984). Reasoning about the ability of self and others: A developmental study. *Child Development*, 55(1990-1999).

Nielsen, J. (1990). Traditional dialogue design applied to modern user interfaces. *Communications of the ACM*, *33*, 109-118.

Parsons, J., Kaczala, C., & Meece, J. (1982). Socialization of achievement attitudes and beliefs: Classroom influences. *Child Development*, *53*, 322-339.

Picard, R.W. (1997). *Affective computing.* Cambridge, MA: MIT Press.

Pinker, S. (1997). *How the mind works*. New York: W.W. Norton & Company.

Pridemore, D. R., & Klein, J. D. (1991). Control of feedback in computer-assisted instruction. *Educational Technology Research and Development*, *39*, 27-32.

Raisler, I. (1976). Differential response to the same message delivered by native and foreign speakers. *Foreign Language Annals*, *9*(3), 256-259.

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media as real people and places*. Cambridge: Cambridge University Press/CSLI.

Schmandt, C. (1994). *Voice communication with computers*. New York: Van Nostrand Reinhold.

Schurick, J. M., Williges, B. H., & Maynard, J. F. (1985). User feedback requirements with automatic speech recognition. *Ergonomics*, *28*, 1543-1555.

Selker, Ted (1994): COACH: A teaching agent that learns. *Communications of the ACM*, *37*(7), 92-99.

Shepard, R. N. (1990). *Mind sights: Original visual illusions, ambiguities and other anomalies*. New York: W. H. Freeman.

Stipek, D. J. (1993). *Motivation to learn: From theory to practice.* Boston: Allyn & Bacon.

Stipek, D. J. (1996). Motivation and instruction. In D.C. Berliner & R.C. Calfee (Eds.),. *Handbook of educational psychology* (pp. 85-112). New York: Simon & Schuster.

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

Wang, P. S. P. (1991). Character and handwriting recognition: Expanding frontiers. *International Journal of Pattern Recognition and Artificial Intelligence*, *5*(1-2), 1-382.

Weary, G., Stanley, M.A., & Harvey, J. H. (1989). *Attribution*. New York: Springer-Verlag.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.

Weiner, B., Graham, S., Stern, P., & Lawson, M. (1982). Using affective cues to infer causal thoughts. *Developmental Psychology*, *18*, 278-286.

Weizenbaum, J. (1976). *Computer power and human reason*. San Francisco: Freeman.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

Note