Can I trust you? Sharing information with artificial companions

(Extended Abstract)

Matthias U. Keysermann
MACS, Heriot-Watt University
Edinburgh, UK
muk7@hw.ac.uk

Ruth Aylett
MACS, Heriot-Watt University
Edinburgh, UK
r.s.aylett@hw.ac.uk

Sibylle Enz
Otto-Friedrich Universität
Bamberg, Germany
sibylle.enz@uni-bamberg.de

Henriette Cramer
Mobile Life @ SICS
Kista, Sweden

Carsten Zoll
Otto-Friedrich Universität
Bamberg, Germany

Patricia A. Vargas
MACS, Heriot-Watt University
Edinburgh, UK

ABSTRACT

This paper discusses an experiment to investigate issues of trust and confidentiality when sharing information with a robot companion in an office context.

An online questionnaire was used to collect opinions about information sharing with a robot companion and preferences for collection and treatment of information. In a subsequent live interaction study, subjects role-played new members of an office team exchanging potentially sensitive information with the robot companion. Evaluated results and their implications are summarised and we suggest generic improvements for HRI systems used for information exchange.

Categories and Subject Descriptors

H.5.m [**Information Interfaces and Presentation**]: Miscellaneous

General Terms

Experimentation, Human Factors

Keywords

Human-Robot Interaction, Trust, Information Sharing, Privacy

1. INTRODUCTION

Trust has long been a significant topic in software agent research [6, 2], relating to topics such as reliability, transparency and provenance in information exchange. However, in the context of embodied agents such as robots, more social reactions come into play [1, 3, 7].

Autonomous agents need information to successfully deliver their services. Such autonomous behaviour, however, is in direct contrast with principles such as user control, privacy and transparency [5], raising the issue of user trust.

This paper relates to a study of trust in the context of long-term robot companions being developed in the LIREC project[1]. This is work in human-robot interaction [4] in which robots are no longer merely machines for achieving tasks but become social actors in real-world human environments.

If a companion is not purely a personal one and interacts with more than one user, then it may hold information relating to one user that ought not to be relayed to another. For these reasons we carried out an experiment looking at issues of trust and privacy when sharing information with a robot companion. To gather general opinions on this topic we first conducted an online questionnaire study; this was then followed by a practical study in which subjects interacted directly with a robot companion.

2. STUDY 1: QUESTIONNAIRE

The questionnaire was designed to address the following research questions:

1. In which situations is the companion considered helpful?
2. Which kinds of personal/team-related information are people willing to share with the companion?
3. What are people's preferences regarding: the collection of information through the companion; the disclosure of information through the companion; interaction modalities with/control of the companion?

The questionnaire provided the participants with a selection of options regarding various types of information to select from and offered them the opportunity to provide additional input via free text responses.

The most important findings were that the companion is considered particularly helpful when working on a joint task, during absences from team, and when working in separate rooms in the same building. According to these functional preferences, information that people are particularly

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), 4-8 June 2012, Valencia, Spain.

[1] http://lirec.eu

willing to share with it concerns meeting dates, important tasks and deadlines, as well as absences from team. Information collected should be reported when collected or only collected when indicated. Before accessing other sources of information (like social networks or other personal internet resources), the companion should ask for permission. As far as disclosure of work-related information is concerned, it should only be disclosed either to classified persons or to team members, but not to others, and it should only be given after authorisation. Personal and private information should never be given away. Information secret to the requesting person should not be given; the companion should indicate that it has no authorisation to give it.

It is very important to control the memory content and treatment of information; however, interactions to exert control should only occur at medium levels of frequency. In order to minimise these interactions over time, sharing preferences should be chosen more and more autonomously by the companion based on previous choices taken by the user. When requesting information, the option to only receive a summarised output of new and currently relevant information can enhance comfort as it provides a quick and effective way of keeping users informed.

3. STUDY 2: LIVE INTERACTION

The second part of the experiment involved live interaction with an actual companion in order to examine how far the issues raised in the questionnaire were translated into interactions within a specific scenario and with a real robot.

The participants were asked to imagine they were a new team member in a team of researchers working on different projects. The goal for the participants was to get to know as much as possible about the other team members and current projects so as to familiarise themselves with their new coworkers and workplace. In the scenario, none of their team members were in the office so that the companion was the only source of information. Participants were able to ask the companion for different kinds of information about the team, and – where considered appropriate – give information about their own role.

A tablet computer was used for requesting information and entering personal data. It offered a simple interface developed in HTML; once logged in users navigated pages using touch buttons and typed on the touch keyboard. The companion responded verbally using a text-to-speech programme that outputs a human-like unit selection female voice.

It could be noticed that not much information was requested about other team members compared to the information the companion held in its memory. We attribute this, along with the limited information participants supplied about their role, to limited engagement of the subjects in the role-play. It was likely that few or none had ever been in the position of entering a new workplace team in real life. We rule out lack of confidence in the system as the reason because participants reported that they felt comfortable when providing information.

Their responses showed that participants gained in trust for the companion because they were informed about who their information could be disclosed to as well as being able to change these authorisation levels. This confirms that data transparency and control over data are very important and should never be neglected.

4. CONCLUSIONS

Control over the information collected is a priority for the great majority of our participants, confirming other research into the importance of transparency, control and user's trust [3]. Expectations, and effectively interactions with the agent, will teach the user which information will or will not be available. In our live interaction study we found the first interactions with the agent set the tone for the further information requests. While participants felt the companion did not give away information they did not expect it to give away, they did have assumptions not necessarily matching the realities of the agent's functionality. Transparency as well as actively finding out the preconceptions potential users hold and ways to counter misconceptions will be crucial.

Allowing for user control, for example implies asking before autonomously trying to collect information from other sources and clearly indicating which information is collected and why. However, participants also indicate they do not want to spend a lot of effort managing control mechanisms. Adaptation to their personal preferences without the need for explicit user input, and taking into account context (such as who else is present possibly 'overhearing' information and reasons why information is requested) would be useful. However, such adaptivity would also be in direct contrast with the control users desire. A clear approach to such issues has not been devised by the research community.

5. ACKNOWLEDGMENTS

This work was partially supported by the European Commission (EC) and is currently funded by the EU FP7 ICT-215554 project LIREC (Living with Robots and Interactive Companions). The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

6. REFERENCES

[1] T. W. Bickmore and J. Cassel. Relational agents: A model and implementation of building user trust. In *Proceedings of SIGCHI*, pages 396–403, 2001.

[2] C. Castelfranchi and R. Falcone. Trust and control: a dialectic link. *Applied Artificial Intelligence*, 14(8):799–823, 2000.

[3] H. Cramer, V. Evers, M. van Someren, and B. Wielinga. Awareness, training and trust in interaction with adaptive spam filters. In *CHI '09*, Boston, USA, 2009.

[4] K. Dautenhahn. Socially intelligent robots: Dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):679–704, 2007.

[5] A. Jameson and E. Schwarzkopf. Pros and cons of controllability. In *Proceedings of AH '02*, pages 193–202, 2002.

[6] J. Lee and K. See. Trust in automation: Designing for appropriate reliance. *Human Factor*, 46(1):50–80, 2004.

[7] B. Reeves and C. I. Nass. *The media equation: How people treat computers, television, and new media like real people and places*. University of Chicago Press, 1996.