Traps, Pitfalls, Swindles, Lies, Doubts and Suspicions in Human-Computer Interaction: A Counter-Case for the Study of Good Etiquette

Jack L. Edwards and Greg Scott
AI Management and Development Corporation
206 Keewatin Avenue
Toronto, Ontario, Canada M4P 1Z8
jle2@sympatico.ca

Sharon McFadden and Keith C. Hendy
Defence Research & Development Canada - Toronto
1133 Sheppard Ave., West
North York, Ontario M3M 3B9

Abstract

To understand and implement the rules of etiquette for human-computer interaction, a counter-case for the study of good etiquette is presented. The Internet is seen as a crucible for studying a wide variety of behavior that violates **veracity**, the foundational rule of etiquette. In the past, deceit and its attendant behaviors were exercised more or less directly but with the advent of this ubiquitous computing environment, creative and subtle forms of practicing many of the old dishonesty tricks have arisen. Further, the emergence of (intelligent) agents for both local and global computing environments has opened up new possibilities for even greater abuse. Focusing on the Internet as a forum for studying the rules of good etiquette and their abuse has many advantages, among them: the vast amounts of data and the scope for studying both human and agent deceptions and their consequences. Studies in such a context cannot help but produce useful insights that generalize to many of the morre technical domains of interest presented in this Symposium.

Background and Problem

The term "etiquette" has a nice image with implications of **trust**, consideration for others and a certain interactive refinement. As defined in the Call for Participation for this Symposium, it is somewhat non-committal with respect to these implications:

> *"By 'etiquette', we mean the defined roles, acceptable behaviors and interaction moves of human and intelligent agent participants in a common setting."*

The use of the terms "defined" and "acceptable" do imply standards of behavior but we know that, although many people behave according to those rules of etiquette, they do find ways to abuse them in many circumstances. What is the appropriate response then, when standards of human behavior are violated: When a boss verbally abuses someone who works for him, when an employee bullies her fellow co-workers or when an acquaintance violates the niceties of pleasant conversation?

It is our contention that simply identifying protocols for good etiquette among humans and computers will not lead to a thorough understanding of the subject and to effective and secure implementations. Concentrating only on positive and supporting aspects of interaction leaves one open to many instances of violation for which no adequate response has been anticipated. It is only by studying deliberate violations of good etiquette and the motives behind them that a comprehensive view of human-computer etiquette will begin to emerge.

Hackers are traditional violators of good human-computer etiquette but many other dangers lie on the horizon. For example, a recent AI ALERT (May, 2002), a semimonthly online news service from AAAI, cited the following from the Boston Globe:

> *"Scientists at the Massachusetts Institute of Technology have created the first realistic*

videos of people saying things they never said - a scientific leap that raises unsettling questions about falsifying the moving image. In one demonstration, the researchers taped a woman speaking into a camera, and then reprocessed the footage into a new video that showed her speaking entirely new sentences, and even mouthing words to a song in Japanese, a language she does not speak. The results were enough to fool viewers consistently, the researchers report. ... But scientists warn the technology will also provide a powerful new tool for fraud and propaganda - and will eventually cast doubt on everything from video surveillance to presidential addresses. ... Previous work has focused on creating a virtual model of a person's mouth, then using a computer to render digital images of it as it moves. But the new software relies on an ingenious application of artificial intelligence to teach a machine what a person looks like when talking."

The last part of our definition of etiquette refers to "participants in a common setting." The most ubiquitous common setting today is the Internet where many communications, whether they get through or not, can be sorted into one or more of the terms in the title of this presentation. Examining those types of communication and associated examples will prove useful but, before proceeding to that, a vision is needed for how studying etiquette on the Internet might justify such an examination.

The Internet as the Major Communication Medium of the Foreseeable Future

Why will the Internet be *the* future of most human-agent-human interaction? There are several rather obvious answers to that question. First, some part of almost everyone's work and leisure time now extends beyond local processing. They may be communicating via a local area network, a wide area network or an Intranet but some portion of their work and often their leisure activity (in some cases the majority) will involve the Internet. The fact that vast amounts of both work and leisure time are spent on the Internet, with the consequent vulnerabilities for individual and corporate resources, means that there are more than a few confidence men, tricksters, vandals and even terrorists willing to violate good communication etiquette or, more accurately, to use aspects of good etiquette for malicious, dishonest and destructive dealings.

Second, there has been a trend toward increasing involvement of technology in exchanges that were formerly person-to-person. Consider the following simple example: Ten years ago if a person wanted to get a message to someone in a hurry, he would most likely have picked up the telephone and placed a call. Today, that same task is very often accomplished by sending email.

Both approaches involve human-human interaction where machines, along with their more recently adaptive software, serve as the communication medium. As technology continues to advance, the medium will evolve with future human-computer-human interaction mediated by ever more intelligent agents.

Even now, with the proliferation of spam, people are resorting to the use of junk-mail filters, which delete messages sent anonymously or not addressed explicitly to the recipient. Such filters can block messages from an entire domain if considerable spam is found to come from a single source. Another use of simple agents involves the use of auto-responders to send replies to emails received while a person is away.

A recent Apple newsletter (August 8, 2002) described the new Mac OS X mail software that uses "adaptive latent semantic analysis" to learn what the user considers to be spam and what is legitimate mail.

Like the weather, everyone complains about junk mail. But we've actually done something about it.

We've made Mail smart. It learns. Thanks to sophisticated mail filtering technology - adaptive latent semantic analysis, to be exact—Mac OS X Mail learns what you

consider to be junk email. And after an initial training period, you can tell Mail what to do with your junk email.

So instead of sending email directly to you, I may be sending it to an autonomous piece of software (an agent) that represents your interests and which decides whether or not to pass that email on to you. The very primitive email agents in these examples will (in the years to come) almost certainly evolve into much more complex filtering agents with the ability to learn what the user considers to be junk mail, what isn't, whether to pass email to a user or discard it, whether to forward select email to a private address when the user is away or hold it for later review, and so on.

Third, examining the kinds of agent currently being proposed and developed, it is clear that many need to be Internet-based in order to function effectively. Indirect attacks on the integrity of systems through their supporting agents (including their security agents) thus become distinct possibilities.

Finally, **traps, pitfalls, swindles**, and the like appear to generalize easily to the Internet and ultimately to intelligent agents that travel its pathways.

Adaptation of classic fraud schemes already have brought many of them into the Internet age. Ponzi schemes (see Smithsonian, 1998), such as pyramid scams and chain letters, have moved from paper to fax and now to computer screens, thereby dramatically increasing their potential for widespread distribution and attendant profit.

It is a certainty that in future intelligent agents will be created that will learn which of many fraudulent techniques work best, with whom, where and under what circumstances. Such agents likely will be refined and targeted to segmented portions of the population ("Mark" categories, if you will) and so, studying the old schemes will give insights into what the future holds and how best to counter intelligent, yet fraudulent agents.

The exploits of Limehouse Chappie, Kangaroo John, Yellow Kid Weil and other confidence men of the last century (Maurer, 1940) are nothing compared to those making a much wider presence felt on the Internet. Ubiquitous computing is giving rise to ubiquitous confidence games, some with deadly consequences.

More recent material for comparison with the older con artists can be found on and off the Internet, which deals with everything from confidence tricks to hacking by (intelligent) agents to terrorist attacks. A wealth of material exists but two instructive examples are Zeltser (2000), who identifies a number of malicious agents and describes both their common and unique characteristics, and Blackhat.com (2002), an organization devoted to promoting understanding among professionals about the security risks to their computers and information infrastructures.

Veracity: The Foundational Rule of Etiquette

Much of the speculation about human-computer etiquette is predicated on a fundamental rule of truthfulness ("Be honest"), and that notion is so fundamental that it often gets only a brief mention before the writer moves on to "more interesting" topics.

Yet so much of human-human interaction violates that assumption that we need to explore in some depth how dishonesty and deceit are making their way into that increasingly common setting for communication, the Internet.

Trust appears throughout the papers in this Symposium and is grounded in a belief that the other fellow is fundamentally honest. Considerable discussion and research has been directed to the topic of **trust** (see Castelfranchi, Falcone, Firozabadi and Tan, 2000a&b), how it is developed, lost and rebuilt. What is rarely studied in the same context are the wide range of techniques for deliberately deceiving, undermining and defrauding others and how the tension of the contradictory concepts can best be understood.

It is instructive that many of the rules of etiquette can serve both good and bad motives. Scam artists may develop considerable skill in conforming to most of the other rules of etiquette in the service of their fraudulent activity.

Imagine yourself to be a scam artist or worse, a terrorist, and think for a moment how you might use the following rules of etiquette to further your sinister motives:

- Be helpful
- Be relevant
- Be brief
- Be pleasant
- Be respectful
- Be prompt
- Protect privacy
- Provide options

Now imagine a fraudulent or terrorist software agent with an effective range of communication skills for communicating with humans and other software agents and ask yourself the same question.

Much of the detailed "Netiquette" rules (e.g., Shea, 1994; RFC 1855, 1994), somewhat more popular in the 1990s, and provided as guides to govern chat rooms, emails, newsgroups, mailing lists and the like, can be accommodated easily by the con artist in the service of both his principal goal to deceive and refinements on that clandestine motive such as: "the mark must not even know when he has been taken."

A Counter-Case for the Study of Good Etiquette

Given an arguably strong justification for examining the Internet as *the* future of most human-agent-human and agent-agent communication, and **veracity** as the foundational rule of etiquette, we are now in a better position to examine the categories of deceit pointed to in the title of this paper and through which a counter-case for the study of good etiquette can be made.

Traps – Although a physical manifestation, we do not mean "golf bunkers" or devices for preventing the passage of water. We do mean "a device for capturing or detecting a person unawares."

Browsers with implemented cookies (simple agents) are just such mechanisms and are made even more effective by web sites that prevent access unless the user accepts their cookies, a **trap** one cannot avoid if the user wants the reward of access.

Shortly before 911, a widely circulated Internet myth was that of a supposed study that examined the IQ's of US Presidents back to FDR. The description of the study included the name of the Research Institute, the names of supposed world-renown research scientists who had conducted the study and a rather professional recounting of the results, with Bill Clinton at the top of the list with an IQ of 182 and George W. Bush at the bottom with an IQ of 91. This kind of **trap** fosters myths that people want to believe and a professor friend of mine, who had sent me a report of the study, upon finding out it was bogus, gave one the best description for having fallen into this kind of **trap**, saying he had been "eagerly gullible."

That **trap** even appeared as a topic in a Garry Trudeau Doonesbury Cartoon (1 September 2001)

Pitfalls – "unsuspected dangers or difficulties." There are many dangers and difficulties in our interactions with other humans that may be transferred to machines. Certainly, viruses are dangers that may be unexpectedly lurking in email and their attachments. Luckily, there is McAfee™, Norton's AntiVirus™ and Personal Firewall™ to help make sure that most such **traps** are never sprung.

Swindles – "cheats, scams or frauds perpetrated on a person." A recent comment in the online publication, "Internet Scambusters" (Scambusters, 2002a), reads:

"In 2001, 5.2% of online consumers fell victim to credit card fraud, according to a recent survey by GartnerG2. Online merchants lost a staggering $700 million to fraud in 2001 - over 1% of total annual online sales. Online fraud losses were 19 times as high as offline fraud, according to the study."

Lies – "statements known by the originator to be untrue."

Although not having the status of a virus, some messages have some of the same effects. Masquerading as good Samaritans, they warn of viruses associated with particular files and suggest that we remove them immediately. In doing so, however, we realize too late that we have deleted a system or application file, which then prevents the system from booting or the application from running. The

insidious nature of this type of **lie** is that it often enlists well-meaning people who act as unwitting agents for the originator by sending copies of those messages to their friends.

Doubts – "feelings of uncertainty or disbelief."

This is the beginning of the collapse of **trust**, a key element in support of that fundamental rule of good etiquette: truthfulness and honest dealings. Something in a message may produce feelings of unease and uncertainty. We may not have any clear idea of the source or nature of that unease but may decide that we don't want to read some message or download some attachment, and so we delete it. Unfortunately, we may delete messages that not only are legitimate but also are important to our work.

Suspicions – "partial or unconfirmed beliefs, often about some negative aspect of a person, event, situation, etc."

After repeated **doubts**, increased **suspicions** about a set of communications can lead to a complete collapse of **trust**. A key concern of Internet retailers is that people will come to distrust "anyone they do not know" on the Internet and will no longer use the medium to buy legitimately manufactured and marketed goods. The more **traps, pitfalls and swindles** we are exposed to, the more likely this state of affairs will prevail. The statistics on **swindles**, cited above, provide considerable cause for their concern.

As is evident in the cited examples above, many contain aspects that permit classifications into more than one category in this paper title. A **swindle** or fraud may also involve a **trap**, like that in the "Nigerian Advance Fee Scam" (or, "4-1-9 Fraud"). The **swindle** lures people with the promise of millions, once "hooked" has them come to Nigeria or a border country for a final meeting, tells the person that they do not need a visa, pays off customs officials to let them into the country, springs the **trap** by telling them they are there illegally and then "bleeds" them of much of their savings with the promise to get them out of the country and avoid prosecution.

Here are a couple of paragraphs from an example of this **swindle**, which was recently sent to one of the authors of this paper:

"I wish to inform you that we have FORTY-SIX MILLION US DOLLARS ($46Million), which accrued from deliberate inflated contracts awarded by the Federal Ministry of Petroleum Resources from my Corporation, Nigerian National Petroleum Corporation (NNPC) during the past military regimes and executed by a consortium of multinational companies in the Oil Industry…

"Consequently, we humbly request your gracious assistance towards the transfer of the above stated sum into a personal off-shore account to be nominated by you…We agree to offer you 20% of the total amount involved for your assistance while 5% will be mapped out to cover expenses made in course of the transfer."

In June of 1995, an American was murdered in Lagos, Nigeria, while pursuing this scam, and several other foreign nationals have been reported as missing.

Recently, Perreaux (August, 2002), writing in one of Canada's two national dailies, reported on a business man who was roped into this scam and who, several years later, still gives the impression that he believes it to be genuine:

> *Paul Blazev admits he has thrown away $500,000. He has lost his life savings, his marriage and his tile business. He is in trouble with Revenue Canada. Still over coffee in a Calgary hotel, he dials his cell phone and sends off another $2,500. (p. A3)*

When the author recently broadcast information on this scam, the following email was received from an acquaintance:

"I must admit that I got sucked into the scam several months ago and kept a whole file on it. The scam email was also attached to the message that I received and it was the very same one that I was involved with just a different name.

"I even called a number in Nigeria and some cell phone in Spain to arrange a meeting with this fellow. The call tipped me off as the fellow I talked with was very unprofessional. I then did some research online and found out about the scam."

Exploring these categories and examples is meant to illustrate the point that a thorough understanding of etiquette is not possible without an active consideration of the many ways in which its fundamental rule can be violated and its other rules enlisted in the service of deception and fraud.

Active consideration promotes a deeper understanding than incidental reflection which often occurs when treating only the positive aspects of a subject. In active contemplation of deceit, fraud and the like, the mind is engaged to construct creative and wide-ranging violations that force broader consideration of issues like (personal) protection and security. Also, considering only positive aspects of etiquette, allows continued assumptions of benevolence to bias the character of the study.

A good example of how this works was illustrated years ago by Norman and Rumelhart (1975) who asked people to look at a building and describe what they saw, probed them until they no longer had anything else to say about the building and then asked them to imagine they were theives and re-describe the building. Subjects had no problem adding substantially to their descriptions.

Ensuring Good Etiquette in Human-Computer Interaction

Security precautions are the principal response to system abusers, whether they are applied over the Internet or some other system, and whether the abusers are human or software agents.

Security is predicated on the notion that there are people and things that you cannot **trust** and whose motives, if successfully realized, could have detrimental effects on the goals you are trying to achieve.

How then can we be sure that **veracity**, that foundational rule of etiquette, is maintained in our communications, and, as technology becomes ever more sophisticated, that it will continue to be maintained?

As local (standalone computers; TVs; LANs) and global (Internet; Intranets; WANs) systems evolve, software applications will assume greater roles in ensuring **veracity** and will take over much of what we do now and many things we likely could never do.

Many software applications already are in the process of evolving into intelligent agents that will serve entire communities of users. It can be seen that security and maintenance software such as Norton Utilities™ , AntiVirus™ and Personal Firewall™ are evolving in that direction.

Related components are being collected into single packages, such as Norton System Works™ and the Internet Security™ packages from Symantec. Live Update™ is what might be called a second-generation, industrial-strength agent now included in those packages. It provides automatic updating of virus definitions, patches and other useful elements and sits on your computer to serve your security and maintenance needs on a daily basis.

Windows XP provides automatic and periodic updating of the operating system and other Microsoft files on your computer and often it is unclear what the nature of those updates include. Such updating unquestionably is useful but other software "vendors" who may offer free software could incorporate similar updating agents that would not be as benevolent as the ones just described.

Secure systems are both defensive and offensive and the above examples are a little bit of both, relying on *active* agents to update a variety of systems with *passive* measures, such as new virus definitions that protect against unwanted attacks.

Security also has considerable scope of meaning. In addition to a protection against intruders that could undermine task competency and system integrity, it

generalizes to issues of commonly held knowledge and values.

In the ubiquitous environment of the Internet, information groups such as Scambusters (2002b) actively broadcast warnings about new frauds and **swindles** as a means of helping to protect consumers. Sites like ConsumerSentinel (2002) provide more passive information allowing users to visit the site and check on prize promotions, work-at-home schemes, telemarketing scams, identity theft and other con games. Urban legends sites such as Snopes (2002), Hoaxbusters (2002) and the like provide users with a way to check out the **veracity** of the information they receive in communications from friends, acquaintances and strangers.

The proliferation of pornography on the Internet is an unwanted intrusion to many, especially as it makes its way into classrooms and homes where children surf the Net. For several years now, governments and private institutions have grappled with how to control information that violates the ethics and values of society, while maintaining sufficient freedom of expression. The US Congress passed the Children's Internet Protection Act in 2000, which forces schools and libraries to use Internet filtering software or lose federal dollars. That law has repeatedly met with objections from private organizations and has been declared unconstitutional in its present state on two occasions by the courts (CIPA, 2002). To date, no satisfactory solution to the problem has been found. Some recent information on CIPA appears on the NTIA (2002) site.

An intriguing question is how to incorporate into an intelligent agent what these sites do to support factual truth and protect societal values. Of course, the effort is already underway through filtering agents that control access to unwanted and potentially harmful sites and through automated alerts about scams and hoaxes, but future work could examine how those functions could be combined into a personal agent and what proactive measures that agent might take against repeat offenders.

These last two areas raise interesting possibilities for the future of agent technology. Disagreements over the truthfulness of shared "knowledge" and the desirability of certain social values could set the stage for what can be termed "agent wars," where some agents seek to disseminate information that other agents have been created to prevent. Added functionality could provide agents with the ability to misrepresent themselves or to hack other agents in an attempt to prevent them from achieving competing goals. The technology that emerges from those "agent wars" will have implications for other, possibly more serious matters of security.

Veracity Agents

Given that the central concern of this paper is with **veracity** as the foundational rule of etiquette, a natural consequence is the development of one or more agents whose job it is to ensure that the communications a user receives are tested for their truthfulness.

A number of questions present themselves: What communications should be tested? What kinds of tests are possible? What are the contingencies of response? What are the vulnerabilities of other agents working on behalf of a user? What role would a **veracity** agent have in helping to maintain the integrity of those other agents? How will such an agent maintain its own integrity?

The last three questions imply the possibility of what might be called second-order **traps, pitfalls and swindles**. They address software agent-to-agent contact and response, with exchange of information occurring in the background of human communications, and out of sight of human users.

Example violations of good agent-to-agent etiquette would include the misrepresentation by an agent of its identity, the masking of its purpose, falsification of the information it is seeking and evasiveness about the information it was programmed to provide. Such deceitful behavior could occur in the presence of filtering agents responsible for protecting personal or corporate information and consequently lead to the loss of that information and other valuable assets.

Larry Fonner at MIT has described this as an "arms race of sorts," a race between good-intentioned agents who must co-exist with their "evil-intentioned" counterparts, which

seek to spy on users for underhanded and nefarious purposes (Hamilton, 1999)

In the presentation for this Symposium, more detailed ideas and work on the design and creation of veracity agents will be presented and discussed.

Project Status

Testbed work described in this report is being conducted by Artificial Intelligence Management and Development Corporation, as an internal research and development project, and is in its early stages. Implications for a wide variety of security concerns make the project interesting to a number of public and private-sector organizations. Work on **veracity** agents has begun with discussions centering on design and range of focus. Prototype work is underway.

References

AI ALERT (May, 2002). *At MIT, they can put words in our mouths*. (reported from The Boston Globe.) A semimonthly online service from the American Association for Artificial Intelligence.

Apple (August 8, 2002). *Coming Attractions: Mail*. Apple eNews.

Blackhat.com (2002). http://www.blackhat.com

Castelfranchi, C., Falcone, R., Firozabadi, B.S., & Tan, H-Y. (Eds.) (2000a). Special issue on "trust in agents." Part 1. *Applied Artificial Intelligence*, *14(8)*. Taylor & Francis: London, UK.

Castelfranchi, C., Falcone, R., Firozabadi, B.S., & Tan, H-Y. (Eds.) (2000b). Special issue on "trust in agents." Part 2. *Applied Artificial Intelligence*, *14(9)*. Taylor & Francis: London, UK.

CIPA (2002). For a copy of the Children's Internet Protection Act passed by the US Congress, see http://www.ifea.net/cipa.html

ConsumerSentinel (2002) http://www.consumer.gov/sentinel/

Hamilton, T. (September, 1999). *Task masters*. Globe and Mail report on business: technology. Canada.

Hoaxbusters (2002). http://hoaxbusters.ciac.org/

Maurer, D. W. (1940). *The big con*. Anchor Books: New York, NY.

Norman, D.A., & Rumelhart, D.E. (1975). *Explorations In cognition*. W H Freeman & Co.: San Francisco.

NTIA (2002). Request for Comment on the Effectiveness of Internet Protection Measures and Safety Policies by the National Telecommunications and Information Administration. http://www.ntia.doc.gov/ntiahome/frnotices/ 2002/cipa_52202.htm.

Perreaux, L. (August 21, 2002). *They will never see a penny: Alleged Nigerian Fraud*. National Post, p. A3.

RFC1855 (1994). *Netiquette Guidelines* http://www.dtcc.edu/cs/rfc1855.html

Scambusters (2002a) Newsletter #50, 14 May 2002.

Scambusters (2002b) http://www.scambusters.org/

Shea, V. (1994). *Netiquette*. Albion Books: San Francisco.

Smithsonian (December, 1998). *In Ponzi We Tru$t*. Smithsonian magazine. http://www.smithsonianmag.si.edu/smithsonian/issues98/dec98/ponzi.html

Snopes (2002). http://www.snopes2.com/

Zeltser, L. (2000). *The evolution of malicious agents*. http://www.zeltser.com/agents/agents.html