Agent Abuse: The Potential Dangers of Socially Intelligent Embodied Agents

Chris Creed
School of Computer Science
University of Birmingham
Birmingham, UK B15 2TT
cpc@cs.bham.ac.uk

Russell Beale
School of Computer Science
University of Birmingham
Birmingham, UK B15 2TT
r.beale@cs.bham.ac.uk

CHI 2006, April 22–27, 2006, Montreal, Canada.
ACM 1-xxxxxxxxxxxxxxxxx.

Abstract
Research into developing socially intelligent embodied agents has increased over the last decade with the main focus being on how they can enhance human-computer interaction. However, little research has concentrated on the potential they have to manipulate our behavior for unethical purposes. A discussion is provided highlighting the main dangers associated with embodied agents. Suggestions for reducing these risks are then provided, along with a brief discussion regarding the need for further research.

Keywords
Interface agents, embodied interfaces, agent abuse.

ACM Classification Keywords
I2.11 [Artificial Intelligence]: Distributed Artificial Intelligence – Intelligent agents. H1.2 [Models and Principles]: User/Machine Systems – Software psychology.

Introduction
Most research related to the use of embodied agents has tended to concentrate on the benefits that such agents might bring to an interface and how they can arouse positive emotional states that enhance cognitive functions (e.g. learning and problem solving). Very little

research has focused on the negative impact that embodied agents have on an interaction. As they are becoming more socially intelligent, there is the increased possibility that they will be able to 'abuse' us in a number of ways.

This position paper will discuss the main issues surrounding this possibility. A brief overview of recent work which has examined human-computer relationships is provided, along with an outline of the dangers this gives rise to. Suggestions for reducing these risks are then provided, as well as a brief discussion regarding the need for further research.

Social-Emotional Human-Agent Relationships

A large number of studies have suggested that we seem to treat computers as social entities [5]. This has motivated a number of researchers to investigate how we can make use of social skills in human-human interaction and use them in HCI to enhance human-computer relations. For example, Bickmore and Picard [1] investigated whether embodied agents can build and maintain long-term relationships with computers by making use of the many relational strategies that humans often use (e.g. small talk and talk of the relationship). They found that people generally liked and trusted agents more that made use of such strategies over agents which did not.

Many other recent studies also appear to be suggesting a similar trend in that we seem to prefer interacting with embodied agents that have some form of social intelligence. This is despite the fact that the level of intelligence demonstrated by such systems is very limited in comparison to our own. However, with computer processing speeds doubling every year, many believe this ability is likely to change drastically in the near future. Kurzewil [3] predicts that by 2010 we will have virtual humans that look and act much like *real* humans, although they will still be unable to pass the Turing Test. By 2030, he believes that it will be difficult to distinguish between virtual and biological humans. This potential increase in agent intelligence and representation raises a number of troubling issues.

Potential Dangers

Our tendency to treat computers as social actors [5] suggests that socially skilled agents may be able to utilize many of the strategies and techniques that humans use to manipulate other peoples' behavior. For example, in human-human interaction, we tend to act on the advice of people we like and trust rather than people we dislike and distrust. It is possible that the same principle might apply in HCI; as mentioned above, a range of studies have suggested that we like and trust socially skilled agents over ones which have no such skills. Therefore, these agents may be able to manipulate human behavior more effectively than agents with no social skills built into them (e.g. [4]).

Socially intelligent agents also have a number of advantages over humans when attempting to manipulate our behavior, including their ability to persistently make use of a wide variety of persuasive techniques without ever becoming tired or deterred (e.g. asking somebody to register for a product every time they start up their computer). They can also make requests at times when it is more likely that the request will be complied with (e.g. a computer game or product that asks children to provide personal details before being able to progress to the next stage).

In some circumstances, users may also trust computers more than they do other humans. Whether deserved or not, some professions have a reputation for being manipulative and deceptive (e.g. salespeople) and people often tend to be cautious when interacting with such people. However, if users were to interact with a computational sales agent, they may drop their guard and be more open to manipulation as computers generally do not have a strong reputation for deception and attempting to manipulate peoples' behavior.

Is it acceptable for agents to manipulate (perhaps deceive) people in this way to, for example, help companies sell more products? Perhaps so, as long as the user believes that they have received good value for their money and do not feel exploited. Human salespeople often present the products they sell in their 'best light', even when they are fully aware that the product may have certain features which are not desirable for the customer. This is a form of manipulation (and deception), and most people are aware that many salespeople are like this. While this may not please people, they are unlikely to mind if they feel they have received value for money and a good service. On the other hand, if customers feel cheated they will be unlikely to return with their money again.

As embodied agents' social skills improve over the coming years, the danger of them being used to manipulate our behavior will increase. In fact, there are many embodied agents available today that attempt to manipulate peoples' behavior in questionable ways. For example, Fogg [2] highlights TreeLoot.com as one such website which employs embodied agents to use a number of social strategies (e.g. displaying *positive* emotions toward to the user) to keep people playing their game and to encourage them to visit their sponsors. The success of agents such as these is yet to be empirically tested, but the potential for them to manipulate user behavior certainly exists.

As we move more towards managing computer systems rather than directly manipulating them, we will work more closely with agents in everyday activities as they undertake tasks on our behalf. This means that people are likely to develop long-term relationships with agent entities in their interactions, who they will grow to know and trust. It may be that these agents are then in a very strong position to alter their behavior and start becoming more and more manipulative over time (like a cult: nice to begin with, drawing a person in, and then changing and starting to abuse the trust that has been created). This may happen by initial malicious design, or more intriguingly, by external people 'attacking' an agent and making it turn on its user! A new form of virus writer may emerge.

Suggestions for Reducing Risks

People need to be warned about the potential dangers associated with agents which attempt to manipulate their behavior and what evasive steps can be taken. They also need to be taught about the different persuasive strategies that computers can utilize and how they should respond to them. Users must also take responsibility for their actions. Just as they would when interacting with a human salesperson, people need to be aware of any subtle manipulation that is taking place and must adjust their behavior accordingly. This may prove difficult for users initially because of the novelty factor associated with embodied agents and the perception of them being 'fun' and 'entertaining' to interact with.

Preventing children from being manipulated by embodied agents will be more problematic. Again, education and raising awareness about the potential dangers is fundamental, but children may be more likely to overlook these dangers than adults. Some form of monitoring body may need to be introduced in the future to assess online content and entertainment products aimed at children, to ensure that no unethical manipulation is taking place.

Design of agents is also a key issue. A balance will need to be found between an agent performing its tasks effectively (which will likely involve attempts to manipulate user behavior) and not taking excessive advantage of users. This will become increasingly difficult to achieve, but it is essential that designers consider the social skills, strategies and techniques that their agents use to fulfill their goals. Introducing an ethical code of practice that designers and producers of agents sign up to may also help reduce some of the main risks associated with socially intelligent agents.

Conclusion

To understand further the extent to which our behavior can be manipulated by embodied agents, it is imperative that a number of areas be researched in detail. We need to understand more clearly exactly what approaches agents can use to manipulate our behavior and how effective they are. Whilst it may not seem the natural course to take, it is important to study the unethical implications of embodied agents. Can they persuade users to spend more money? Can they influence which candidate we decide to vote for? Are children more likely to give their personal details to a socially intelligent embodied agent that claims to be their 'friend'?

This type of research will not only help us understand how users can be manipulated for unethical gain, but also how agents might be able to manipulate user behavior for beneficial purposes. In fact, we are currently investigating whether emotional embodied agents can help motivate people to eat more healthily than unemotional agents and are looking to conduct our initial experiments over the coming weeks.

It is vital that we begin studying in more detail how socially intelligent agents can manipulate our behavior. Other issues also need to be debated and discussed, such as finding a balance between an agent effectively performing its duties and not taking advantage of a user. A deeper understanding of these areas will enable us to take steps toward avoiding agent abuse against users, both now and in the future.

References

[1] Bickmore T, Picard R. Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Transactions on Computer-Human Interaction (TOCHI) 12*, 2, (2005), 293-327.

[2] Fogg B. *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufman, San Francisco, USA, 2003

[3] Kurzewil R. *The Singularity is Near*. Penguin, New York, USA, 2005.

[4] Picard R, Klein J. Computers that recognise and respond to user emotion: theoretical and practical implications. *Interacting with Computers 14*, 2 (2002), 141-169

[5] Reeves B, Nass C. T*he media equation: How people treat computers, televisions, and new media like real people and places*. Cambridge University Press, New York, USA, 1996.