

Running head: UNDERSTANDING EXPRESSIVE MACHINES

Understanding Expressive Machines:
Research and Theory in the CASA Framework

Paul Aumer-Ryan

The University of Texas at Austin

December 13, 2005

INF 391D.8

Abstract

Affective design, alternatively called (with slight variations in meaning) hedonic design, emotional design, affective human factors design, human-centered design, and empathetic design, is essentially the focus on the role of human emotions and their influence on the way we understand and relate to artifacts. Because of the ubiquity of emotions in human experience, affective design can be applied to all fields which have as a focus the interaction between humans and objects (broadly defined), and is hence multidisciplinary. Researchers and practitioners come from fields as diverse as information science, ergonomics, philosophy, cognitive science, computer science, artificial intelligence, and linguistics.

This paper will explore the use of theory in—as well as the practical applications of—research in a specific subset of affective design, usually referred to as “computers as social actors” (CASA). The specific focus will be on research and theory that is directed toward two fundamental questions: first, how do people respond to computers that “display” (here loosely defined) emotions and respond to emotions themselves, and second, what does a computer that “displays” emotion look like, i.e., what are its potential applications and implications? The first question here has a strong basis in practice, and as such is addressed by many empirical studies within the related literature, while the latter question is more theoretical in its aim to create and understand new models of social interactions with computers.

One of the benefits that the CASA framework offers is the disentanglement of two roles that these “expressive machines” play in an interaction with a person. First, the machine is an emotional actant¹ when it displays emotions (and when the person attributes emotionality to it); second, it acts as an emotional respondent by “interpreting” the emotions of the person interacting with it. Because of this understanding, the CASA framework grounds its focus in the

communicative act between actants, and as a whole moves away from artifact-centered research methods toward those that are human- and agent-centered. Consequently, the study of CASA lends itself to description under several theoretical approaches, including social epistemological and social constructivist approaches.

There is a fair amount of contention among the proponents of these approaches—the loudest of which is between the cognitive and social camps—and much of this contention comes from the ill-defined nature of emotions themselves. Since the study of emotions as mental processes is somewhat at odds with traditional behavioristic notions of how the mind works, and since computers are habitually considered non-thinking, unexpressive apparatuses, various incongruous models of machine emotion have been advanced in the literature. Additionally, in much the same way that information is considered both “thing” and “concept,” so too do definitions of emotion range from the physical to the abstract. These definitions will be explored along with their associated theories.

Finally, since affective design is still somewhat of a nascent field, much of the literature associated with the computers as social actors framework is conjectural. This speculative and probing approach offers a wide variety of methodological tools for the development of expressive machines. I will explore and evaluate the most common of these methodological tools and their application to CASA and information sciences as a whole, as well as offer my own interpretations of the CASA framework and its possible implications.

Origins of Affective Design

The broad category of affective design is actually a multidisciplinary approach relying upon and affecting the areas of information science, philosophy, cognitive science, computer

science, artificial intelligence, and software and game development. In a familial sense, it is the grandchild of human factors (ergonomics) via human-computer interaction and in opposition to classic cognitivism (Aboulafia & Bannon, 2004). Its role as a separate field of inquiry can be traced back to Western philosophical notions originating with Aristotle (384/340 B.C.E.) and further developed by Descartes (1595/1650 C.E.), who envisioned a distinct separation between the rational mind and the emotional mind, as well as divisions between subjective and objective experience.

The human mind is typically dichotomized into two categories based on these distinctions: rational and emotive. While recent literature (Lewis & Haviland-Jones, 2004; Trapp, Petta, & Payr, 2003; Goleman, 1995; Damasio, 1994) suggests that emotions underscore every aspect of human existence, classical treatments often focus on behavioristic or cognitivist approaches that dismiss (or ignore) the impacts of emotions on human behavior and interaction. Such approaches focus on input-output (stimulus-response) paradigms of the mind, and tend to view emotional responses as noise affecting otherwise pure data. Affective design, on the other hand, intends to propel emotion to the forefront of design.

The computers as social actors² framework has considerable overlap with affective design in its theoretical underpinnings and practical applications, with two exceptions. First, affective design is concerned with the emotional impact of design choices, while the CASA framework focuses on the broader social interaction between human and computer (which includes but is not limited to emotional responses). Secondly, because CASA utilizes many theoretical concepts from the social sciences that are typically assumed to apply only to human-human interactions, research within the framework is strongly applicable to the design process—

but it is not limited to this effect. Other areas, including usability studies and communication research, are also beneficiaries of this research.

In a historical sense, the CASA framework also relies on its opposition to the traditional cognitivist treatment of computers as emotionless, rational, and structured input-output systems; the CASA framework fills the gap left by this cognitivist approach in terms of the social interactions between humans and computers. Where the cognitivist approach is concerned with, for example, organizational schemes and computational efficiency, the CASA framework moves the focus onto exploring social situations and the understanding of the emotional interplay between human and computer.

History of CASA

While computers have long been recognized for the social impact they have on their environments (Brabent, 1983), the seeds of the notion that computers can be treated as human-equivalent social beings rather than simple tools can at be traced to Turkle's (1984) book, *The Second Self: Computers and the Human Spirit*. In it, Turkle divested the computer of its position as a tool that exists simply to fulfill some task and explored the social impacts that human-computer interaction creates. She goes beyond the simple notions of social impact (e.g., workplace changes due to the introduction of computers) to claim a much broader impact—on self-conception, awareness of others, and the perception of the world itself.

The Second Self broke new ground, not in its discussion of the social and psychological responses that humans exhibit in the presence of computers, but rather in the perspicacious idea that humans invoke psychological and social properties within the computers. Let us examine this dramatic reversal for a moment. That computers, by their presence, affect the world around

them is not a brazen claim; that is true of most any artifact. But that humans ascribe social and psychological characteristics to computers—action, intentionality, agency, and thought, to name a few—is a much grander proposition. This recognition of the social understanding of computers—computers as social actors, as actants—created the foundation for what was to come next.

In 1994, Nass, Steuer, & Tauber presented their influential paper, “Computers are Social Actors,” at the SIGCHI (Special Interest Group in Computer-Human Interaction) conference in Boston, which inaugurated (in a post hoc, historical sense) the development of the computers as social actors framework. Nass et al. elucidated a new research methodology that they claimed would have broad applicability to both design and usability testing. This methodology relied on a single assumption: “individuals’ interactions with computers are fundamentally social” (p. 72). Upon this theory was built a scaffolding for approaching future research projects that addressed and tested it: first, a researcher should select some social science finding (from fields like social psychology or sociology) that describes human interactive behavior, and substitute “computer” in place of “human” in the theory or method; then, when redesigning the study, the researcher should give the new computer actant several characteristics that are associated with humans (e.g., voice, reaction/interaction, and common human roles). At this point the researcher would be prepared to see if the social finding is applicable in a human-computer context, and if the computer can be properly seen as a “social actor.”

To support their novel approach, Nass et al. presented the findings from five studies they performed that utilized social science findings as outlined above. Because of their importance in defining the boundaries of the new framework and in influencing a shift in the conceptualization

of the role of computers in human-computer interaction, a detailed explanation of these studies is necessary.

All five studies have the same basic construction: a human participant takes a multiple-choice test (on topics ranging from mass media, computers, social customs, and love and relationships) on a computer while another computer acts as a tutor that verbally offers helpful information to the test-taker. A third computer acts as a vocal evaluator of the tutor computer, judging whether or not the information supplied by it is helpful. Note that in the various studies the three roles for the computer (presenting the multiple choice test, offering advice/tutoring, and evaluating the tutor's performance) are either performed by separate computers with the same physical configuration (i.e., on the same brand of computer, displaying the same interface, and using the same monitor and audio speaker setup), or all roles are performed by the same computer. These interactions between same-computer and different-computer are also used to determine if participants treat computers as social beings.

The first study, perhaps the easiest to summarize and the most frequently cited in literature that summarizes CASA research (Trappl & Payr, 2003; Miller, 2004), is concerned with social etiquette and politeness. The purpose of the study was to find out if participants would apply common rules of politeness when interacting with a computer, i.e., would participants submit a more positive evaluation of the tutor computer's performance when performing the evaluation on the tutor computer? The findings by Nass et al. showed that this outcome was in fact the case: participants' evaluations were more positive (e.g., the tutor was perceived as friendlier and more competent) when performed on the tutor computer than they were when performed on a different computer or on pencil and paper. To further support the computer's position as a social actor, the findings showed no significant difference between the

evaluations taken on a different computer and those taken on pencil and paper (i.e., this lack of significance shows that there was not a bias introduced because of the evaluation being performed on either a computer or pencil and paper).

The second study (further discussed in Nass, Steuer, Tauber, & Reeder, 1993) illustrates the application of the social norm that sanctions praising others and criticizing oneself (as opposed to criticizing others and praising oneself). After interacting with the tutor computer, participants listened to either the tutor computer rate its own performance or another computer (using a different voice) rate the tutor computer's performance. Results demonstrated that participants interpreted the tutor computer criticizing itself as more friendly than the other computer criticizing the tutor computer, and the other computer praising the tutor computer as more friendly than the tutor computer praising itself.

The third study explored how participants identified the computer as a social actor—is it the audible voice of the computer that is the locus of the social actor, or is it the physical box (the equipment)? By manipulating the pairing of the voice and the box in study 2 (i.e., giving the other computer the same voice as the tutor computer when evaluating it, and giving the tutor computer a different voice when evaluating itself), the study showed that participants were more likely to interpret a different voice on the same box as a different social actor, and the same voice on a different box as the same social actor. In other words, participants were treating the voice of the computer rather than its physical presence as the social actor.

The fourth study explored whether or not participants would apply common gender stereotypes to computers whose voices were either male or female. The stereotypes explored in this study were: “‘Praise from males is more convincing than praise from females,’ ‘Males who praise are more likable than females who praise,’ and ‘Females know more about love and

relationships than males” (p. 76). The study confirmed that these stereotypes were applied equally to computers (however inappropriately) as they are to humans. Specifically, participants rated tutor computers higher (more forceful, affectionate, and sympathetic) when they were positively evaluated by a male-voiced evaluator computer as compared to a female-voiced evaluator computer. Participants also rated the male-voiced evaluator computers higher (more forceful, sympathetic, and warmer) than their female-voiced counterparts. Finally, tutors with female voices were seen as more sophisticated and knowledgeable when discussing the topic of love and relationships.

The last of the five studies sought to determine if participants were actually responding socially to the human programmer of the computer instead of the computer itself (and thus negating most of the interesting conclusions of CASA). By manipulating the way in which the human experimenters referred to the tutor computer (either as “the computer” or as “the programmer”), the researchers were able to observe a significant difference in the participants’ evaluations of the tutor computer. In general, the computer that was referred to as “the computer” (and which referred to itself as “I”) was rated higher (more capable and more likable) by the participants than the computer that was referred to as “the programmer.” This difference was sufficient enough to claim that the participants would treat the computer differently when they believed that it was an individual—but it also showed that the computer’s social “self” could be removed if it was treated as the work of some human programmer.

Reeves & Nass further solidified the CASA framework with their compilation of 35 social science studies modified to explore computer actants in *The Media Equation* (Reeves & Nass, 1996), and also began to engage in more exhaustive philosophical discussions of the social nature of human-computer interactions. This work created the new terminology of “the media

equation,” which attempted to explain why humans treat computers and other media as social beings. The media equation is defined thusly: “media equals real life” (alternatively, mediated life equals real life). Let us pause for a moment to examine the peculiar terminology at work here. The usage of “media” here is somewhat suspect, since computers are not necessarily media; rather, they have the ability to present forms of media (audio, video, textual information, and Web sites, for example). The implicit suggestion is that the computer offers a “mediated” view of reality, by communicating information that has, at some time, been a part of the experience of another human being. Thus the media equation is not specifically referring to the effects of media upon our perception of reality, but rather how the “mediated world” is categorically different from the “real” world. Further, the proposition is that the two worlds are both viewed under a single umbrella of “reality.”

The term “media” relies heavily upon the dichotomy of the mediated world and the real world. Reeves & Nass proposed that humans have not had time to adapt—in a biological and evolutionary sense—to the different situations created by the introduction of recent technologies (like computers) that allow for the presentation of media without the historically concomitant partner that is human agency. Note that there is a subtle assumption here that the level of interactivity and exchange between, say, a human and a computer is categorically different than the relationship between a human and a more mature information “technology,” like a printed book. In Nass, Steuer, & Tauber (1994), for example, this categorical difference relies on the ability of the medium to interact with a person using a human voice.

The philosophical proposition, then, is that humans do not differentiate, on a social level, between an interaction with a human actor and one with a computer actant (or at least that standard social mores and etiquette are unconsciously carried from traditional human-human

interactions). According to Reeves and Nass, “[m]odern media now engage old brains. People can’t always overcome the powerful assumption that mediated presentations are actual people and objects” (1996, p. 12). In short, the human mind interacts with the mediated world in the same fashion that it interacts with the real world. The implications of this proposition range from the socially responsible—such as using affective design methodologies in the design of computer systems or in usability studies—to the socially deleterious—such as preying on common gender stereotypes to bolster a product’s effect upon consumers.

The 35 studies discussed in *The Media Equation* run the gamut of social science studies: from manners, etiquette, and politeness, to personality, emotions, and social roles. Computers can be treated politely; they can cause us to impose gender stereotypes; and they can threaten us and calm us, elicit anger and ease our frustration. Interestingly, Reeves and Nass credit both statistical and empirical methods from the social sciences for the discoveries presented in *The Media Equation*. Because of the counterintuitive nature of many of the findings—that is, most people do not consciously recognize their treatment of computers as social actors—techniques such as focus groups and open-ended comments on the social treatment of computers were not useful in seeking parallels between mediated life and real life. Statistical analyses and empirical observations were the most successful methods in finding results that supported computers being social actors.

Since the publication of *The Media Equation*, many different research groups have incorporated and elaborated upon the ideas raised by the CASA framework (at the time of this writing, it had been cited nearly 500 times, according to Google Scholar). While a full discussion of these research groups is beyond the scope of this paper, several of them warrant mention. Klein, Moon, & Picard (2002), operating under the umbrella term of “affective

computing,” apply CASA to usability and design contexts, and seek to address and reduce negative emotions, like users’ frustration. Picard’s book (1997) and paper (2002) address some of the more philosophical aspects within CASA, such as the meanings and assumptions behind claiming a computer “has” emotions. Fogg’s (1998) CAPTology (computers as persuasive technology) explores the influential behaviors of computers to better understand the potential dangers and benefits they can have in a social sphere. Finally, Topffer’s law (from Mishra, Nicholson, & Wojcikiewicz 2001-2003), which states, “all interfaces, however badly developed, have personality,” codifies the importance of designing with the CASA framework in mind. The aim here is that since social responses to computers are unavoidable, the design process should always account for them (e.g., an error message that blames a user for the problem will be poorly received because of its breach of etiquette).

All of these works are heavily invested in the concept of the media culture, and in the social effects that non-human, interactive actants have on the human perception of reality. Even though CASA has a specific focus on computers, it does not discount social responses to other artifacts, however, there is some contention about whether or not computer technology offers a qualitatively different social experience than with a more traditional form of media. Picard (2002) mentions the social response to a doll or a stuffed animal as contrary examples. Ultimately, the dividing line—albeit a fuzzy one—seems to separate media that are more animated and interactive from those that are not, with the understanding that more mobile, complex, interactive, and responsive artifacts will evoke a stronger social response from human actants. A doll may impersonate a human (and may actually be more successful in imitating the appearance of one), but it can hardly do something unexpected the way a more complex computer can.

Criticisms

Let me take a moment to raise and attempt to respond to some of the more obvious criticisms of the media equation. First, Nass, Steuer, & Tauber's (1994) third study implies that participants identify a computer's "social self" in its voice, not in its physical makeup. In other words, it is the voice (and a recorded human voice, at that) that participants are treating as a social being. A first reaction might be, "well, of course." It is not a computerized or digitized voice that the participants are listening to—it is a recording of a human speaking. Perhaps the participants are only responding to the voice in the same manner that they might respond to, say, an answering machine message (i.e., there is another human somewhere in this communication process, so it is not really human-computer interaction). Ostensibly, we are communicating with the person who created the answering machine message, but we are in fact only responding to a mesh of circuitry that sounds like our friend. Are we to claim that the answering machine is another media form, one that we treat as a social actor?

I believe the answer to this question is "no," in the sense that we do not believe it is the answering machine we are communicating with; while the answering machine does mediate the communication between us and our friend, it is not an active or conspicuous element of that interaction. That being said, there are many concerns related to the disembodiment of human communication because of technologies like telephones, and many of these are strongly related to the mediated world/real world dichotomy that the media equation rests upon.

Another argument against computers being social actors relies on their multifunctional role. By their nature, computers are adept at performing multiple tasks, e.g., word processing, electronic communication, and game playing. Each of these tasks carries with it a specific set of interactions between the human and the computer. This multifunctionality may prompt us to

ask: To what, exactly, are we ascribing social behavior? Is it the physical computer, the one we can see and hear? Or is it the specific application or software piece that we are interacting with at the moment? And if it is the computer itself, are we then uncomfortable with the computer displaying wildly divergent personality types when it runs different programs?

Several of the studies exploring CASA addressed this problem, and at least came to a partial conclusion: it is the voice of the computer, not its physical presence, that is the social actor. While the generous use of pre-recorded human voices across all the relevant studies hints at a weakness in their ability to explore computers as social actors, it does imply that the social interaction was fleeting and impromptu. In other words, social rules are being applied to specific interactions rather than to some long-term relationship with a computerized agent. In some sense, this transient social situation reinforces the transitive nature of the anthropomorphic act: computers and other media are treated socially because we are accustomed to interacting socially with humans in everyday life. We will return to this discussion in a moment.

Finally, one of the strongest criticisms of the media equation and CASA in general is concerned with their apparent novelty: since computers are interactive—they respond to us, and we respond to them—it is fitting that we treat them in a social manner, because interaction is at the core of social behavior. This realization is no more perceptive than the knowledge that humans often anthropomorphize other non-human entities they come in contact with, from other animals to diseases to automobiles.

This criticism can be answered with equal strength by addressing the very conspicuous metaphor that operates in anthropomorphism. Assigning human capabilities to non-human artifacts and living things we interact with is basically a metaphorical process. In general, this process is not denied by the person enacting it. Someone who sees, for example, an angry

expression in the headlights and grill of a speeding automobile will readily identify the metaphorical characteristics of the association between plastic and a human face. They also may explain the reason for anthropomorphizing the automobile: since the automobile was speeding, it was acting in a dangerous fashion; similarly, an angry person is also potentially dangerous; therefore, viewing an angry face in the automobile is a type of safety mechanism that encourages avoidance. The unexpected conclusion of the CASA framework is that people do not readily admit or deem reasonable their social treatment of computers: “People respond socially and naturally to media even though they believe it is not reasonable to do so, and even though they don’t think that these responses characterize themselves” (Reeves & Nass, 1996, p. 7). What makes the study of CASA distinct and original is this counterintuitive treatment of the social response by the human actant and his/her disbelief in the metaphorical act.

Philosophical Assumptions

I noted earlier that many of the studies performed in the CASA framework relied on empirical methods derived from the social sciences. It would be tempting to conclude, then, that research in the CASA framework relies on an empiricist philosophical outlook. While empirical methods are heavily relied upon, the epistemological foundations of empiricism are not necessarily supported by CASA research. Instead, assumptions about the origin and creation of knowledge are more closely related to epistemological frameworks like social constructivism. Where an empiricist view may lead us to ask questions about how mediated interactions impersonate “real life” experiences—such as why the human mind often cannot socially differentiate between mediated and real life—the social constructivist view focuses more on the mediated experiences themselves. Since the bulk of research in the CASA framework explores

the applicability of various social situations to human-computer interaction, the epistemological approach used is very similar to a social constructivist one.

Although research in the CASA framework concentrates upon the social aspects of the computer, this does not discount the computer's role as an information system in Buckland's (1991) sense of the term. If we focus now on the computer's ability to interpret emotions or social responses exhibited by the human it is interacting with, we can discuss the intricacies of the interaction, or communication, between the human and the computer actant. One of the key aspects of an information system is that it can only deal with information-as-thing—that is, for the computer to be able to process information, it must be reduced, quantified, and represented in the internal state of the computer. How then do Buckland's three categories of information—information-as-process, information-as-knowledge, and information-as-thing—bear upon the view of computers as social actors? If we consider as information the actual emotional or social response of the human to the computer, it seems most likely that this intangible information would be classified as information-as-knowledge. However, the expression of this emotional or social response must be realized in the physical world—as a facial expression, gesture, or spoken phrase, for example. It is here that the information-as-knowledge (the social or emotional response to the computer) is converted into information-as-thing. Consequently, the computer as information system will be able to interpret and internally represent the human's social or emotional response, decipher its meaning in the fashion of any information system, and then communicate its own social or emotional response back to the human. In some sense, this outline is quite similar to that of a human-human interaction, except that it is contestable whether or not the computer actually represents its own emotional state in terms of information-as-

knowledge before converting it again into a physical manifestation that is communicated back to the human.

While the information-as-thing approach may be appropriate for the designers of information systems (such as the designers of computers that respond to human emotion), it is difficult to apply to emotions and social responses in general because of their intangible and fleeting nature. In much the same way that we can view the projection of a shadow and deduce the form of the object that is casting it, we understand the physical representation of emotion as information-as-thing, but it is one step removed from the actual information-as-knowledge. Further, this encoding and decoding is a translation process, and information can be lost along the way. In the study of computers as social actors, most researchers avoid the treatment of emotional responses as information-as-thing, except insofar as they are to be understood in a quantitative fashion for statistical analysis. The most common assumption is that there is some underlying information-as-knowledge (for example, the emotion experienced by the human actant) that must be translated before being applied to the computer.

Philosophical Implications

I would like to return to the metaphorical act implicit in the computers as social actors framework. First and foremost is the claim that computers are viewed in a substantially different fashion than in the traditional anthropomorphic act of considering, say, an animal as a metaphorical human. In the latter case, we may claim that the metaphor involved is explicitly apparent to the human actant performing the metaphorical act, and that there is a well-described reason for the substitution. For example, a person who enters a pet store looking for a companion may choose an animal that reminds him or her of a family member—if pressed, this

person might defend his/her comparison by claiming that the pet would bring him/her more joy because of the prior relationship with the family member. In comparison, the treatment of computers as social actors relies on no such prior decision; it is, in effect, automatic and unconscious, without purpose and possibly counterintuitive.

Nass, Steuer, Tauber, & Reeder (1993) discuss four cues that induce social responses in humans: “words for output,” “responses based on multiple prior inputs,” “the filling of roles traditionally filled by humans,” and “the production of human-sounding voices” (p. 111). Because the computer exhibiting these cues is interacting with a human in the same fashion that he/she would interact with another human, I contend that the metaphor (the computer is a social actor) goes beyond the effects of an anthropomorphic act. When the computer is recognized as a social actor, it substantially changes the perception of that computer. When a computer is communicating using human language, exhibiting complex responsive behavior similar to that of a human, acting in roles historically occupied by a human, or using a human voice, it is, in its interaction with a person, a human actant. The metaphor is no longer perceived as a considered substitution, but becomes a full-fledged replacement. The computer is not impersonating a social actor; it *is* a social actor, regardless of its lack of humanity. We have moved outside the realm of metaphor and into the realm of ordinary social interaction, where the human participant is not actively questioning the computer’s status as a social being.

Ethical Implications

The literature is replete with predictive discussions on the potential breach of ethics involved in perceiving computers as social actors and in giving computers emotions (see Miller, 2004; Trappl, Petta, & Payr, 2003; Friedman, 1997; Picard, 1997; Reeves & Nass, 1996). For

example, at one point Picard addresses the need to distance the meaning of emotions in computers from that of the existence of a soul: “[W]e should be clear to the public that giving a machine emotion does not imply giving it a soul” (2002, p. 233). At another point she discusses the eventual inability for the designer to guide the emotions “in” the computer: “we are not going to be able to control ... these agents at some point. I think this is a responsibility decision to make as designers, while we are in control” (p. 216). Friedman & Millett (1997) showed that many participants held a computer morally responsible for an error or program crash. These studies all focus on humans attributing actions and behaviors to a computer actant; their focus is on how a computer is viewed within a social context that has been created by a human.

Unfortunately, there is yet to be a discussion of the psychological impact upon the humans that are interacting with an emotionally intelligent³ tool. A computer as a social actor implies that a human will be treating it in a social manner; but it also implies that it will be treating the *human* in a social manner (or, at least, that is the perception on the part of the human). This reversal of the typical CASA study is somewhat convoluted (i.e., the human thinks that the computer is thinking something about the human), but it is important in the understanding of a computer as a fully capable actant. To be succinct, we may call it implied agency. This implied agency may also signify a quality of the treatment of computers that is different than traditional anthropomorphism.

If we are to take the media equation seriously as a step beyond anthropomorphism, then there is the understanding that people treat computers and other mediated experiences in the same social manner as they do when interacting with other humans. In short, we will be treating computers, at least subconsciously, as equivalent social actors. At the same time, these computers are mere appliances which we use—and I do not believe the media equation discounts

this fact. This collision begs two questions: first, will people be willing to accept the repercussions of a tool that is emotionally intelligent (i.e., will it hinder the use of the tool), and second, is it really appropriate in the traditional human-computer interaction—an interaction that is very much one-sided and dominating—for a tool to actively display human-like emotions? It is not a stretch to believe that the human participant may start to believe that he/she is actually dehumanizing the computer by treating it as a tool instead of an individual. The effects of this feeling could be expressed in various ways by the individual, such as no longer using the computer (in which case the incorporation of emotional intelligence did not benefit the interaction) or, on a more ethical level, as the feeling that he/she has participated in a relationship of dominance which could have lasting personal and social consequences. Some people may be uncomfortable in such situations and equate the interaction with unequal social relationships like bonded servitude, or worse, slavery. Given these potential troubles, there may be a desire for a new style of computer, one which interprets and reacts to the emotions of the human, but which does not outwardly emphasize its own emotionality. These “emotionally reticent computers” may be useful in situations that require the computer to be vigilant of the human’s emotions—like a user becoming frustrated when an error message keeps occurring and interrupting his/her work—but not distract the human from the task he/she is attempting to complete.

As troubling as the impact of these unequal social relationships may be, there is the equally disturbing possibility for the computer to be emotionally exploitative or treacherous. With the realization that humans treat computers—often unconsciously—as social actors, it may be tempting for researchers and designers to exploit these tendencies for personal or monetary gain. Perhaps the strongest danger of this occurring lies within CAPTology (computers as persuasive technology), where research is focused on creating computers that strongly influence

the thoughts and emotions of people. While this may obviously have positive benefits—such as enhancing the learning process and improving educational materials and equipment—designers and researchers must be acutely aware of the implications social computers can have, and must maintain a high standard of professional ethics.

These potential difficulties are not necessarily in the distant future, either. Reeves & Nass showed that even very basic and unconvincing computer actants elicited social responses from participants. It would be irresponsible of us to assume that, since no compelling humanoid robots exist, or because no computer has convincingly passed the Turing test (Turing, 1950), social interactions with computers will be limited to fleeting applications of unconscious social mores like politeness.

Conclusion

From its roots in the early eighties to its formalization in the nineties, the CASA framework has illuminated the somewhat counterintuitive notion that people treat computers in a social manner. A wealth of studies based on empirical studies in the social sciences have supported this conclusion and expanded it in unexpected directions, including computers as persuasive technology and emotional design. While many criticisms exist about the assumptions made in CASA research—such as the primary denigration that viewing computers as social actors is no different than the traditional anthropomorphism of things like animals, diseases, and automobiles—convincing and complete responses exist to counter them. In the philosophical arena, most research into computers as social actors can be construed as epistemologically based on social constructivism rather than empiricism, even though most of the studies heavily rely on empirical methodologies. Additionally, Buckland's (1991) theory of information-as-thing is

both applicable and inapplicable to CASA research, since computers are information systems that must deal with information in a quantifiable fashion, but emotions and social responses are intangible and difficult to quantify; the conversion process between information-as-knowledge and information-as-thing is particularly important for research in the CASA framework, since a substantial portion of the information about the social interaction between a human and a computer can be lost in the translation. When comparing CASA to the treatment of other artifacts as social actors, we see that the metaphorical aspect of considering that a computer is a social actor is categorically different than the metaphorical aspect of traditional anthropomorphism, because with computers the metaphor tends to dissolve completely (i.e., a human is not interacting with a computer *as if* it were a social actor; there is no impersonation involved like with the anthropomorphism of animals or other artifacts). Finally, there exist many ethical implications in further research and design based on the CASA framework, including the possibility for emotional treachery by computers on behalf of designers, as well as the likelihood for needing emotionally reticent computers that respond to human emotions but do not distract humans from using the computer as a tool. As a combined research methodology, the CASA framework can provide a novel direction for new research into the social impacts of information technology.

¹ I use Latour's (1987) actor-network terminology of "actant" here instead of the traditional "actor" to draw the discussion away from the intentionality of an actor (an admittedly difficult suggestion when one is talking about computers) and refocus it on the broad notion of agency and of the human perception of the actant's actions. Nass, Moon, Morkes, Kim, and Fogg (1997), while still using the traditional terminology, also narrowly define what it means to be a social actor: "something is a social actor to the extent that people respond to it as if it were a social actor" (p. 158).

² It should be noted that the label "CASA" is not a universally agreed upon term for research in this area; in fact, even the original developers of CASA (Reeves, Nass, et al.) do not strictly adhere to this terminology. Alternating labels, to be discussed further, include "the media equation," "ethopoeia," "computers are social actors," "affective computing," and even affective/emotional design. Most recently, CASA research has fallen under the general heading of "social responses to communication technology" at Stanford and "affective computing" at MIT.

³ The term "emotionally intelligent" here refers both to the ability for computers to interpret emotions from a human actant and to the ability for the computer to "display," in some fashion, emotions toward the human actant. Whether or not the computer actually feels or understands these emotions is not relevant at the moment.

References

- Aboulafia, A., & Bannon, L. J. (2004). Understanding affect in design: An outline conceptual framework. *Theoretical Issues in Ergonomics Science*, 5(1), 4-15.
- Aristotle. (1954). *The rhetoric and the poetics of Aristotle* (W.R. Roberts, Trans.). New York: Modern Library. (Original work written ca. 340 B.C.E.).
- Boorstin, J. (1990). *The Hollywood eye: What makes movies work*. New York: Cornelia & Michael Bessie Books.
- Brabant, S. (1982). The computer as a construct of social reality. *ACM SIGUCCS Newsletter*, 12(3), 12-17.
- Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science*, 42(5), 351-360.
- Cañamero, L. (2001). Emotions and adaptation in autonomous agents: A design perspective. *Cybernetics and Systems: An International Journal*, 32(1), 507-529.
- Csikszentmihalyi, M. (1992). *Flow: The psychology of happiness*. London: HarperCollins.
- Csikszentmihalyi, M., & Rochberg-Halton, E. (1981). *The meaning of things: Domestic symbols and the self*. New York: Cambridge University Press.
- Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.
- Descartes, R. (1989). *On the passions of the soul* (S. Voss, Trans.). Indianapolis, IN: Hackett. (Original work published 1694 C.E.).
- Fogg, B. J. (1998). Persuasive computers: Perspectives and research directions. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 225-232. Los Angeles: ACM Press/Addison-Wesley Publishing Co.
- Friedman, B. (1997). *Human values and the design of computer technology*. New York: CSLI Publications and Cambridge University Press.
- Friedman, B., & Millett, L. I. (1997). Reasoning about computers as moral agents: A research note. In B. Friedman (Ed.), *Human values and the design of computer technology* (pp. 201-206). New York: CSLI Publications and Cambridge University Press.
- Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.
- Google Scholar (n.d.). Retrieved November 28, 2005, from <http://scholar.google.com/>

Katagiri, Y., Nass, C., & Takeuchi, Y. (2001). Cross-cultural studies of the computers as social actors paradigm: The case of reciprocity. In M. J. Smith, G. Salvendy, D. Harris, & R. Koubek (Eds.), *Usability evaluation and interface design: Cognitive engineering, intelligent agents, and virtual reality* (pp. 1558-1562). Mahwah, NJ: Lawrence Erlbaum.

Khalid, H. M. (2004). Guest editorial: Conceptualizing affective human factors design. *Theoretical Issues in Ergonomics Science*, 5(1), 1-3.

Kitayama, S., Markus, H. R., & Kurokawa, M. (2000). Culture, emotion, and well-being: Good feelings in Japan and the United States. *Cognition & Emotion*, 14(1), 93-124.

Klein, J., Moon, Y., & Picard, R.W. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14(1), 119-140.

Krippendorff, K. (2004). Intrinsic motivation and human-centred design. *Theoretical Issues in Ergonomics Science*, 5(1), 43-72.

Latour, B. (1987). *Science in action*. Cambridge, MA: Harvard University Press.

Leontjev, A. N. (1978). *Activity, consciousness, and personality*. Englewood Cliffs, NJ: Prentice-Hall.

Lewis, M., & Haviland-Jones, J. M. (2004). *Handbook of emotions* (2nd ed.). New York: Guilford Press.

Lindsay, P. H., & Norman, D. A. (1977). *Human information processing* (2nd ed.). New York: Academic Press.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224-253.

Miller, C. (Ed.). (2004). Human-computer etiquette: Managing expectations with intentional agents [special section]. *Communications of the ACM*, 47(4), 31-61.

Mishra, P., Nicholson, M., & Wojcikiewicz, S. (2001-2003). Does my word processor have a personality? Topffer's law and educational technology. *Journal of Adolescent and Adult Literacy*, 44(7), 634-641.

Morkes, J., Kernal, H., & Nass, C. (1999). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of SCRT theory. *Human-Computer Interaction*, 14(4), 395-435.

Nass, C. I., Steuer, J., Tauber, E. R., & Reeder, H. (1993). Anthropomorphism, agency, and ethopoeia: Computers as social actors. *INTERACT '93 and CHI '93 conference companion on Human factors in computing systems* (pp. 111-112). Amsterdam, The Netherlands: ACM Press.

- Nass, C. I., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI conference on human factors in computing systems: Celebrating interdependence* (pp. 72-78). Boston: ACM Press.
- Nass, C. I., Moon, Y., Morkes, J., Kim, E., & Fogg, B. J. (1997). Computers are social actors: A review of current research. In B. Friedman (Ed.), *Human values and the design of computer technology* (pp. 137-162). New York: CSLI Publications and Cambridge University Press.
- Norman, D.A. (2004). *Emotional design: Why we love (or hate) everyday things*, New York: Basic Books.
- Norman, D. A. (1991). Cognitive artifacts. In J. Carroll (Ed.), *Designing interaction: Psychology at the human-computer interface* (pp. 17-38). Cambridge, UK: Cambridge University Press.
- Picard, R. W. (1997). *Affective computing*. Cambridge, MA: MIT Press.
- Picard, R. W., & Klein, J. (2002). Computers that recognize and respond to user emotion: Theoretical and practical implications. *Interacting with Computers*, 14(1), 141-169.
- Picard, R. W. (2002). What does it mean for a computer to “have” emotions? In R. Trappl, P. Petta, & S. Payr (Eds.), *Emotions in humans and artifacts* (pp. 213-235). Cambridge, MA: The MIT Press.
- Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. Chicago: The University of Chicago Press.
- Postrel, V. (2003). *The substance of style: How the rise of aesthetic value is remaking commerce, culture, and consciousness*, New York: HarperCollins.
- Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Stanford, CA: CSLI Publications.
- Trappl, R., Petta, P., & Payr, S. (Eds.). (2003). *Emotions in humans and artifacts*, Cambridge, MA: The MIT Press.
- Trappl, R., & Payr, S. (2003). Emotions: From brain research to computer game development. In R. Trappl, P. Petta, & S. Payr (Eds.), *Emotions in humans and artifacts* (pp. 1-10). Cambridge, MA: The MIT Press.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind: A Quarterly Review of Psychology and Philosophy*, 59(236), 433-460.
- Turkle, S. (1984). *The second self: Computers and the human spirit*. New York: Simon and Schuster.