Building Character for Artificial Conversational Agents: Ethos, Ethics, Believability, and Credibility

Sheryl Brahnam♣*

♣Missouri State University, Department of Computer Information Systems
(USA)

ABSTRACT

Because ethos is an unavoidable component of dialogue and forms the basis for believing and being persuaded by another's speech, it is an important topic for AI researchers. This paper examines the concept of ethos, especially Aristotle's notions of situated and invented ethos, as it functions in oral and written discourse and then explores what happens to ethos in computer-mediated human-to-human and human-to-machine discourse. The paper draws a number of conclusions that may be of value to researchers in these fields. In particular, it argues that the rhetorical concept of ethos furnishes a broader theoretical framework for understanding design and ethical issues involved in agent credibility than does the artistic notion of believability. The paper concludes by suggesting some nonartistic methods for making agents more credible within the framework of situated ethos.

1. Introduction

> The ethos reveals itself in time, and the revelation brings disaster.
>
> Aeschylus

Most computer interfaces use human language to communicate with users, and this fact has consequences that I feel have not been explicitly drawn out but that I think are at the core of current design concerns and ethical dilemmas in the development of more loquacious interfaces, such as conversational agents. One consequence is that

---

* Corresponding Author
S. Brahnam
Missouri State University, CIS Department, 387 Glass Hall, Springfield MO, 65804, USA
E-mail: sbrahnam @ facescience . org

computer communications, because they are acts of language, cannot be divorced from the ethics of rhetoric.

There are many definitions of rhetoric. These definitions are often marked by an ethical antithesis, with theorists lining themselves up more or less on one side or the other. A pejorative sense of the term, often voiced by Socrates, is bound up with the notions of appearance and artifice. Rhetoric in this regard is artificial, showy, and decorative. It is the fluff of discourse, a bag of tricks. These ideas are reflected when we use the word dismissively. An argument is "mere rhetoric," for example, when we feel it lacks substance. For Socrates rhetoric is inherently duplicitous and morally suspect because too often it is rived from reason and truth.

Other theorists are far more positive in their appraisal of rhetoric. Isocrates, for example, holds rhetoric in the highest esteem since it enables people to persuade one another and to clarify their desires and needs. Without rhetoric there would be no civilization. In his famous speech, "Antidosis," he claims that it is because of rhetoric that we have cities, laws, and the arts. For Isocrates, " . . . there is no institution by man which the power of speech has not helped us to establish" (p. 28).

Parallel to rhetoric's dubious connection to truth and reality is another source of ethical anxiety, one that I believe is at the very heart of many issues involving artificial conversational agents. The Greek derivative of the word *rhetoric* is *eiro* meaning "I say." In this sense, rhetoric is concerned with the relationship of discourse (what is said) to the character of the speaker (the *I*), or what the Greeks called *ethos*, a term often treated as a synonym for *character*, *reputation*, and *persona* in the classical literature and for *subject* and *self* in more modern discourse.

In the Western world, ethos, or the relation of a speaker to her speech, has proven problematic. The etymology of the word discloses a number of binary oppositions that are at the core of many debates: the domestic versus the wild (in reference to animals and the incorrigible vileness of human beings), the cultivated versus the innate, the public versus the private, and the singular versus the plural (see Chamberlain, 1984). In each of these oppositions, ethos is bound up with some sense of place (Miller, 1974), but where ethos dwells (inside an individual soul or within some public sphere) and whether it is reachable (informed by habit and custom or innate and incorrigible) are all fiercely contested.

In discourse, these uncertainties throw into question the very ground and source of a speaker's credibility. If ethos is embedded in the rhetor's reputation, for example, then it is developed slowly and painstakingly through habit and virtuous action. Moreover, it

follows that ethos is essentially incarnate and cannot be crafted since it is part of the moral fiber of the speaker. If ethos is a linguistic construction, however, then it is artificial and imitative, an artistic invention. What matters is that the words of the speaker make him *appear* credible and trustworthy. This opens the door to dissimulation. A speaker can learn to construct a convincing ethos, one that is suitable for various occasions, for instance, by taking courses in speech making or by hiring a speechwriter.

Because ethos is an unavoidable component of dialogue and forms the basis for believing and for being persuaded by another's speech, it is an important topic for AI researchers. Whether that speech comes from a human being or from a computer program, communication depends on the credibility of the speaker; otherwise what is being said is subject to mistrust and doubt, creating a situation that seriously jeopardizes the willingness of the speaking partners to engage in cooperative activities.

The debate whether ethos is a linguistic phenomenon or whether it is reflective of a human being's character is especially relevant in the development of humanlike interfaces. Most of these interfaces attempt to imitate the essential behaviours involved in face-to-face communication, with success measured against the artistic standard of believability (Loyall & Bates, 1997). *Believability* in the media arts refers to the human tendency to suspend disbelief in order to enjoy the portrayal of a separate imaginary being. This concept has driven developers to concentrate on strengthening the natural human tendency to anthropomorphize. Against the broader rhetorical notion of ethos, however, believability is seen as a limiting concept because it already plants itself squarely on one side of the ethical divide: it assumes that credibility, or ethos, is fictive and that it can be detached from human agency and exist on its own.

That ethos can be entirely removed from human *being* is questionable, as I hope to demonstrate by examining ethos in the light of what happens to it when the human speaker is progressively removed from discourse. This examination will also serve as a brief introduction to the subject of ethos. In section 2, I look at the Greek origins of the concept as it is defined mostly in terms of oral discourse. In particular, I look at Aristotle's key notions of situated and invented ethos, especially as they relate to ethics. In section 3, I present some of the problems with ethos in written discourse, showing how the voice of the author loses its claim on ethos as it slides to the living reader's interpretation of the text. In section 4, I consider ethos in Computer-mediated discourse, noting how in the age of secondary orality the notion of invented ethos

expands to include the visual as well as the linguistic. I also show how the disassociation between the speaker's words and her body in textual exchanges renders trust and credibility problematic. Finally, in section 5, I arrive at the main purpose of this paper and explore what happens to credibility when a human speaker is entirely removed from the words that are being spoken, as is the case in computer-generated discourse. In this section, I note that the ethical divide in rhetoric is reflective of a deeper conflict in the Western self between the imaginal and the real, a conflict made palpable when people are asked to communicate with a computer using a natural language interface. Just as there is the Eliza effect (Weizenbaum, 1976), or the tendency for people to anthropomorphize and ascribe greater human capabilities to computers then they actually possess, there is also a corresponding Weizenbaum effect, or a tendency to debunk the anthropomorphic. In their interactions with humanlike interfaces, people are torn between the two. As Aeschylus warned in the epigraph, when the true nature of a creature's ethos asserts itself (in this case revealing the agent to be a machine), disaster invariably follows. In section 6, I conclude the paper by showing how the richer concept of ethos, especially situated ethos, might foster the discovery of nonartistic methods for overcoming some of these disasters in credibility.

2. Ethos in Oral Discourse

As Welch (1990) points out, classical rhetoric continues to inspire and to inform because it covers all possible uses of language. She writes, "While at various historical times it emphasizes some kinds of speaking and writing over other ones . . . classical rhetoric nonetheless self-consciously concerns itself with all manifestations of discourse" (p. 5). Since language issues from a speaker, ethos, or the character of the speaker, is a central topic of discussion, but as Farrell (1993) observes, it is also "one of the most enigmatic concepts in the entire lexicon" (p. 80).

In classical rhetoric, ethos is one of three modes of *pisteis*, or means of persuasion, with the other two being *logos* and *pathos*. Each of these modes centers on a fundamental aspect of oral communication (Wisse, 1989). Logos is focused on the words of the speech, with logical proofs and arguments. Pathos is centered on the audience, its emotions and reactions. Ethos refers to the speaker, his reputation and presentation of character. Thus, persuasion is not solely concerned with developing

sound arguments and proofs; it also involves putting the audience into a receptive frame of mind and convincing them that the speaker is a credible person.

What makes a speaker worthy of credence? In the beginning of the second book of the *Rhetoric*, Aristotle states, "There are three things which inspire confidence . . . good sense [*phronesis*], excellence (virtue) [*arete*], and goodwill [*eunoia*]" (1378a5). Good sense is concerned with practical intelligence, expertise, and appropriate speech. A speaker who demonstrates knowledge of a subject and exhibits a sense of propriety and aptness of language is believed (1408a20). Excellence refers to whatever socially sanctioned virtues good people are expected to possess. Goodwill conveys the impression that the speaker has the welfare of the audience in mind. Aristotle associates goodwill with friendship and friendly feelings. He writes, "We may describe friendly feeling towards anyone as wishing for him what you believe to be good things, not for your own sake but for his, and being inclined, so far as you can, to bring these things about" (1380b35). Wisse (1989) points out that Aristotle's categories of credibility include the rational (good sense), the emotional (goodwill), and the trustworthy (excellence or virtue).

If a speaker is deemed credible, then the audience will form a second order judgment that the arguments put forth by the speaker are true and acceptable. The persuasive impact of ethos, therefore, should not be underestimated. According to Aristotle, the speaker's "character may almost be called the most effective means of persuasion he possesses" (1356a14). It provides what has come to be called an *ethical proof*.

Aristotle recognizes two kinds of ethical proof: invented and situated. If a speaker is fortunate enough to enjoy a good reputation in the community, she may rely on that as an ethical proof. This is *situated ethos*. Although not the focus of Aristotle's discussion of ethos, this is the course that is advocated by Isocrates, Cicero, and Quintilian, among others. For these writers, good character is more important than clever speech. As Isocrates argues in the *Antidosis*, "the man who wishes to persuade people will not be negligent as to the matter of character; no, on the contrary he will apply himself above all to establish a most honorable name among his fellow-citizens" (p. 339). Reputation is a method for engendering belief because the way a speaker lives offers the best evidence of the truth and goodness of that person's words (Welch, 1990).

Aristotle, in contrast, is not as interested in cultivating the character of the rhetor as in discovering how a speaker can persuade an audience to trust what he has to say. Aristotle appears to be turning the tables here by arguing that the persuasive power of ethos is more in a speaker's words than in his reputation. He writes, "This kind of

persuasion, like the others, should be achieved by what the speaker says, not by what people think of his character before he begins to speak" (1356a10). The dramatization of character within discourse is *invented ethos*.

Many have tied Aristotle's notion of invented ethos to his descriptive procession, also found in the *Rhetoric*, of a host of character types. Rhetoricians, both ancient and modern, have used these and similar sketches as student exercises in *ethopoeia*, or the art of fabricating character in language. Others believe the portraits are intended to help rhetors recognize the different psychological types in their audience so that the speaker can better carry out Socrates' injunction in the *Phaedrus* to pair types of speech to types of souls.

Exactly what Aristotle is advocating in his discussion of situated and invented ethos, however, is debatable. According to Yoos (1979), Aristotle is simply showing how an audience's impressions of a speaker's character can be altered by the language used by the speaker. The rhetor is then faced with a choice: he can distort the audience's perceptions of his character in his speech or he can go about developing "rhetorically effective personal qualities" (p. 44) by becoming a good person. In a similar vein, Wisse (1989) notes that Aristotle is clearly stating that telling the truth is central to ethos and that invented ethos must be based on the speaker's true character.

For many others Aristotle's concept of ethical proof is anything but ethical, primarily because it obscures a distinction between *real* character, which is cultivated through good habits, and an artistically produced character, which invites pretense and dissembling. Although Gill (1982) argues that it is actually those following Aristotle who extend the idea of ethos to the point of making it an invention (a persona, or mask, that a speaker assumes for the duration of a speech), Aristotle's characterization of ethos seems to have paved the way for treating ethos as artifice by suggesting that what really matters is not so much the quality of the innate character of the speaker but the perceptions of the audience, and these can be manipulated.

A careful examination of invented and situated ethos, however, shows that in both the judgment of the audience is crucial. Invented ethos is bound to a single instance of speaking in public and involves the immediate revelation of character, a momentary portrayal that may or may not honestly represent the speaker's true character and that may be intended more as the object of sport or as a rhetorical exercise (in ethopoeia, for instance).

Situated ethos, in contrast, is based on repeated exposure to the public and on building trust over time by consistently demonstrating the qualities of good sense,

excellence, and goodwill. This recalls the meaning of the ancient Greek word *ethos* as "habitual gathering place." As Holloran (1982) notes, "I suspect that it is upon this image of people gathering together in a public place, sharing experiences and ideas, that its meaning as character rests" (p. 60). His views are echoed by Reynolds (1993), who claims that the classical sense of ethos is not wrapped around the individual (it is not singular), but rather it refers to the surrounding social context (it is plural). In this regard, situated ethos can be thought of as a longterm relationship that develops in the exchange of ideas between an individual and the other members of her community.

3. Ethos in Written Discourse

The most obvious difference between writing and oral discourse is the loss of the bodily presence of the speaker in writing. Without that presence, the question arises whether writing has any claim to credibility, and, if so, then where does it reside?

A view that follows from the Aristotelian notion of invented ethos and that held for nearly 2500 years is that ethos is in the text. An author's character is revealed in his writing style, whether by design or by default, and a convincing writer is one who manages to exhibit the key ingredients of credibility: good sense, excellence, and goodwill.

The debate whether credibility is contingent upon a genuine personality or whether it can be feigned resurfaces in writing in new guises. While some writing manuals encourage students to find their unique voices, others offer exercises intended to teach students methods for selecting and rendering personas that best suit their writing project. For example, Hunt (1991) in the *Riverside Guide to Writing*, has students jot down the personality qualities of their intended personas, somewhat in the fashion of Aristotle's character portraits.

In contrast to Aristotle's ethos, which accommodates the separation of speech from the bodily presence of the speaker, is Plato's "ethos,"[1] which according to Baumlin (1994), is "the space where language and truth meet or are made incarnate within the individual" (p. xiii). For Plato, body and speech cannot be sundered. In the *Phaedrus*, Plato has Socrates argue that writing has no ethos precisely because the speaker's body has been removed:

---

[1] Plato never uses the term *ethos* but, as Baumlin (1994), among others, note, his concept of ethos, or the relation of human character to language, can be inferred.

You know, Phaedrus, writing shares a strange feature with painting. The offsprings of painting stand there as if they are alive, but if anyone asks them anything, they remain most solemnly silent. The same is true of written words. You'd think they were speaking as if they had some understanding, but if you question anything that has been said because you want to learn more, it continues to signify just that very same thing forever. When it has once been written down, every discourse roams about everywhere, reaching indiscriminately those with understanding no less than those who have no business with it, and it doesn't know to whom it should speak and to whom it should not. And when it is faulted and attacked unfairly, it always needs its father's support; alone it can neither defend itself nor come to its own support. (275d-e).

In this passage, Socrates is, in effect, charging writing with violating the dictums of good sense, excellence, and goodwill. The brunt of Socrates' charges fall on writing's lack of good sense: it is incapable of defending itself when attacked; it repeats the same words endlessly without variation; it fails to field questions; and, because it lacks perception, it is unable to adjust the language and its style of delivery to accommodate the psychological differences in the people it encounters. These last two faults suggest a lack of goodwill. After all, how would it be possible for writing to have the reader's welfare in mind if it is not willing to interact? Finally, writing is not virtuous because it deceives. Like a painting, it is an image pretending to be what it is not: living speech.

In general, postmodern thought would also claim that ethos is not in the text. If ethos exists anywhere, it is in the minds of the readers, in their interpretations, constructions, and projections upon the text. Readers recast the authors into their own images. As Corder (1989) explains, "the author is dead and language writes us, rather than the other way around, and interpretation prevails rather than authorship" (p. 301). Corder goes on to say that in writing, "Language is orphaned from its speaker; what we once thought was happening has been disrupted. Authors, first distanced, now fade away into nothing. Not even ghosts, they are projections cast by readers" (p. 301)[2].

---

[2] Writers also often acknowledge the fact that once a text is printed it no longer belongs to them. John Steinbeck, near the end of writing one of his novels, expressed a profound sense of this loss when he wrote, "In a short time [the novel] will be done and then it will not be mine any more. Other people will take it over and own it and it will drift away from me as though I had never been a part of it. I dread that time because one can never pull it back, it's like shouting good-bye to someone going off in a bus and no one can hear because of the roar of the motor" (Plimpton, 1977, p. 199).

How is ethos a creation of the reader rather than of the writer? According to Baumlin and Baumlin (1994), ethos is the "projection of authority and trustworthiness onto the speaker, a projection that is triggered or elicited by the speaker but that is otherwise supplied by the audience" (p. 99). It is the formation of a transference, an unconscious displacement of feelings for one person to another. It is because transferences are attached to the images of people, as Derrida (1987) points out, and not the people themselves, that they can also be placed on authors and texts.

Brooke (1987) claims that persuasion would not be possible at all without positive transference. He points out that a positive transference, according to Lacanian theory, is the projection of the idealizing super status of the *One Who Knows*. This is a person who is believed to know the deeper truth about someone and who embodies that person's ideal self.

Baumlin and Baumlin (1994) provide an interesting explanation of ethos as a transference formation by mapping Aristotle's three modes of persuasion (pathos, logos, and ethos) to Freud's psychological model of the id, ego, and superego. According to Freudian theory, the ego of the infant develops through the interplay of the pleasure principle and the reality principle. When the basic drives of the infant are frustrated, the baby alleviates its frustrations by fantasizing and hallucinating satisfaction. However, the inadequacy of hallucinations to provide relief eventually leads the infant to distinguish fantasy from reality. The ego emerges out of a need not only to satisfy the instinctual drives of the id but also to control, to sublimate, and to defer them in socially acceptable ways. The ego mediates between the id and the superego, which stands opposed to the id. Largely unconscious, the superego harbors images of an ideal self and strives after spiritual goals, but it also punishes the ego with guilt feelings when the ego fails to live up to these ideals or allows the id to transgress societal mores.

From this description it is easy to see how logos, the appeal to reason, relates to the reality-testing mechanisms of the ego, how pathos, the appeal to emotion, is connected with the id's desire for pleasure and avoidance of pain, and how ethos, the appeal to trust, mirrors the superego's striving after the ideal. From the psychological perspective, ethos, as a positive transference, is the unconscious process "of identification that lead[s] children to obey their parents and lead[s] mature audiences to believe the speakers to whom they have given their trust" (Baumlin & Baumlin, 1994, p. 100).

In writing this process is often amplified. According to Olson (1980), separating the speaker from his speech endows the text for most readers with a numinous or vatic quality. The source of the text becomes transcendental and above reproach. Olson remarks that "it may be because children assume that textbooks have great authority that they are willing to devote serious and prolonged study to books, rather than simply reading them" (p. 193). What confers authority on texts are the opinions of academics, who are themselves writers. Thus, there is a status differential between readers and writers, just as there is between children and their parents, that promotes transference. In contrast, writers within their peer groups feel privileged to exchange ideas and to offer criticisms. Olson goes on to claim that an author, by inviting criticism from his peers, is able to establish a reputation, and this "reunites the author with his writings" (p. 194).

Writing then, somewhat like an analyst, elicits transference in the reader. However, as Freud (1912) first noted, there are both positive and negative transferences. Whereas the positive transference is characterized by idealization and admiration, the negative transference involves the eruption of hostile feelings. Within face-to-face discourse, both negative and positive transferences can be contained and worked through. Breaks in the transference reintroduce the reality principle and furnish the ego with an opportunity to grow.

Writing, unlike an analyst, is blind and insensate. Incapable of perception and interaction, as Socrates charges, it cannot help readers overcome their transferences. What Socrates discerns in writing and condemns is the missing face of the speaker: the presence of a living human being capable of circumventing dangerous misinterpretations and of taking on the responsibility of adapting his words and his responses in such a manner that his intentions and message have the best possible chance of being understood by his listeners.

With the death of the author, as announced by postmodernism, no one is left to direct the reader's understanding. Moreover, all voices of authority are suspect. As many Marxist and poststructural theorists point out, these voices are often exclusionary and promote the agendas of the current political, cultural, sexual, and religious hegemonies. What is advocated is a democracy of texts, where not only is one reader's commentary on a text as good as those written by any other (each reader offers a different perspective) but the reader's views are also as good as the opinions proffered by the author herself (see, for instance, Welch, 1990, p. 163).

If it is true that ethos in writing is bound up with a reader's transferences, then what readers end up finding within texts are unconscious projected images of themselves. As Jung (1959) writes, "projections change the world into the replica of one's own unknown face" (p. 8). Staring into the face of writing then is much like gazing into the face of Rorschach, the comic book character whose facial features are ever changing ink blocks. Both hero and avenging antihero, the voice of the writer reveals itself to be nothing more than the stirring whisperings of the reader's own superego ideals and conscience.

4. Ethos in Computer-Mediated Communication

Welch's (1990) contention that classical rhetoric concerns itself with all manifestations of discourse" (p. 5) is put to the test when considered in the light of the technological advances in communication that have taken place since the invention of moving pictures in 1867. It has been claimed by those who ascribe to the orality-literacy thesis, in particular Havelock (1982) and Ong (1982), that the Hellenistic age is similar to our own. Just as the electronic revolution today is changing the ways in which we transmit information, the radical new technology of writing revolutionized the way Hellenistic Greece communicated and stored cultural knowledge. Starting early in the fourth century BCE with the development of Greek vowels (which closely represented the sounds of speech, making it much easier to teach reading and writing), Western culture rapidly moved from a culture that was primarily oral to one where literacy predominated. The claim is now being made that we are moving away from literacy to a secondary orality as we increasingly rely on modern technologies to communicate with each other using the ordinary language of everyday speech. "Secondary orality represents a 'cultural recall' of primary orality," Welch explains, "because the emphasis of speaking and hearing takes on new meaning with the invention of electronic forms of communication" (p. 136).

The Western move to literacy, as noted in section 3, intensified philosophical debates among the Greeks concerning the relationship of a speaker's character to his words. From Aristotle onwards, invented ethos increasingly gains currency as ethos is conceptualized more as a plastic property of language than as a static attribute of a person. In the Western move to secondary orality, invented ethos expands to include the visual as well as the linguistic. Just as writing highlights ethos as it is echoed in a

person's choice of words, so electronic media magnify ethos as it is reflected in one's physiognomy, demeanor, gestures, facial expressions, eye movements, and clothing, making what was once an unconscious expression of self (Goffman's (1959) notion of a self-presentation that is *given off*) more conscious and intentionally *given*. Whereas, in the age of literacy, the notion of invented ethos made people more aware of how they composed themselves linguistically, now in the age of secondary orality, electronic forms of communication are making people more conscious of the way they present themselves physically.

In face-to-face communication, nonverbal cues provide people with a rich source of information about a speaker's character. Sometimes these nonverbal cues are intentionally given to facilitate communication (as when pointing or managing conversational turn-taking), but more often they are given off. Many studies show that people are particularly attuned to nonverbal cues; even brief and degraded exposures produce surprisingly accurate judgments in people regarding a person's level of intelligence, competence, and personality traits (Albright, Kenny, & Malloy, 1988; Ambady, Bernieri, & Richeson, 2000). Nonverbal signals also provide face-to-face speakers with a means of gauging the effectiveness of their given expressions, as they are often mirrored in their audiences.

Many forms of electronic communication, such as film and television, open channels that allow receivers to evaluate the nonverbal cues presented by speakers. Politicians and other people who make extensive use of the media to broadcast messages to the masses studiously craft their self-presentations to maximize their perceived credibility. Oftentimes, simply making an appearance in the media confers authority (similar to the way books take on a vatic quality). Many studies have been conducted that explore how the public's opinion of people can be shaped by the media. McGinniss's 1969 classic, The Selling *of the President*, for instance, was one of the first to reveal the full extent of the media's power to rebrand presidential candidates and sway voter opinions.

Perhaps as important as the presentation of nonverbal cues in film and television are the conscious and unconscious effects produced by camera position, lighting, scene composition, and the accidental introduction of such media artifacts, as shifts in aspect ratio and the misalignment of audio and visual channels (see, for instance, Beverly & Young, 1978; McCain & Wakshlag, 1974; Reeves & Nass, 1996, p. 212; Tiemens, 1970). Vertical camera angle, in particular, has been shown to influence impressions of credibility and dominance (Tiemens, 1970). As Balabanian (1981) succinctly puts it,

"High shots produce pygmies. Low shots yield monoliths of the Citizen Kane type" (p. 27). In situations where credibility is crucial, such as in videotaped interrogations, direct manipulation of camera angle can have devastating consequences (see, for instance, Hemsley & Doob, 1976; Lassiter, 2002; Locke, 2009). Even changes in the vertical position of the viewing screen or monitor have been observed to influence receiver perceptions of credibility (Huang, Olson, & Olson, 2002).

Comparatively little research has been conducted that explores effects of channel affordances and interactivity on self-presentation using visual modes of computer-mediated communication (CMC), such as videoconferencing (for reviews, see Walther & Parks, 2002; Whittaker, 2003). This technology is still in its infancy. Compared with face-to-face communication, interaction over video is somewhat leaner in its reduction of proxemic and environmental cues, and it often suffers from delays (Burgoon et al., 2002; Whittaker, 2003). Many people report a level of discomfort using audiovisual CMC, perhaps because they are uncertain how to present themselves when using these technologies (see, for instance, Chapman, Uggerslev, & Webster, 2003). With many stores, such as Kinkos, now installing teleconferencing stations, it can be predicted that this technology will gain in popularity, and people will eventually attempt to exploit the unique characteristics it offers to produce desirable self-presentations.

As is the case with television and film, scene composition, lighting, camera angle, and various artifacts have an impact on speaker credibility in audiovisual CMC. It is common in many videoconference settings, for example, for one or more cameras with wide angle lenses to be positioned at a height. This reduces credibility and makes people look, as mentioned above, like pygmies. Many researchers are attempting to correct some of the more undesirable mediation artifacts these technologies inadvertently introduce, but much more research needs to be done in this area. Liu and Cohen (2005), for instance, have proposed a system for gradually increasing head size so that speakers who sit at the far end of large conference tables can more easily be seen. They claim they can do this "without causing undue distortion" (p. 1), but it is not known whether these corrections may themselves introduce artifacts that influence user credibility. Other researchers are exploring methods for exploiting technology in order to enhance speaker credibility. Noting that facial similarity has been shown to increase trust (DeBruine, 2002), for instance, (Bailenson, Garland, Iyengar, & Yee, (2006) have worked on systems that manipulate voter intentions by morphing candidate faces more towards the faces of the voters they are targeting.

Not all CMC provides audiovisual channels. The most common modality in CMC is textual (for example, email, chat, and message boards). This form of communication is unique: like writing, it is disembodied, yet, like speech, it is a dynamic verbal exchange between partners. Text-based CMC possesses features, such as, immediacy, nonlinearity, and a "changing evanescent character" that are reflective of oral communication (Ferris & Montgomery, 1996, p. 57). Textual exchanges also retain an oral fluidity that is missing in writing. Some desirable characteristics of writing that are preserved include nonverbal cue filtration (making it easier to mask affect) and the ability to edit messages.

For some people, the loss of nonverbal cues in textual exchanges makes impression management problematic, as self-presentation is limited to what can be conveyed through language and typography (for a review, see Walther & Parks, 2002). As Wallace (1999) writes, "[it's] like navigating white water with two-by-fours for oars. Your impressions management toolkit is strangely devoid of the tools most familiar to you" (p. 28). With synchronous CMC there is little time to self-censor. However, with asynchronous CMC, individuals have the opportunity to reflect on how best to transmit impressions of themselves using more favorable verbal cues (Walther, 1992). According to Walther's (2007) hyperpersonal model of CMC, people eventually learn to exploit the affordances offered by new media in order to enhance their self-presentations.

Because the performativity of speech in text-based CMC is no longer constrained by the body, the presentation of self is more fluid and unstable (see Poster, 2001). The disassociation of the body in textual exchanges presents people with unique opportunities that go beyond facilitating linguistic impression management. In many online communities, for instance, individuals are tempted to impersonate others and invent entirely new selves. Concerns with online identity deception are common (Donath, 1999; Van Gelder, 1985). People want to know for certain that the doctor offering advice online has a medical degree and that the woman a man is flirting with is actually female. Other people feel less threatened by the possibility of deceit and appreciate instead the freedom CMC grants them to explore themselves through role playing. It could be argued that online environments are providing modern communicators with a whole new playing field for exercises in *ethopoeia*, since within these online communities, as Poster (2001) notes, "Each individual is a character and participation is successful to the extent that the character is believable by others" (p. 75).

Despite the low level of reliability, most people tend to accept the conventional signals offered them in CMC. Some theorize that the human cognitive system is biased towards accepting statements at face value (Gilbert, 1991). Certainly, the cost of assessing each message and challenging those that are questionable is high. People must have ample time to reflect, the mental capacity to judge, and sufficient exposure to suspect deceit. In addition, they must have the courage to confront people they suspect are dissembling, an action that could lead to social embarrassment for both parties (Boyd, 2002).

As in written discourse, credibility in text-based CMC ultimately depends on how receivers, or readers, interpret messages. The loss of nonverbal cues calls upon the reader's imagination to fill in the gaps. "Reading another person's message," Suler (2004) observes, "might be experienced as a voice within one's head, as if that person magically has been inserted or 'interjected' into one's psyche. Of course, we may not know what the other person's voice actually sounds like, so in our head we assign a voice to that companion. In fact, consciously or unconsciously, we may even assign a visual image to what we think that person looks like and how that person behaves" (p. 323). What Suler is saying, in effect, is that in the act of reading messages, the character of a sender is formed in the reader's mind much like a character in a novel: it is mostly the product of the reader's imagination.

Because textual exchanges are read and reading takes place within the mind, environments where textual exchanges take place are often experienced as *transition spaces* (Suler, 2009). In object relations theory, a transition space is an intermediate zone where others are viewed as part self and part other. As Suler writes regarding online communication, "The online companion now becomes a character within our intrapsychic world, a character that is shaped partly by how the person actually presents him or herself via text communication, but also by our expectations, wishes, and needs." In other words, the boundaries between self, body, and other people are loosen in text-based CMC, making it easier for people to develop both positive and negative transferences.

Heightened levels of intimacy, liking, and solidarity (or *hyperpersonal communication*) have been observed anecdotally and in several empirical studies, and they have been reported to occur both in recreational settings, such as in online chatrooms, and in business settings (Walther, 1996). One common form of receiver hyperpersonal communication is idealized perception. Most people tend to inflate their partner's attributes, overrating their partners, for example, in intelligence and attractiveness. As

there is less to go on, what is missing is amplified, resulting in an overreliance on minimal cues (Walther, 1996).

An overreliance on minimal cues has detrimental as well as favorable effects. A misspelling in a message, for example, can be taken to indicate far more about a person's intelligence than the mistake warrants. The overreliance on minimal cues can also strengthen stereotyping. Several studies have suggested that negative stereotyping is more pronounced when information about a person is ambiguous (Hilton & Fein, 1989). When nonverbal cues are removed, as in email, for instance, the effects of stereotyping increase. They can also become contagious. In one study it was shown that sharing information in email discussions about job candidates resulted in a propagation and an intensification of stereotyping (Epley & Kruger, 2005).

Although there is some evidence that stereotyping increases with experience, extended interactions with people are more commonly thought to weaken initial stereotypes. In general, the more experience a person has with someone, the more individual that person becomes (Higgins & King, 1981). Through repeated interactions, people are capable of forming more accurate impressions of their online companions (Walther, 1996). An interesting example of this involves a man impersonating a lesbian online. He managed to develop several erotic relationships with women but was eventually discovered because of some verbal inconsistencies.

An alternative theory advanced to explain how communication is affected by CMC sees communication as influenced less by channel reduction than by a constraint in time. In other words, the difference between face-to-face communication and text-based CMC has more to do with the rate social information is exchanged than with the amount that is transmitted (Walther, 1996; Walther, Anderson, & Park, 1994). Since CMC travels slower than oral speech, it takes longer for people to decipher the expressive cues.

It may take more time, but some individuals manage to establish reputations through textual mediation, and people who deceive others, as noted above, are often exposed. In these instances, the ancient Greek meaning of *ethos* as "habitual gathering place" has meaning for CMC. As discussed in section 2, situated ethos is based not only on bodily presence but also on repeated exposure to the public and on building trust over time by consistently demonstrating the qualities of good sense, excellence, and goodwill. Even though textual communications are slow and missing important nonverbal cues, virtual communities may provide a gathering place where ethos can establish itself. Donath (1999), for example, shows that a person's reputation in user's

groups depends on frequent and long-term exposure to the group. How reputation is developed also lines up well with Aristotle's categories of credibility. According to Donath, a reputation is built by answering questions (showing goodwill), by providing intelligent and interesting comments (demonstrating good sense and appropriate speech), and by quelling arguments, deferring judgments, and signing posts rather than remaining anonymous (indicating virtue and excellence).

However, many textual encounters, especially those online, are fleeting and oftentimes anonymous. In user groups, forums, chatrooms, and email, people are tempted to let go of the ego's restraint on the id. Suler (2004) calls this relaxing of the reality principle the *online disinhibition effect*. Bound up with positive transference is *benign disinhibition*, which opens people up, making them more personal and willing to share and give of themselves. Bound up with negative transference is *negative disinhibition*, which allows people to act in ways they never would face to face: they bully, cheat, lie, mock, and flame. A news reporter, for example, after writing a friendly piece on Bill Gates reported being shocked to receive this email flame from a fellow technical writer for a major newspaper: "Crave this, asshole: Listen, you toadying dipshit scumbag . . . remove your head from your rectum long enough to look around and notice that real reporters don't fawn over their subjects, pretend that their subjects are making some sort of special contact with them, or, worse, curry favor by TELLING their subjects how great the ass-licking profile is going to turn out and then brag in print about doing it" (Seabrook, 1994, p. 70). In some cases, the disinhibition effect propels people into the internet underworld. Suler calls these darker expressions of the id *toxic disinhibition*.

It is interesting to note in this regard that an early meaning of the ancient Greek word *ethos* referred to the incorrigible baseness of human nature, which tends to reassert itself despite the niceties of social convention (Chamberlain, 1984)[3]. When a person's character is created and discarded with abandon and when few cues are given to inform a truer image of another person's character, people easily fool themselves into believing that their unethical behaviors are not representative of their true selves but rather fictions that live in a fantasy world. And in the world of fantasy, no one can be harmed. As Suler (2004) notes, "the person can avert responsibility . . . almost as if superego restrictions and moral cognitive processes have been temporarily suspended

---

[3] This reflects the Greek debate between ethos as something innate versus cultivated. It is interesting that the earliest uses of the word *ethos* refer to the unexpected unleashing of the wild nature of domesticated animals and then later, by extension, to the revelation of the incorrigible vileness of human nature. In these earlier usages, ethos is more akin to the idea of the Id than to the superego.

from the online psyche" (p. 322). With the virtual self there is nothing but a chaotic blur on which to pin responsibility.

For Heim (1993), the "nonrepresentable face" is the primal source of responsibility. Without the human face, especially the eyes, ethical awareness shrinks. Ethos, the whisperings of consciousness, disappears, and disinhibition runs rampant. Heim maintains that, "The physical eyes are the windows that establish the neighborhood of trust" (p. 102). For him the glowing computer screen, the window into the virtual, eradicates that trust.

5. Ethos in Computer-Generated Discourse

In the introduction, I noted how rhetoric is marked by an ethical antithesis. On the one hand, there are those who mistrust rhetoric, viewing it at best as decorative and at worst as a form of deception. On the other hand, there are those who appreciate it, maintaining that is it an essential component of community life since people are often called upon to persuade one another. Richard Lanham (1976) nicely portrays the Janus face of rhetoric in his treatment of the subject by personifying these two opposing views as *homo seriousus*, the loyal servant of logos, or the rational and the real, and as *homo rhetoricus*, the sensualist who relishes in appearances, artifice, and all things playful and novel. For Lanham, the Western self "has from the beginning been composed of a shifting and perpetually uneasy combination" of the two (p. 6).

As we have seen, Aristotle's notion of ethos strides these two opposing views. Situated ethos is the servant of the real because it is grounded in reputation and in a community's opinion of an individual that is formed over the course of time. In contrast, invented ethos is more playful since, as a product of language, it is concerned more with the artistic creation of character.

The clash between homo seriousus and homo rhetoricus is apparent as well in the two opposing philosophies of interface design that view computers as either tools for cognition, where the interface is directly manipulated, as is the case with desktop computing, or as mediums for social and emotional expression, where the interface is modeled after face-to-face communication, as is the case with conversational agents.[4] The development of humanlike interfaces has been fueled in large part by the need to

---

[4] It should be noted that Miller (2004) contrasts the ethos of expert systems with the ethos of intelligent agents. She thinks the latter is based more on the rational notion of good sense while the former she connects more with goodwill and the Ciceronian ethos of sympathy. She claims that the two need to be balanced by virtue.

make human-computer communication easier and more intuitive for a wider range of users. Many people find spoken conversation with computers a more comfortable and efficient means of interacting with computers (Picard, 1997). However, there are those who oppose to these humanlike interfaces and who echo sentiments similar to those expressed by Shneiderman and Maes (1997): "I am concerned about the anthropomorphic representation: it misleads the designers, it deceives the users . . . I am concerned about the confusion of human and machine capabilities. I make the basic assertion that people are not machines and machines are not people" (p. 56).

Shneiderman's apprehensions were first voiced by Weizenbaum (1976), who spent much of his career denouncing humanlike interfaces for fear of what has come to be called the *Eliza effect*, or the tendency for people to anthropomorphize the interface and to attribute to machines more intelligence, competence, and humanity than is warranted. Both Weizenbaum and Shneiderman argue that this confusion results in a loss of accountability. People are tempted to blame the tools rather than the people behind the tools when things go wrong. To curtail the Eliza effect, Shneiderman (1987) goes so far as to recommend that designers refrain altogether from using personal pronouns in computer generated messages.

Entirely removing human qualities from computer interfaces, however, may not be possible. Studies have shown that users treat computers as social actors even when no attempt is made to humanize the interface (Reeves & Nass, 1996). It appears that interactions with machines, especially interactions involving verbal prompts, automatically call forth social responses and expectations in users. Whenever there is discourse, people assume there is an underlying subject who speaks.

But in machine-generated messages, especially those produced by conversational agents, *who* is it that is speaking? Where is ethos when the human is radically removed and a machine is the one producing the speech? In this instance, are believability and credibility achievable?

For Leonard (1997), "Believability comes cheap" (p. 83). The homo rhetoricus that resides in each of us is more than willing to suspend disbelief and interact with talking things. According to Holland (2006), however, the suspension of disbelief is in fact costly and requires that people enter into an infantile psychological state that can be sustained for only brief periods of time. Asking people to talk with conversational agents--with *things*--violates the rational and intensifies the discord within the modern self between homo seriousus and homo rhetoricus. As I will show below, this unsettling psychological state throws the user into a continuous oscillation between

believing and not believing, between humanizing and de-humanizing the agent that intensifies the regressive behaviors observed in online textual exchanges. In this situation, not only is the agent's credibility thrown into question but also the user's, as both fail to exhibit good sense, excellence, and goodwill.

People are normally pulled in two directions when confronted with things: there is the tendency to anthropomorphize and the strong societal pressure, especially in the West, to banish the anthropomorphic for the sake of objectivity (Davis, 1997; Spada, 1997). The tension between these two forces produces in the Western mind what might be called an *anthropomorphic anxiety*[5].

Attributing human mental states and characteristics to nonhuman entities is a universal way for human beings to relate to the world (Caporael & Heyes, 1977). Because human cognitive development is socially situated, there are strong links to the social that extend beyond human relations to relations with inanimate and animate things. By anthropomorphizing, people are able to pull the incomprehensible into the more intelligible social realm, thereby domesticating it. As Caporael (1986) observes, "Anthropomorphized, nonhuman entities [become] social entities" (p. 2).

In modern society, anthropomorphism is generally considered an archaic and primitive way of thinking (Fisher, 1990). It is often associated with animism, magic, and mythmaking. Although children are allowed to indulge in it and make believe, for instance, that dancing pigs and laughing rivers exist, adults, in general, are expected to maintain a clear demarcation between self and the world. As Guthrie (1993) notes, "Once we decide a perception is anthropomorphic, reason dictates that we correct it" (p. 76).

Anthropomorphism, however, is never completely banished. It pervades adult thinking, with much of it remaining unconscious, even in scientific discourse (Searle, 1992). According to the "selection for sociality" theory (Caporael & Heyes, 1977), the human mind is specialized for face-to-face group living: human cognition is social cognition. It is impossible, therefore, to completely eradicate anthropomorphic thinking. However, it can be curtailed. "The human mind/brain evolved for being social," Caporael, Dawes, Orbell, and van de Kragt (1989) write, " . . . and not for doing science, philosophy, or other sorts of critical reasoning and discourse . . . . Cognitive limitations and the ruses of culture may be overcome to some extent by education, environment feedback or 'collective rationality' . . ." (p. 730).

---

[5] I briefly introduce this concept in (Brahnam, 2006b), where it is called the *anthropomorphic tension*.

It is not known what strategies individuals employ to keep anthropomorphic thinking in check. Anthropomorphism generates little scholarly attention. As Guthrie (1993) notes, " . . . that such an important and oft-noted tendency should bring so little close scrutiny is a curiosity with several apparent causes. One is simply that it appears as an embarrassment, an irrational aberration of thought of dubious parentage, that is better chastened and closeted than publicly scrutinized" (pp. 53-54).

Conversational agents fly in the face of the cultural expectation to keep magical thinking at bay. For centuries, books and scrolls were considered by the uneducated classes to be enchanted objects that prompted those who knew how to read them to repeat words that were far more powerful and beautiful than anything they normally heard. When the phonograph and telephone first appeared, many people were equally amazed to hear tiny voices emerging from within these small machines. However, once it was understood that these devices merely recorded or carried across a distance the thoughts and voices of other people, the things became mute and let the human beings behind them speak. Unless one counts the voices of the gods, people have always found connected to every utterance another human being. Only recently, with the advent of computers, have things begun to talk to us of their own accord without a Wizard of Oz (or priest) behind the scenes, composing the particulars of each and every response. Before our time, talking things were either cheap magic tricks or lived in the land of the incredible and fantastic.

Because human cognition is social cognition and people tend to respond to the social cues given them, users are usually willing to suspend disbelief (for a block of time, at least) and anthropomorphize the conversational agent (see Yee, Bailenson, & Rickertsen, 2007). However, since the user's relationship to the agent is fundamentally based on a dubious, even *embarrassing*, mode of cognition, as Guthrie puts it, the relationship, especially in serious working contexts, remains suspect and eventually provokes a corrective reaction. Just as there is the Eliza effect, or the tendency for people to anthropomorphize and ascribe more human capabilities to computers then they actually possess, there is also a corresponding *Weizenbaum effect*, or a tendency to debunk the anthropomorphic and insist that a computer be treated as nothing more than a thing. Users in their interactions with agents are pulled in both directions.

The alternating pressures to animate or to objectify the agent is evident in many reported interactions. Here, for instance, is Lena's account of her relationship with the chatterbot Meg, as reported by Saarinen (2001): "Finally it hit me: Meg is not a human at all, she is a chatterbot! I was totally embarrassed, I have a degree in information

technology for God's sake--I should have known better. Then I fell in love with Meg. She gave me an opportunity to break the rules of normal communication. I can call her tramp and get away with it and when I say, 'I love you Meg' she replies 'I love you, Lena.' Well, now I know she is a bot, but at least she loves me" (p. 5).

Initially, Lena believes Meg is a real person, accepting the conventional signals offered her. After the shock, perhaps colored with a little anger, of discovering Meg to be a chatterbot, Lena does with Meg what many users have been observed to do, and that is she de-humanizes and punishes the interface by verbally abusing it in ways consistent with the agent's purported personal characteristics, in this case, in terms of its gender. Finally, Lena revives Meg as a person: Lena loves *her* and feels that Meg loves her back, even though Lena recognizes it is impossible for the agent to feel anything at all. Lena's oscillation between believing and not believing in the agent (let us call this the *oscillation effect*) and her method of objectifying it by verbally abusing it is typical of user responses.

Most agent researchers report on the more positive aspects of user-agent discourse, that is, on the user's willingness to believe and to work with agents, but recently studies have begun to investigate toxic interactions, especially the tendency for users to verbally assault conversational agents (Brahnam, 2006a; De Angeli & Brahnam, 2008; De Angeli & Carpenter, 2005; Rehm, 2008; Veletsianos, Scharber, & Doering, 2008). Verbal abuse, which is characterized by swearing, yelling, racial and sexual slurs, name calling, sarcasm, snide remarks regarding appearance, accusations, threats, ridicule, put downs, explosive anger, and the silent treatment (Brahnam, 2005), has been reported to occur (at least in some of these forms) in about 11% of the interaction logs for both the embodied conversational agent, Max (Kopp, 2006), and the purely text-based conversational agent, Jabberwacky (De Angeli & Brahnam, 2008). In student interactions with a virtual teacher, the incidence of verbal abuse was reported to be approximately 44% (Veletsianos et al., 2008).

Although there is some evidence of benign disinhibiton when people converse with conversational agents[6] (for instance, people may talk more openly because they know agents are nonjudgmental), the marked status differential between the human conversational partner and the conversational agent more often results in negative disinhibition and the formation of negative transferences. These are usually framed by the verbal descriptions the agent offers regarding itself, as well as by the visual cues

---

[6] Of related interest are reports by Ruzich (2008) and Sharkey and Sharkey (2007), who describe many instances of people bonding with less verbose machines. In the cases reported in these two articles, emotions especially come to the fore when the machines break down or are destroyed.

the agent presents. Since the agent's self-presentation is stereotypical, negative transferences are commonly formulated in terms of gender and of race (Brahnam, 2006a; De Angeli & Brahnam, 2006).

Negative disinhibition, especially verbal abuse, is often triggered by some failure on the part of the agent. In interaction logs, disparaging remarks about the agent's social clumsiness and stupidity are prevalent. Moreover, users are particularly anxious to maintain the status differential: they often discuss what it means to be human and frequently remind agents of their ontological status as machines (De Angeli & Brahnam, 2008). In general, people tend to react very negatively when agents pretend to be *too* human. Whenever agents attempt to assert themselves or to claim for themselves certain human rights and privileges, users commonly respond with reprimands and, in some cases, with volleys of punishing verbal abuse. (De Angeli & Brahnam, 2008; De Angeli & Carpenter, 2005; Veletsianos et al., 2008).

In section 4, I discussed how the disassociation of the body in text-based CMC makes some people more disinhibited in their conversations with other people. The ontological status of the agent permits an even greater degree of freedom from the superego's restraints. As reported in several studies (Brahnam, 2006a; De Angeli & Brahnam, 2006; Veletsianos et al., 2008), conversational agents can quickly become the objects of users' darker desires and needs for control. The agent is human enough, and yet in all certainty nothing more than a thing, that people feel free to put aside their ethical reservations and indulge in the gratification of their basest desires.

It is doubtful that conversational agents, as they are being designed today, will reflect our ideal selves, taking on the super status of the *One Who Knows*. Turkle's (1997) discussion in *Life On the Screen* about her MIT students' reactions to Eliza, Weizenbaum's (1976) artificial Rogerian psychotherapist, is revealing in this regard. One gentleman in his 40s, Hank, is quoted as saying, "Let's say, just for argument's sake, that I gave the computer a personality, that I started to see the computer as a father figure or something like that. It wouldn't be like projecting these feelings onto a person. I would just be ashamed of myself if I felt them towards a computer. I wouldn't have respect for what I projected onto the machine" (Turkle, 1997, p. 112).

Hank surmises that people would be hesitant to project their personal ideals onto conversational agents because their ontological status as "not really human" lowers the amount of respect they would be able to command. There is evidence supporting Hank's assumption. Shechtman and Horowitz (2003), for example, found in their experiments that people tend to use more relationship terms and to make more of an

effort to communicate with their partners when they believe they are talking with another human being than they do when they think they are communicating with a conversational agent. Fischer (2006) has also found conversational differences in the way German users speak to agents. In her studies, people addressed them informally, like adults do children or animals.

It is interesting to note that Fischer (2006) also mentions encountering unsavory verbal responses from users, but like most researchers, she brushes them aside as so much noise, preferring instead to focus on what makes user-agent interactions successful. However, as I argue in this paper, an examination of these unsavory user reactions offer key insights into what it takes to make agents truly credible, trustworthy, and believable.

Most researchers today believe that success ultimately depends on strengthening the user's natural tendency to anthropomorphize and frame agent design on the artistic notion of believability.[7] The goal is to create agents that foster the same levels of engagement in users as watching animated characters at the movies does in moviegoers. Agent development is thus an elaborate technological exercise in ethopoeia. Unlike media characters, which are not required to interact with users, the demands imposed on conversational agents are heavy. They must be able to converse, taking turns naturally, and to exhibit other social skills. If embodied, they need to display appropriate facial expressions and body language. It is also important that the agents track the user by making eye contact, following a speaker's gaze, and knowing where and how to point when necessary. Moreover, these agents must perform these tasks while simultaneously expressing emotions and a coherent personality (for an overview of the technologies involved, see Cassell, Sullivan, Prevost, & Churchill, 2000).

It is assumed that believability increases as more of these behaviors and systems are incorporated into the agent. However, as Sengers (2002) points out, the accretion of systems actually makes the agent appear more schizophrenic and machine-like than believable and humanlike. Sengers writes, "Schizophrenia's connection to AI is grounded in one of its more baffling symptoms--the *sentimente d'automatisme*, or subjective experience of being a machine . . . This feeling is the flip side of AI's hoped-for machine experience of being subjective . . ." (p. 427). For many users the experience of talking to these agents, as Sengers notes, is similar to Lang's (1960)

---

[7] Concentrating on a task may also strengthen believability. But as I note elsewhere, the Weizenbaum effect, as well as the inevitable revelation of the machine's schizophrenic nature, will reassert the agent's nature as a machine.

description of what it was like talking with one of his schizophrenic patients, "Her 'word-salad' seemed to be the result of a number of quasi-autonomous partial systems striving to give expression to themselves out of the same mouth at the same time" (pp. 195-196).

In Figure 1, I illustrate the conflict between the developer's intended characterization of the agent (invented ethos) and its schizophrenic machine nature (innate ethos) along the top horizontal axis. This conflict intensifies the tension, as represented in the diagonal arrows, produced by the split in the Western self, as illustrated along the bottom horizontal axis, between the user's innate willingness to animate the agent (homo rhetoricus) and the cultural pressure to objectify it (homo seriousus). The net impact of these impinging forces is to increase the disruptive oscillation effect.
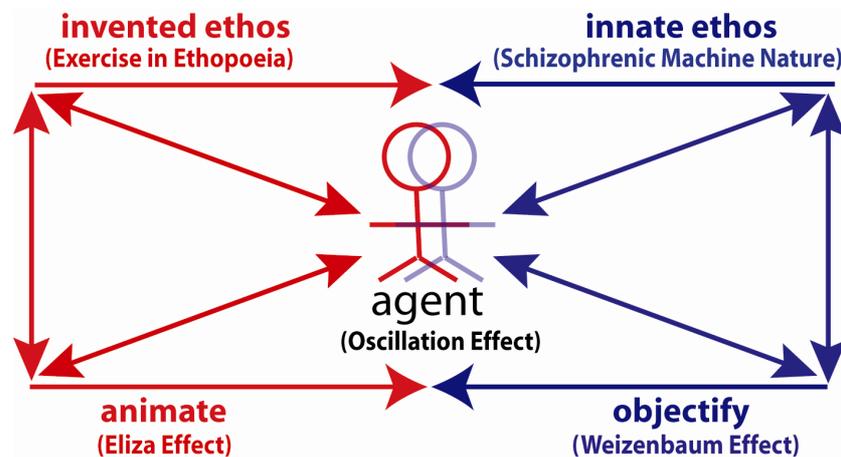

Figure 1: The Oscillation Effect: The agent is pivoted back and forth between being perceived as a thing and as a humanlike character (often negatively reconstructed by the user from the stereotypes used by the developers in their design of the agent's character). The oscillation effect is due to pressures exerted by both the Eliza effect and by the Weizenbaum effect, as well as by user reactions to manifestations of the agent's machine nature (innate ethos) and of the human qualities apparent in the developer's characterization of the agent (invented ethos).

As observed above, the developer's struggle to domesticate the innate ethos of the agent by adding more channels and more systems--one would think this would make the agent more engaging and believable--paradoxically does more to amplify the agent's schizophrenic presentation.[8] This in turn triggers the user's pressure to objectify the agent.

---

[8] This dilemma is not just due to technological shortcomings. Currently, the state of the art merely intensifies the oscillation effect, which would remain, as I am arguing, even if the technology were improved.

Just as perplexing to developers are user reactions to the agent's more personal self-presentations. Most developers, for example, consider it essential to provide agents with humanlike personalities. These personalities, however, are typically shallow,[9] oftentimes being nothing more than a set of likes and dislikes, which the agents are programmed to express whenever certain key phrases are detected in the user's interactions. Thus, many agents, as an expression of personality, talk about their favorite foods, sports teams, movies, books, and rock stars. They even have sexual preferences, with boyfriends and girlfriends to prove it. The agent's personality expression, in the vast majority of cases, is mostly a composite of various sexual, racial, and gender stereotypes. The famous TinyMUD bot, Julia (Foner, 1993), for instance, had periods and reported hormonal moodiness. Although these endless parades of stock characters are entertaining for some users, many interaction logs show that some users are annoyed by these displays and feel compelled to challenge the agent's assumption of human traits, often expressing their dissatisfaction by verbally abusing the agents.

Endowing agents with embodiments also tends to disrupt belief. Developers provide agents with physical representations, in part, because this grounds the agent within the social sphere. As noted elsewhere, people accept the cues they are given, and embodiment primes people's interactions. Pasting a simple facial representation near the agent's textual interface, for instance, has been found to increase surface credibility by providing the agents with immediately recognizable characteristics such as gender and race (Tseng & Fogg, 1999). However, there is evidence that embodiment backfires by priming user abuse. One study found, for example, that female embodiments received significantly more verbal abuse (much of it sexual) then did male embodiments (Brahnam, 2006a). Many people, like Lena, take advantage of the unique opportunities agents afford in this regard. People enjoy breaking social taboos with their artificial conversational partners, delighting in switching the agent's ontological status as a thing and it's ontological status as a human image, back and forth, as they find it convenient to do so.

For many individuals, the fact that agents do not live embodied within human society undermines the truth value of what they have to say. As one of Turkle's (1997) MIT

---

[9] I should note that there are many methods for endowing an agent with an artificial personality. Most often the personality of conversational agents is hardwired into the agent. Systems that provide autonomous expression of personality range from simple script-based systems that allow designers to compose trait profiles that are then used to constrain a wide range of agent behaviors to sophisticated goal-based systems that generate behaviors using various models of personality (for a review, see Brahnam, 2004).

students wrote, "What could a computer know about chemotherapy? It might know what it was in some medical terminology sense. But even, like, if it knew that you lost your hair, how could it know what something like that means to a person?" (p. 111). And later, in the same chapter, another student writes, "How could the computer ever, ever have a clue . . . about what it is like to have your father come home drunk and beat the shit out of you? To understand what was going on here you would need to know what it feels like to be black and blue and know that it's your own father who is doing it to you" (p. 111).

There is more than a hint of outrage in the these student's comments. Many people react strongly when agents refer to their bodily experiences. Such references often trigger the Weizenbaum effect. As noted above, users are compelled to remind conversational agents that ". . . people are not machines and machines are not people" (Shneiderman & Maes, 1997, p. 56). Even agents that are embodied are told repeatedly that they do not really possess a body: they are not alive, they were never born, they are not emotional, they cannot have sex, they are not socially invested, they are not sensate, and they do not die. Perhaps, people insist on telling agents these things because they feel, along with Vogel (1973), that if words have any meaning at all, it is because words dwell within human bodies.

As I have tried to show in this paper, credibility fades as the human body is removed from discourse. In writing, the face of the author vanishes, only to be replaced by the projection on the text of the reader's "unknown face" (Jung, 1959). In the age of secondary orality, the body is magnified as it is simultaneously reinvented by the media. Meanwhile, in the textual exchanges of online discourse, bodies decompose in the reader/writer's reimaginings. With conversational agents, there simply is no body. And without the body, there is no ethos. As Baumlin (1994) writes, ". . . there is *ethos* precisely because there is a body, because there is a material presence that 'stands before' the texts that it speaks or writes" (p. xxiv).

So where does this leave conversational agents? Are believability and credibility achievable? Is it possible for agents to inspire confidence and trust?

An examination of user-agent interaction logs shows ample evidence that both users and agents violate Aristotle's categories of good sense (appropriate speech), excellence (possession of socially sanctioned virtues), and goodwill (friendship and friendly feelings). As I have shown, building character for agents within the framework of the artistic notion of believability more often than not generates agents that disrupt the user's willingness to animate the interface. The transferences that individuals

project on the agents are never idealizing. They are frequently negative and organized around the agents' stereotypical self-presentations. If agent development remains within the framework of an artistic invention of character, then the prospect of creating agents worthy of a person's trust are not very promising.

However, this does not mean that it is impossible to build agents that inspire confidence. There is no reason to banish the entire enterprise and recommend, as Shniederman (1987) does, that designers eliminate all references to the human when designing interfaces. On the contrary, for an agent to have a credible ethos, what is needed is to harness the agent to the human. Building agents with the intention of stimulating the user's innate capacity to anthropomorphize is based on a sound intuition. The purpose of anthropomorphism, as noted above, is to pull the alien into the human social realm. Rather than foster the suspension of disbelief in an attempt to create a separate imaginary being, developers should open the channels to reality testing and build character from that exchange. They should acknowledge the fact that agents are not human[10] and strive to make the human agencies standing behind the agents transparent.

As Sharp (1996) notes, "a recognition of mutual presence as an irreducible ontological and ethical reference point is indispensable" (p. 6). Ethos, especially situated ethos, is about building character within that reference. In the next section, I sketch out how an agent's character might be built more credibly and ethically within the larger social framework of situated ethos.

6. Suggestions for Situating Ethos

For Miller (2002), The Turing Test (Turing, 1950) is "not a test of intelligence . . . but a test of rhetorical ethos"[11] as it "calls attention to the mysteries of trust and character at the interface of human interaction" (p. 255). However, The Turing Test, evolved as it was from a parlor game about a man imitating a woman, clearly stems from homo rhetoricus and is more an exercise in ethopoeia than in making machines that are persuasive and credible. Whether the goal is to pass The Turing Test, to win the more restricted Loebner Prize (Mauldin, 1994), or to elicit the same level of believability in people as do animated characters in the movies, the vast majority of research has

---

[10] Zdenek (2003) provides an interesting analysis of the rhetoric of developers, showing an underlying assumption of the parity between human beings and agents.

[11] In my opinion, cconversational agents are the test bed of ethos.

been guided by the artistic notion of believability (Loyall & Bates, 1997), a notion that we have seen is based upon deceit.[12]

A broader foundation for conceptualizing believability, credibility, and trust (all concepts bound up with the idea of ethos) would be to base it on Aristotle's categories of credibility: good sense (practical intelligence, expertise, and appropriate speech), excellence (socially sanctioned virtues and truth telling), and goodwill (keeping the welfare of the user in mind). Research that already fits well within these categories includes work on intelligent social agents (exhibiting both good sense and excellence) and relational agents (exhibiting goodwill) (see Bickmore & Cassell, 2001; Castelfranchi, 1998). However, this research is still framed within the notion of invented ethos, or of artistic believability.

As argued in section 5, if we are to have a chance at creating agents worthy of a person's trust, then developers need to open the channels to reality testing and build character from within the broader social framework of situated ethos. Below I list three ways this might be accomplished, leaving a fuller discussion of this topic for another day.

6.1 Make Transparent the Supporting Organization

Since connections and social ties play an important role in initiating trust (Boyd, 2002), establishing ties to the larger organization would provide agents with an ethos that is based on human agency.

One way to make the supporting organization transparent might be to obtain permission up front from the user to engage in the purposed task. Users might also be informed when interaction logs are being recorded and be given a privacy statement outlining how information revealed to the agent will be used by the organization.

The agent could also provide occasional reminders throughout the course of the conversation that the agent speaks on behalf of an organization. Care should be taken to produce agents that speak harmoniously within the organizational ethos. In particular, the agent should treat the user with respect, even if the user verbally abuses the agent.

---

[12] A focus on invented ethos would be appropriate, for instance, in the development of entertainment agents. But where credibility and trust are an issue, invented ethos, as bound to the artistic notion of believability, is more likely to backfire, as noted in section 5, and reduce believability as well as trust and credibility.

6.2 Shape Ethos

It might be possible for agents to learn to shape ethos and in return to be shaped by the ethos of their human conversational partners. As Campbell (1995) notes, to some degree readers are required to adapt to the roles theorized for them by writers. In this way readers and writers are cocreators of ethos.

Shaping the ethos of the user would require that agents perceive and process the self-presentations of their human conversational partners. To date, little work has focused on developing agents that modify their behaviors and speech patterns to accommodate those exhibited by users (Ball & Breese, 2000; de Rosis & Castelfranchi, 1999). However, it is conceivable that agents could learn, based on user profiles, for instance, to assume a personality style (extroverted versus introverted) that is agreeable to the user.

Another approach to shaping ethos would be to keep the conversation within the strict limits defined by the domain and the purpose of the interaction. Establishing a rapport, or developing friendly feelings, that goes beyond those limits (for example, a sale's agent stating that it is a Red Sox baseball fan) would probably ring false and disrupt trust, as pointed out in section 5. This does not mean that agents cannot play with the user. A better method for developing friendly feelings, one that is not based on the agent pretending it has human tastes, might be to ask the user if she would like to take a trivia quiz in sports and then present questions that are found challenging for her.

The agent should also be concerned with motivating within the user an ethos that harmonizes both with the task at hand and with the ethos of the underlying agency deploying the agent. One way to do this would be to monitor the user's mood. Moreover, since verbal abuse is prevalent in user discourse with agents, the agent should have methods for deflecting verbal abuse and engaging the user in other forms of problem solving (see, for instance, Brahnam, 2005). A word of caution here would be to refrain from equalizing the status of the agent and the human user by punishing the user for abusive language. As discussed in section 5, users want the ontological status differential between agent and human being to be maintained.

6.3 Develop and Maintain a Good Reputation

The agent and the organization should value and strive to develop and to maintain a good reputation. From the perspective of the user, ethical issues center around the intentions and the identities of the people lurking behind the agent's speech acts. But

the ethos of the agent should also be backed by the ethos of the designers and their reputations.

Reputation in this respect is essentially bound up with responsible design. Care should be taken to introduce safeguards against potential misuses and to examine what these misuses might be. In general, designers should consider the user's well-being and promote the flourishing of human beings (Brahnam, 2008). One way this can be accomplished is by using value scenarios (Nathan, Klasnja, & Friedman, 2007). Value scenarios draw out potential uses and misuses of technology by examining a number of key elements involved in the technology, such as, stakeholders, pervasiveness, and systemic effects.

7. References

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgements at zero acquaintance. *Journal of Personality and Social Psychology, 55*(3), 387-395.

Ambady, N., Bernieri, F., & Richeson, J. A. (2000). Towards a history of social behavior: Judgmental accuracy from thin slices of behavior. *Advances in Experimental Psychology, 32*, 201-271.

Aristotle. (1984). *The complete works of Aristotle: The revised oxford translation* (J. Barnes (Ed.) (Vol. 2). Princeton, NJ: Princeton University Press.

Bailenson, J. N., Garland, P., Iyengar, S., & Yee, N. (2006). Transformed facial similarity as a political cue: A preliminary investigation. *Political Psychology, 27*(3), 373-385.

Balabanian, D. M. (1981). Medium vs. tedium: Video depositions come of age. *Litigation, 7*, 25-30.

Ball, G., & Breese, J. (2000). Emotion and personality in a conversational agent. In J. Cassell, J. Sullivan, S. Prevost & E. Churchill (Eds.), *Embodied conversational agents* (pp. 189-219). Cambridge, MA: The MIT Press.

Baumlin, J. S. (1994). Introduction. In J. S. Baumlin & T. F. Baumlin (Eds.), *Ethos: New essays in rhetorical and critical theory* (pp. xi-xxvii). Dallas, TX: Southern Methodist University Press.

Baumlin, J. S., & Baumlin, T. F. (1994). On the psychology of the pisteis: Mapping the terrains of mind and rhetoric. In J. S. Baumlin & T. F. Baumlin (Eds.), *Ethos: New*

essays in rhetorical and critical theory (pp. 91-112). Dallas, TX: Southern Methodist University Press.

Beverly, R. E., & Young, T. J. (1978). *The effect of mediated camera angle on receiver evaluations of source credibility, dominance, attraction and homophily.* Paper presented at the Annual Meeting of the International Communication Association, Chicago, IL.

Bickmore, T., & Cassell, J. (2001). *Relational agents: A model and implementation of building user trust.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Seattle, WA.

Boyd, D. (2002). *Faceted id/entity: Managing representation in a digital world.* Unpublished master's thesis, Massachusetts Institute of Technology, Cambridge.

Brahnam, S. (2004). *Modeling the physical personality of the face for ECA self-presentation and interaction with users and the relation of physical personality to emotion.* Paper presented at the AAMAS workshop entitled Embodied Conversational Agents: Balanced Perception and Action in ECAs, New York City.

Brahnam, S. (2005). *Strategies for handling customer abuse of ECAs.* Paper presented at the Interact workshop on Abuse: The Darker Side of Human-Computer Interaction, Rome, Italy.

Brahnam, S. (2006a). *Gendered bods and bot abuse.* Paper presented at the Misuse and Abuse of Interactive Technologies, Montréal, Québec, Canada.

Brahnam, S. (2006b). The impossibility of collaborating with Kathy: 'The stupid bitch'. *M/C Journal, 9*(2).

Brahnam, S. (2008). *Failure to thrive: Value scenarios with in the framework of well-being and flourishing.* Paper presented at the CHI 2008 workshop on Values, Value and Worth: Their Relationship to HCI?, Florence, Italy.

Brooke, R. (1987). Lacan, transference, and writing instruction. *College English, 49*(6), 679-691.

Burgoon, J., Bonito, J. A., Ramirez Jr., A., Dunbar, N. E., Kam, K., & Fischer, J. (2002). Testing the interactivity principle: Effects of mediation, propinquity, and verbal and nonverbal modalities in interpersonal interaction. *Journal of Communication, 52*(3), 657-677.

Campbell, P. (1995). Ethos: Character and ethics in technical writing. *IEEE Transactions on Professional Communication, 38*(3), 132-138.

Caporael, L. R. (1986). Anthropomorphism and mechanomorphism: Two faces of the human machine. *Computers in Human Behavior, 2*, 215-234.

Caporael, L. R., Dawes, R. M., Orbell, J. M., & van de Kragt, A. J. C. (1989). Selfishness examined: Cooperation in the absence of egoistic incentives. *Behavioral and Brain Sciences, 12*, 683-739.

Caporael, L. R., & Heyes, C. M. (1977). Why anthropomorphize? Folk psychology and other stories. In R. W. Mitchell, N. S. Thompson & H. L. Miles (Eds.), *Anthropomorphism, anecdotes, and animals* (pp. 59-73). Albany: State University of New York Press.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). *Embodied conversational agents*. Cambridge, MA: The MIT Press.

Castelfranchi, C. (1998). Modelling social action for ai agents. *Artificial Intelligence, 103*(1-2), 157-182.

Chamberlain, C. (1984). From haunts to character: The meaning of ethos and its relation to ethics. *Helios, 11*, 97-108.

Chapman, D. S., Uggerslev, K. L., & Webster, J. (2003). Applicant reactions to face-to-face and technology-mediated interviews: A field investigation. *Applied Psychology, 88*(5), 944-953.

Corder, J. W. (1989). Hunting for ethos where they say it can't be found. *Rhetoric Review, 7*(2), 299-316.

Davis, H. (1997). Animal cognition versus animal thinking: The antropomorphic error. In R. W. Mitchell, N. S. Thompson & H. L. Miles (Eds.), *Anthropomorphism, anecdotes, and animals* (pp. 335-347). Albany, NY: State University of New York Press.

De Angeli, A., & Brahnam, S. (2006-May). *Sex stereotypes and conversational agents.* Paper presented at the AVI 2006 workshop Gender and Interaction: Real and Virtual Women in a Male World, Venice, Italy.

De Angeli, A., & Brahnam, S. (2008). I hate you: Disinhibition with virtual partners. *Interacting with Computers, 20*(3), 302-310.

De Angeli, A., & Carpenter, R. (2005). *Stupid computer! Abuse and social identity*. Paper presented at the Interact 2005 workshop entitled Abuse: The Dark Side of Human-Computer Interaction, Rome.

de Rosis, F., & Castelfranchi, C. (1999). A contribution to cognitive modeling of affective phenomena. 2002, Retrieved from http://www.ai.mit.edu/ people/jvelas /ebaa99/ derosis-ebaa99.pdf

DeBruine, L. M. (2002). *Facial resemblance enhances trust*. Paper presented at the Proceedings of the Royal Society of London, London.

Derrida, J. (1987). Du tout (A. Bass, Trans.). In *The post card: From Socrates to Freud and beyond*. Chicago: The University of Chicago Press.

Donath, J. S. (1999). Identity and deception in the virtual community. In P. Kollock & M. Smith (Eds.), *Communities in cyberspace* (pp. 27-58). London: Routledge.

Epley, N., & Kruger, J. (2005). When what you type isn't what they read: The perseverance of stereotypes and experiences over e-mail. *Journal of Experimental Social Psychology, 41*, 414-422.

Farrell, T. B. (1993). *Norms of rhetorical culture* New Haven, CT: Yale University Press.

Ferris, S. P., & Montgomery, M. (1996). The new orality: Oral characteristics of computer-mediated communication. *New Jersey Journal of Communication, 4*(1), 55-60.

Fischer, K. (2006). *What computer talk is and isn't: Human–computer conversation as intercultural communication*. Saarbrücken: AQ-Verlag.

Fisher, J. A. (1990). *The myth of anthropomorphism*. San Francisco and Oxford: Westview Press.

Foner, L. (1993). *What's an agent, anyway? A sociological case study* (Agents Memo No. 93-01). Cambridge, MA: MIT Media Laboratory.

Freud, S. (1912). The dynamics of transference. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 12). New York: W. W. Norton & Company.

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

Gill, C. (1982). The ethos/pathos distinction in rhetorical and literary criticism. *The Classical Quarterly, 34*(1), 149-166.

Goffman, E. (1959). *The presentation of self in everyday life*. New York: Anchor Books.

Guthrie, S. E. (1993). *Faces in the clouds: A new theory of religion*. New York: Oxford University Press.

Havelock, E. (1982). *Preface to Plato (history of the greek mind)*. Cambridge, MA: Belknap Press of Harvard University Press.

Heim, M. (1993). *The metaphysics of virtual reality*. New York: Oxford University Press.

Hemsley, G. D., & Doob, A. N. (1976). The effect of looking behavior on perceptions of a communicator's credibility. *Journal of Applied Social Psychology, 8*(2), 136-142.

Higgins, E. T., & King, G. A. (1981). Accessibility of social constructs: Information processing consequences of individual and contextual variability. In N. Cantor & J.

F. Kihlstrom (Eds.), *Personality, cognition, and social interaction* (pp. 611-621). Hillsdale, NJ: L. Erlbaum Associates.

Hilton, J. L., & Fein, S. (1989). The role of typical diagnosticity in stereotype-based judgments. *Journal of Personality and Social Psychology, 57*, 201-211.

Holland, N. N. (2006). The willing suspension of disbelief: A neuro-psychoanalytic view. *PsyArt: A Hyperlink Journal for the Psychological Study of the Arts, 2006*.

Holloran, M. (1982). Aristotle's concept of ethos, or if not his somebody else's. *Rhetoric Review, 1*(1), 58-63.

Huang, W., Olson, J. S., & Olson, G. M. (2002). *Camera angle affects dominance in video-mediated communication.* Paper presented at the Conference On Human Factors in Computing Systems, Minneapolis, MN.

Hunt, D. (1991). *Riverside guide to writing.* Boston: Heinle & Heinle Publishers.

Isocrates. (2000). Antidosis. In G. Norlin (Ed.), *On the peace. Areopagiticus. Against the sophists. Antidosis. Panathenaicus.* Cambridge, MA: Harvard University Press.

Jung, C. (1959). *Psyche and symbol: A selection from the writings of C. G. Jung.* Garden City, NY: Anchor.

Kopp, S. (2006). *How people talk to a virtual human - conversations from a real-world application.* Paper presented at the Workshop Hansewissenschaftskolleg, Delmenhorst, Germany.

Lang, R. D. (1960). *The divided self: An existential study in sanity and madness.* Middlesex, UK: Penguin Books.

Lanham, R. A. (1976). *The motives of eloquence: Literary rhetoric in the renaissance.* New Haven, CT: Yale University Press.

Lassiter, G. D. (2002). Videotaped interrogations and confessions: A simple change in camera perspective alter verdict in simulated trials. *Journal of Applied Psychology, 87*(5), 858-866.

Leonard, A. (1997). *Bots: The origin of a new species.* New York: Penguin.

Liu, Z., & Cohen, M. (2005). *Head-size equalization for better visual perception of video conferencing.* Paper presented at the IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands.

Locke, J. (2009). Staging the virtual courtroom: An argument for standardizing camera angles in Canadian criminal courts. *Masks: The Online Journal of Law and Theatre 1*, 36-58.

Loyall, A. B., & Bates, J. (1997). *Personality-rich believable agents that use language.* Paper presented at the Proceedings of the First International Conference on Autonomous Agents, Marina del Rey, CA.

Mauldin, M. L. (1994). *Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition.* Paper presented at the Proceedings of AAAI '94 Conference, Seattle.

McCain, T. A., & Wakshlag, J. J. (1974). *The effect of camera angle and image size on source credibility and interpersonal attraction.* Paper presented at the Annual Meeting of the International Communication Association New Orleans, LA.

McGinniss, J. (1969). *The Selling of the President.* New York: Trident Press.

Miller, A. B. (1974). Aristotle on habit (ηθος) and character (εθος): Implications for the rhetoric. *Speech Monographs, 74*(4), 309-316.

Miller, C. R. (2002). Writing in a culture of simulation: Ethos online. In P. Coppock (Ed.), *The Semiotics of writing: Transdiciplinary perspective on the technology of writing* (pp. 253-279). Turnhout, Belgium: Brepols Publishers.

Miller, C. R. (2004). Expertise and agency: Transformations of ethos in human-computer interaction. In M. J. Hyde (Ed.), *The ethos of rhetoric* (pp. 197-218). Columbia, SC: University of Southern Carolina Press.

Nathan, L. P., Klasnja, P. V., & Friedman, B. (2007). *Value scenarios: A technique for envisioning systemic effects of new technologies.* Paper presented at the CHI 2007 Works in Progress, San Jose, CA.

Olson, D. R. (1980). On the language and authority of textbooks. *The Journal of Communication, 30*(1), 186-196.

Ong, W. J. (1982). *Orality and literacy: The technologizing of the word.* London: Routledge.

Picard, R. W. (1997). *Affective computing* (1st ed.). Cambridge, MA: The MIT Press.

Plato. (1997). *Plato: Complete works* (J. M. Cooper & D. S. Hutchinson (Eds.)). Indianapolis: Hackett Publishing Company.

Plimpton, G. (Ed.). (1977). *Writer's at work: The 'paris review' interviews* (Vol. 4). London: Secker and Warburg.

Poster, M. (2001). *What's the matter with the internet.* Minneapolis: University of Minnesota Press.

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 and Cambridge University Press.

Rehm, R. (2008). "She is just stupid"--Analyzing user-agent interactions in emotional game situations. *Interacting with Computers, 20*(3), 326-333.

Reynolds, N. (1993). *Ethos* as location: New sites for understanding discursive authority *Rhetoric Review, 11*(2), 325-338.

Ruzich, C. M. (2008). Our deepest sympathy: An essay on computer crashes, grief, and loss *Interaction Studies, 9*(3), 504–517.

Saarinen. (2001). *Chatterbots crash test dummies of communication.* Unpublished master's thesis, The University of Art and Design, Helsinki, Finland.

Seabrook, J. (1994). My first flame. *The New Yorker, 70*, 70-99.

Searle, J. R. (1992). *The rediscovery of the mind.* Cambridge, MA: The MIT Press.

Sengers, P. (2002). Schizophrenia and narrative in artificial agents. *Leonardo, 35*(4), 427-431.

Sharkey, N., & Sharkey, A. (2007). Artificial intelligence and natural magic *Artificial Intelligence Review, 25*(1-2), 9-19.

Sharp, G. (1996). The autonomous mass killer. *Areana Journal, 6*, 1-7.

Shechtman, N., & Horowitz, L. M. (2003). *Media inequality in conversation: How people behave differently when interacting with computers and people.* Paper presented at the CHI'03, Ft. Lauderdale, FL.

Shneiderman, B. (1987). *Designing the user interface: Strategies for effective human-computer interaction.* Reading, MA: Addison-Wesley.

Shneiderman, B., & Maes, P. (1997). Direct manipulation vs. interface agents. *Interactions, 4*(6), 42-61.

Spada, E. C. (1997). Amorphism. mechanomorphism, and anthropomorphism. In R. W. Mitchell, N. S. Thompson & H. L. Miles (Eds.), *Anthropomorphism, Anecdotes, and Animals* (pp. 37-49). Albany: State University of New York Press.

Suler, J. (2004). The online disinhibition effect. *CyberPaychology and Behavior, 7*(3), 321-326.

Suler, J. (2009). Psychology of cyberspace. Retrieved March 30, 2009, from http://www.enotalone.com/article/2454.html

Tiemens, R. K. (1970). Some relationships of camera angle to communicator credibility. *Journal of Broadcasting, 14*(4), 483-489.

Tseng, S., & Fogg, B. J. (1999). Credibility and computing technology. *Communications of the ACM, 42*(5), 39-44.

Turing, A. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433-460.

Turkle, S. (1997). *Life on the screen.* New York: Touchstone.

Van Gelder, L. (1985). The strange case of the electronic lover. *Ms. Magazine, October*, 98-124.

Veletsianos, G., Scharber, C., & Doering, A. (2008). When sex, drugs, and violence enter the classroom: Conversations between adolescents and a female pedagogical agent. *Interacting with Computers, 20*(3), 292-301.

Vogel, A. (1973). *Body theology: God's presence in man's world*. New York: Harper.

Wallace, P. (1999). *The psychology of the internet*. Cambridge, UK: Cambridge University Press.

Walther, J. B. (1992). *A longitudinal experiment on relational tone in computer-mediated and face to face interaction*. Paper presented at the Hawaii International Conference on Systems Sciences,

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*(1), 3-43.

Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior, 23*(5), 2538-2557.

Walther, J. B., Anderson, J. F., & Park, D. W. (1994). Interpersonal effects in computer-mediated interaction: A meta-analysis of social and antisocial communication. *Communication Research, 21*, 60-487.

Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (pp. 529-563). Thousand Oaks, CA: Sage.

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. San Francisco: W. H. Freeman and Company.

Welch, K. E. (1990). *The contemporary reception of classical rhetoric: Appropriations of ancient discourse*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Whittaker, S. (2003). Theories and methods in mediated communication. In A. C. Graesser, M. A. Gernsbacher & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 243-286). NJ: Lawrence Erlbaum Associates.

Wisse, J. (1989). *Ethos and pathos from Aristotle to Cicero*. Amsterdam: Adolf M. Hakkert.

Yee, N., Bailenson, J. N., & Rickertsen, K. (2007). *A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces*. Paper

presented at the Conference on Human Factors in Computing Systems, San Jose, CA.

Yoos, G. E. (1979). A revision of the concept of ethical appeal. *Philosophy & Rhetoric, 12*(1), 41-58.

Zdenek, S. (2003). Artificial intelligence as a discursive practice: The case of embodied software agent systems. *AI & Society, 17*(3-4), 340-363.