IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP5248

This paper appears in *Managing Modern Organizations Through Information Technology*, Proceedings of the 2005 Information Resources Management Association International Conference, edited by Mehdi Khosrow-Pour.

Software Agents: Ethical Issues Concerning Agent Autonomy

Kane Baxter Bignell

School of Multimedia Systems, Monash University, Clyde Rd., Berwick, Victoria, 3930, Australia, bignell@netspace.net.au

ABSTRACT

This paper establishes a need for software agents in an Information Age where even competent computer users suffer information overload. The social consequences of agent use is examined, that to this time have not been adequately addressed by developers or governing bodies. Thus, there is a need for the establishment of a universally accepted, effective code of ethical agent practice in order to avoid agent misuse and threats to individual privacy. Consequently the code would protect users and ensure the future of the technology itself, as people become increasingly informed and suspicious of these intrusions such as spyware agents. The author proposes an Agent Ethical Code of Conduct, noting the possible future developments of such a Code should further research be undertaken in this area.

INTRODUCTION TO AGENT TECHNOLOGY: ESTABLISHING A NEED FOR AUTONOMY

The Internet, the most complete representation of information in western society, has fundamentally changed the way society utilises and interacts with computers. Business and individuals find themselves in transition to a networked, information society. The continued digitisation of content and networking of business and private computer users forms the basis of this information society where the computer is becoming increasingly ubiquitous now we are all connected through electronic networks. Consequently, the amount of data and software applications is continually increasing and growing in complexity. Users now have an abundance of information at their fingertips, yet the consequential diversity and complexity of how this data is represented can overwhelm even competent computer users finding it difficult to productively search through the abundance of information to find relevant material. Western society's current technology related dilemma seems to be 'information overload' and the resulting stress and orientation loss.

Thus, there is a need for assistance in this highly complex, digital environment from autonomous software agents where new methods of information dissemination and filtering are required to manage this overload of information. However, due to agents' autonomous nature, there are potential hazards with their use. Because of this, agent developments must involve not only attention to technical details, but also to the ethical concerns relating to their resulting impact. In the excitement of new and emergent technology such as agents, Information Technology (IT) developers often forget to examine the impact new technology will have on users. In fact the social dimension of all technology is the driving force and the most important consideration of technology itself (Rieder, 2003).

This paper encompasses analysis of the technical aspects of agent development evolving from fields of computer science, together with ethical issues deriving from the humanities and social science fields of IT. The technical analysis of agent software is necessary to form a basis for critical evaluation of the social ramifications of agent implementation, determining whether agents are a viable solution to the well documented problems of information overload, along with how current agent development techniques will affect the way users interact with computers in the future.

EVOLUTION OF AGENT TECHNOLOGY: CURRENT APPLICATION OF AGENTS

Software agents are one of the fastest growing areas of information technology, however, they are not a new idea. They are developments stemming from research in artificial intelligence (AI) which flourished in the 1980's and 1990's. AI is a field of computer science and engineering concerned with the computational understanding of what is commonly called 'intelligent' behaviour. To this extent, software agents are currently being developed with the expectation they will one day possess the same level of intelligence as humans, to actively learn from and assist computer users. However, to this point researchers have not been successful in creating an 'intelligent' agent, as they still do not know exactly how the human brain functions, yet many authors still wrongly label this category of software as intelligent agents. It is also inappropriate to regard agents as intelligent as measuring and defining intelligence is considered especially subjective due to our own expectations, therefore it is almost impossible to label software as 'intelligent'. Bradshaw's (1997, p.16) argument that "one person's 'intelligent agent' is another person's 'smart object'; and today's 'smart object' is tomorrow's 'dumb program'" emphasises this subjectivity where intelligence is only reflected by the activity or role that it plays in assisting users. Without a direct application or use, software intelligence is meaningless.

To date the only software agents in operation use pattern recognition algorithms to assist people performing routine, or personal tasks. These agents can be seen in operation through such electronic commerce sites as amazon.com or Barnes & Noble's online bookstores (http://www.bn.com) where user demographics are analysed to give personalised service to users. Pattern recognition agents are also used for e-mail prioritisation, in education and also to aid Internet users in searching for information such as news articles. Microsoft has recently (July, 2004) launched an agent-based news filtering application they call 'Newsbot'. The agent scours 4,800 news outlets to find the most relevant news headlines that would be of interest to the user (AustralianIT, 2004a). From these applications the author identifies that there is a fundamental difference between software agents currently used over the Internet, compared to the agents which are being developed to replicate human intelligence.

SOFTWARE AGENT CHARACTERISTICS: APPLYING A UNIVERSAL DEFINITION

It has not yet been possible to agree on a universally acknowledged, comprehensive definition of a software agent, as their developments are influenced by criteria from various research fields, notably artificial intelligence and information communications. Ultimately the name 'agent' is an umbrella term which covers many different applications. Other aliases often utilised include *personal assistant agents, autonomous agents, information agents, intellibots, chatterbots, databots, knowbots, softbots, userbots, taskbots, personal assistants, avatars, personal agents* and *network agents.* The reason these different labels exist is that agents promise such diverse applications and therefore authors have been hesitant to restrict research by applying a universal

definition. At present, there is every indication that there are more definitions than there are working examples of systems that are said to be agent-based. With no definitive description, Wooldridge and Jennings (1995) fear agents are in danger of becoming a simple buzz term for new software applications, due to the term's overuse.

For the purposes of this paper, the author has constructed the following definition derived from descriptions given by several researchers. It is broadly recognised that agents are interactive entities that exist as part of an environment shared by other agents. They are conceptual entities that are autonomous, operating without direct intervention of humans or others (Russel, 1995) in a proactive or reactive manner (Wooldridge & Jennings, 1995) in pursuit of one or more objectives (Jennings et al., 1996), within an environment where other agents exist and interact with each other (Shoham, 1997) based on shared knowledge of communication and representation (Finin, Labrou & Mayfield, 1997). When a new task is delegated by the user, an agent should determine precisely what its goal is, evaluate how the goal can be reached in an effective manner, and perform the necessary actions by learning from past experience and responding to unforseen situations with its adaptive and continuous reasoning strategies. It is important to add to this definition that agents are not self-motivating and therefore are restricted to operate on behalf of users.

SOCIAL & ETHICAL IMPLICATIONS: PLACING OUR TRUST IN AGENTS, AND THEIR DEVELOPERS

From a philosophical point of view, technology itself is created by human beings, for human beings, thus any implementation should contribute to improving society's quality of life. However, the widespread implementation of agent software raises a number of issues that have clear ethical, even moral dimensions, including the balance between autonomy and control, as well as questions relating to trust, responsibility and privacy. While the value of agents performing certain tasks is undeniable, this utility perhaps comes at a certain cost that users may or may not consider acceptable. With agents acting on behalf of the user, many questions are raised as to whether the agent will always act in the correct manner in accordance with the user's objectives. These issues (outlined below) need to be addressed if agents are to have a beneficial contribution towards our networked, digital society.

Agent Trust

User-agent trust must be built through the user's confidence with using the technology. Agents are useful only to the degree which users can trust them to perform a task autonomously, without repeated, direct instruction and constant supervision. It is imperative users understand that in delegating tasks to an agent, as the metaphor suggests, the agent may not perform the tasks exactly the same way the user would have completed it, had they accomplished the task themselves. Also, the user must understand that an agent might at different points in time react differently in an identical situation. Trust between the agent and the user is an evolving state that should grow gradually as agents demonstrate themselves to be more capable (just as with people). Developers must also respect individual user differences and capabilities when creating agents, as acceptance rates will vary between users.

Delegation of Control

Trust and control are interrelated issues, in extending trust to an agent the user will be willing to give up some amount of control. What is important is that control not be taken from users involuntarily. The delegation of control could thus make the user feel intimidated by their computer as it starts making decisions on their behalf. Loeffler (1996) notes that the unpredictability resulting from significant autonomy might well result in agents who are less helpful to us than we might hope or indeed expect. Further, Reider (2003) highlights that more than in the case of any other technology before, the agent's functioning disappears into the black box, where their functions and motivations are unknown. If the user does not know the structure of the agent - the underlying algorithms, which are too complex for even an expert to adequately evaluate - users are forced to place agents in a position of authority, ultimately making them vulnerable to the agent's actions.

Ethical Responsibility

With agents possessing the characteristic to act autonomously, there is a distinct need to address the moral responsibilities of software agent programmers, as their role in our online society becomes increasing more important. Hypothetically, if an agent participates in a scheme found to defraud or deceive, questions are certain to be raised as to whether the issue of responsibility belongs to the developer of the agent, or the user who may have configured the agent for the specific application, or both. The agent developer or programmer is the easier option to place blame, as they essentially created the agent's framework and features. They can decide what, if any, safeguards to include. Also, they are in the best position to understand the problems the agent may cause and have the most significant potential to minimise this harm (Heckman & Wobbrock, 2000).

Currently, many programmers are developing agents that are not generally in the best interests of society, yet they do not come under any scrutiny predominantly due to the fact that computer programming is not considered as a profession, consequently programmers are not held to the same standards of liability that professionals are. The legal system has been reluctant to give programming a professional status. In the Fair Labour Standards Act of 1938 it states that programmers are analogous to drafters in that both perform mechanical functions (Murch & Johnson, 1999). The author believes that software programmers should be considered professionals as they possess a technical skill that should be used with great care and thought, therefore it would give them a higher duty of care to sustain the purity of society. It seems inevitable that outcomes resulting from agent decisions, which are out of user's control, precipitate the need for renewed examination, both of community, and legal understandings of liability. At present, it comes down to a moral obligation from the agent programmers to ensure that the agent-enabled applications of tomorrow act in the best interests of our networked information society.

Information Privacy

As previously described, users can delegate tasks, responsibilities, and competence to agents. Therefore, agents may need to capture specific, sensitive, personal data about their respective users so they can complete a delegated task. To protect the privacy of the persons involved, it is important that this personal data is used with care, that the information collected is totally necessary and is captured for a legitimate purpose. The user must also know that the data will not be disclosed to other sources, and it is also imperative that personal data is not processed without the knowledge of the persons concerned (Borking, Van Eck & Sieler, 1999). The idea that agents could have access to personal records and financial activities is disturbing to all computer users, no matter how helpful the agents may be.

Computer Monitoring

Currently, there are several applications on the market that utilise agent technology to enable users to monitor who is using their computer. There are differing opinions about the ethics of this type of monitoring, as an organisation could realistically monitor their employees' actions, which could be seen as a violation of employee privacy. Further, with the recent introduction of spyware agents (programs unintentionally downloaded onto a computer to monitor browsing habits, profiling shopping preferences, collect keystroke and password information, without the users knowledge) there is growing concern that agents will be increasingly implemented by companies and individuals with malicious, unethical intent.

To combat this, a privacy Standard, known as the Open Profiling Standard (OPS) was developed by internet technology leaders, Netscape and Microsoft among others, and was accepted by the World Wide Web

Consortium (W3C) in April 2002 (Cox, 2002). The Standard, which constitutes part of a larger privacy effort called P3P or the Platform for Privacy Preferences Project, specifies how personal information particular to a certain user can be stored in an Internet browser. It also specifies that information would be the property of the user, and the user is able to specify that every time a site asks for certain information. Yet many software agent developers still do not comply with this Standard as there are no legal ramifications if it is not adhered too.

DEVELOPING AN AGENT ETHICAL CODE OF CONDUCT: AGENT EFFECTS ON SOCIETY

The formulation of the P3P Standard has played an important role in aiding transparency on the Internet, providing users with more control over their computers and personal information. However, as reflected with the recent infiltration of spyware agents, the P3P Standard has not entirely been adhered to and thus does not seem adequate to curb the intent of rogue agent developers. Imposing legal legislation is another way to ensure the best interests of society are maintained through software development, yet the difficulty encountered defining the term 'software agent' makes it very hard to draft legislation that directly addresses the problem.

Nevertheless, due to the severe ramifications of spyware agents, a bill has recently (October 5, 2004) been passed by the House of Representatives in the United States (U.S.) that makes the use of spyware agents an illegal offence. If accepted by the Senate the 'I-SPY (Internet Spyware Prevention) Act' would become law, allowing the courts to send offenders to jail for up to five years (Leyden, 2004). The bill provides guidelines for technology companies which distribute software capable of most types of monitoring. It requires that consumers explicitly chose to install such software and agree to the information collected (AustralainIT, 2004b). Another bill unanimously passed (399-1) by the House days earlier, the 'SPY ACT (Securely Protect Yourself Against Cyber Trespass)' authorises the courts to issue up to $3 million (U.S.) to spyware agent violators (Mark, 2004).

However, the author further believes it should not be entirely up to governments to impose these restrictions on software development, where a stronger emphasis on ethical responsibility of developers would be a much more constructive approach. While legislation can set out the framework of requirements for privacy and protection, self-regulatory regimes such as codes of conduct can provide critical flexibility as they can be dynamic and responsive to changes in technology in ways legislative processes fail. If the mentioned spyware agent Acts are passed by the U.S. Senate and become law, which may not be until early next year (2005), then the bans against spyware would only begin 12 months after the bill is approved and would automatically expire after 2009 (AustralianIT, 2004b). Further emphasising the inflexibility of legislation is the fact that if spyware agents are developed and programmed in another country which crosses jurisdiction from the U.S., then the offenders cannot be held accountable unless their respective country imposes the same legislative enforcement.

Thus, an Agent Ethical Code of Conduct (AECC) has been proposed by the author to assist agent software developers in understanding the importance of placing the user, and society as a whole, as the priority of any agent application. The goal of this Code is to protect the interests of Internet users and ensure the developers do not see agent software as a means to exploit the potential to capture sensitive, personal information from users. The Code has been drafted by the author upon review of several software user license agreements that often include terms and conditions which would be difficult for most users to understand. The Code was further developed from analysis of the current operations of spyware agents analysing how companies have been able to install this software on user's computers without their knowledge. A further goal of this Code is to establish a means of quality control so that ideally when software is published it would include some verification that the software complies with the AECC, giving users assurance that the software will act in their best interests.

Elements of the Agent Ethical Code of Conduct (AECC)

- **Agent Disclosure** Once installed and operational, the agent software must present a simple, accessible, concise explanation of the software's purpose, intent and use of the user's information.
- **Vendor Interaction** Agent software developers can in no way interact with the user's computer to gather information without notice or consent other than for the purpose of determining whether a user is an authorised user of the agent software, or to perform software updates.
- **No Unnecessary Information Gathering** No information other than what is required for achieving the purpose of the delegated task may be communicated. After collection, information may not be retargeted or reused for any other purposes other than those specified by the user.
- **Formal Online Privacy Statement** The developer of any agent technology must provide a comprehensive privacy statement, completely and clearly specifying all issues in regard to the ability of the agent to capture information and how that information will be used. This statement must detail the limitations of the agent along with the potential dangers of agent use (if any).
- **Pre-emptive Request for Consent** Agent entities must ask the user for permission to transmit sensitive data over the Internet to other agents or third-party organisations, specifying exactly what information is to be passed over the network.
- **Complete Control** Users must have complete control over the software agents installed on their computer and over the use of their Internet connection. Users must have the ability to remove agent software such as spyware from their systems for any reason and at any point in time. This would eliminate the current problems where users are unable to locate and delete invasive software agents from their computers.
- **Ability to Trace Steps** Users must be able to access the path taken by agents to complete their delegated task. This may include information retrieval techniques, or automated processes such as conducting financial transactions. The way agents operate cannot be perceived as a black box by users, therefore they must be able to see the methods used by their respective agent in order for them to gain trust in giving control to the agent to perform further tasks.

With Continued research the AECC would ideally be further expanded and specialised to target specific aspects of agent development, particularly concerning spyware agents. Further expansion of this Code through contribution and review from IT experts, ethicists and consumer advocates would give this proposal further credibility ensuring it reflects the current digital climate. Ultimately, after rigorous review and approval from third parties including software developers themselves, the author intends to issue the Code to the W3 Consortium for consideration for inclusion to the P3P, possibly expanding upon, or replacing the Open Profiling Standard.

CONCLUSION

While agent technology has the potential to be useful, many fundamental problems, both technical and social or ethical, remain to be solved. It is imperative these issues be addressed so that computer users do not continue to be unsuspectingly exploited. This agent technology is designed to automate computer processes for users, making it easier to gather desired information, yet at present, with the introduction of spyware, it is making users extremely exposed and vulnerable as a networked, information society.

To combat the continued infiltration of spyware and other malicious agents, the author has proposed a new branch of applied computer ethics, in the form of an Agent Ethical Code of Conduct be employed which addresses the profound implications of the prospect of agents roaming our networks. The proposed Code reflects the motivations in agent design and attempts to combat many of the underlying ethical issues relating to agent application. Agent developers should evaluate the features of the agent applications in terms of user welfare, where users

should be treated as an end and not as a means to profit or information. It is obvious that the technology has made significant progress but has much further to go before it should be accepted as a tool that improves the quality of people's experiences when interacting with computers.

REFERENCES

AustralianIT (2004a) *'Microsoft unveils Newsbot beta'*, http://australianit.news.com.au/common/print/0,7208,10269634%5E15306%5E%5Enbv%5E,00.html [Accessed 2004, September 12].

AustralianIT (2004b) *'US Approves Spyware Fines'*, [online] Available: http://australianit.news.com.au/common/print/0,7208,10997813^15318^^nbv^,00.html [Accessed 2004, October 07].

Borking, J.J., Van Eck, B.M.A. & Siepel, P. (1999) *'Intelligent Software Agents: Turning a Privacy Threat into a Privacy Protector'*, Information and Privacy Commissioner, Toronto, Ontario, Canada.

Cox, B. (2002) *'Finally, Agreement on P3P'*, [online] Available: http://www.internetnews.com/dev-news/article.php/10_1010361 [Accessed 2004, August 21].

Finin, T., Labrou, Y. & Mayfield, J. (1997) *'KQML as an Agent Communication Language. In: Software Agents'*, J.M. Bradshaw (Ed.), Menlo Park, Calif., AAAI Press, pp. 291-316.

Heckman, C. E. & Wobbrock, J. O. (2000) *'Put Your Best Face Forward: Anthropomorphic Agents, E-Commerce Consumers, and the Law'*, [online] Available: http://www-2.cs.cmu.edu/~jrock/pubs/agents-00.pdf [Accessed 2004, August 21].

Leyden, J. (2004) *'House Approves Anti-Spyware Bill'*, [online] Available: http://www.theregister.co.uk/2004/10/08/us_anti-spyware_laws/ [Accessed 2004, October 6].

Loeffler, C. E. (1996) *'Artificial Life of Agents'*, Advanced IT Tools, Proceedings of the IFIP World Conference on IT Tools, Canberra, Australia, Sept. 2-6, Chapman & Hall.

Maes, P. (1998) *'Intelligence Augmentation; A Talk With Pattie Maes'* Edge Magazine [online] Available: http://www.edge.org/3rd_culture/maes/ [Accessed 2004, March 21].

Murch, R. & Johnson, T. (1999) *'Intelligent Software Agents'*, Prentice Hall PTR, Upper Saddle River, New Jersey.

Rieder, B. (2003) *'Agent Technology and the Delegation-Paradigm in a Networked Society'*, The London School of Economics & Political Science [online] Available: http://www.lse.ac.uk/collections/EMTEL/Conference/papers/Rieder.pdf [Accessed 2004, March 21].

Shoham, Y. (1997) *'An Overview on Agent-oriented Programming. In: Software Agents'*, J.M. Bradshaw (Ed.), Menlo Park, Calif., AAAI Press, pp. 271-290.

Tuohey, J. (2004) *'Is Distributing Spyware a Crime?'*, Medill News Service, [online] Available: http://www.pcworld.com/news/article/0,aid,118105,00.asp [Accessed 2004, October 11].

Wooldridge, M. & Jennings, N. (1995) *'Intelligent Agents: Theory and Practice'*, Knowledge Engineering Review, June, p. 10.