Emotional Computing and the Open Agent Society

Abe Mamdani, Jeremy Pitt, Asimina Vasalou, Arvind Bhusate
Intelligent Systems & Networks Group
Dept. Electrical & Electronic Eng.
Imperial College London
SW7 2BT, UK
{a.mamdani, j.pitt, a.vasalou, a.bhusate}@imperial.ac.uk

Abstract

This position statement outlines a programme of research in computing with emotions in a certain type of distributed multi-agent system that we call an open agent society. We summarise recent results in agent-mediated human-human interaction, human-agent interaction, and agent-agent interaction, describe a platform for estimating an emotive state from physiological signals, and consider a number of short-term and long-term challenges.

Keywords: Emotions in man-machine interaction and interfaces; Emotional computer agents.

1 Introduction

The boundary between virtual reality, electronic territories or *cyberspaces* [5] and "the real world" is being increasingly blurred by Web 2.0 – facilitating commerce, interactivity and collaboration online. The main research issue though remains valid: how do we represent, and interact with, pure data, computational processes, and each other, in an Internet which has effectively ceased to exist as a meaningfully distinct entity (e.g. "the information superhighway"), whereby the offline and online 'self' has entirely converged, and there is now a wide range of haptic modalities available as interaction channels?

The idea that not just virtual spaces, but also physical spaces, could be transformed and treated as electronic environments. has led to innovative concepts such as territory-as-interface and inhabited information spaces, and pioneering research in connected communities and futuristic schools [8]. Ubiquitous computing [20], meanwhile, has focussed on the software and hardware technologies required to realise such concepts, creating physical environments which are saturated with computing devices and wireless communications, yet appear to be seamlessly integrated to the human user(s), and indeed, seamlessly integrate the human users themselves.

However, research in ubiquitous computing also needs to take into consideration the networking problems that arise from purely local information, partial knowledge and inconsistent union; and from decentralised control. What each network node sees is the result of actions by (possibly millions) of actors, some of which are not known, and whose motives may also be unknown. Furthermore, there is no single central authority that is controlling or coordinating the actions of others. The emphasis is on local decision-making based on locally available information and the perception of locally witnessed events.

We now have ubiquitous computing and communication systems where there are unpredictable components (i.e. humans 'in the loop'), which, as a consequence of this, and coupled with uncertain information and decentralised control, will inevitably operate sub-ideally. However, the very nature of such systems means external intervention to re-

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 1575–1582
Torremolinos (Málaga), June 22–27, 2008

store correct operation is beyond their functionality: in other words, it requires its own autonomic (self-repair) capability. Our argument here is that one possible way to provide such an autonomic function is from the development and deployment of systems that can register, understand, respond to, manipulate, and even display emotional states (cf. [11, 7]).

This position statement outlines a broad spectrum of research in computing with emotions, complementary to a line of research in norm-governed multi-agent systems which make up what we call the open agent society [1]. We summarise recent results in three dimensions of interaction: agent-mediated human-human (emotional) interaction in Section 2, human-agent (emotional) interaction in Section 3, and agent-agent (emotional) interaction in Section 4. In a fourth dimension, we describe a platform for estimating emotive state from physiological signals in Section 5. We conclude in Section 6 and consider a number of short-term and long-term challenges.

2 Agent-Mediated Interaction

Web 2.0 makes it possible to conduct everyday economic and social activities online, with a far larger pool of possible collaborators, and in far more unexpected ways, using applications whose designers had perhaps neither intended nor imagined for these purposes. Despite the increase in 'degrees of freedom', there are inherent risks, especially in interactions between anonymous parties in one-off encounters, where a simplistic game-theoretic analysis might conclude that the optimal strategy was to defect.

Previous research has focused on designing computer-mediated communication (CMC) systems which encourage users to act in a reliable manner: formal mechanisms inspired by socio-cognitive concepts such as trust and reputation are therefore built into the interface. Nevertheless, either wilfully, accidentally, or by necessity, the outcome of an interaction can be unfavourable to one or other party (or even both parties). In cases of accident or exigency, though, it is not good enough to zero the trust rating, damage the reputation of the other party, and/or seek redress in a court of law. The system ought to be capable of self-repair, and trust alone will not achieve this. Instead trust breakdowns need to be redressed by some other mechanism, and we argue that one such mechanism can be developed by considering the reciprocal emotional response that occurs as a result of a trust break-down.

We have followed two line of investigation in this respect, one in self-awareness, and one in forgiveness. In particular, based on a review of the relevant psychological literature, we contend that lower self-awareness fostered in certain online settings hinders an offenders experience of shame, guilt or embarrassment; and as a result of this, rather than anonymity (which actually has some beneficial effects), the offender is motivated to perpetuate the anti-normative behaviour.

We have conducted an experiment which substantiates these theoretical conclusions by considering the effect of interface mechanisms that activate self-conscious emotions which in turn motivate the reversal of offensive behaviour [18]. For example, private self-awareness can be manipulated with evaluative cues: this concept was built into an avatar which autonomously expresses visible embarrassment/shame upon its owners offence. When expressing these emotions, the avatars gaze shifts sideways, the posture droops and the face blushes (see Figure 1). The sequence and type of animations chosen are based on expressions that are reported as common for both embarrassment and shame [9].

Figure 1: Emotional Avatar

In a second line of investigation, again based on the relevant psychological literature, we contend that forgiveness is a putative mechanism that repairs trust. We have then developed a formal framework for forgiveness which is used to construct an intelligent intervention mechanism which informs the victim of reasons for forgiveness [19]. We conducted an experiment that hypothesised and showed that systems designed to stimulate forgiveness, such as the forgiveness intervention mechanism, can restore a victim's trust in the offender [17].

We propose in the formal framework that there are four main factors that motivate forgiveness, each of which is comprised by several constituents (ten consituents in all). The judgment of the offence includes the severity of the current offence, the frequency/severity of previous offences and the offenders intent. The offenders efforts to repair contain apologies and reversal of the action. Empathy experienced from the victim towards the offender is motivated by apologies, the offenders visible acknowledgment, familiarity and similarity between the two parties. Finally, beneficial historical relationship with the offender includes frequency and utility of prior benefits. The forgiveness mechanism was implemented using Fuzzy Logic, as illustrated in Figure 2. Space constraints preclude a fuller exposition but all the details can be found in [17].

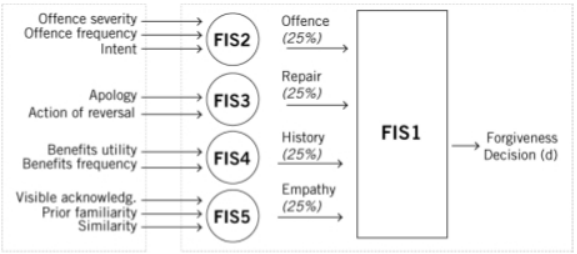

Figure 2: Fuzzy Forgiveness System

The particular import of this research is the observation that the design and development of CMC systems for Web 2.0 are too important, and too sensitive, to be left to technologists alone. An inter-disciplinary approach embracing psychology, philosophy, technology, etc. is necessary, in order to shape online behaviour in such a way that the the community will thrive.

3 Human-Agent Interaction

In addition to adding greater depth and richness to computer-mediated communication, emotional computing can enhance interaction with otherwise quotidian objects, if they are laced with sensors and 'wired' (typically of course, connection will be wireless) into a ubiquitous computing environment. We have been looking at one specific situation, in the context of London's Science Museum.

The Science Museum, like any other public collection of cultural heritage, faces four major challenges in maximising the potential of its resources. These are:

- missed opportunities in exhibition delivery: there is an opportunity to leverage technological advances in, for example, sensor networks and multi-agent systems to provide a more intelligent and/or opportunistic approach to tracking, planning and coordinating exhibits and exhibitions according to anticipated visitors (e.g. school groups);
- missed opportunities in the visitor-museum relationship: there is a tendency for each visitor to think in terms of 'doing', say, the Science Museum on a single trip of a single day. Yet, as stated, there is still too much information to be comfortably assimilated on a single visit to the museum;
- missed opportunities in the visitors-museum relationship: there is signicant benefit to be had from the social aspects of the experience, i.e. the shared experience of groups of school-children which can improve learning and recall. At present there is limited scope for sharing group explorations of the museum;
- missed opportunities in the visitor-exhibit relationship: in any visit, there will be certain exhibits that especially capture a visitors interest and there will be occasions when the user has direct ex-

perience of exhibits. This interest and knowledge is potentially valuable to both museum and visitor but is currently unrecorded.

We therefore seek to leverage computing and communication technologies in order to enhance the quality of experience of a visitor to the museum. To this end, we have developed the concept of the Virtual Clipboard [2], and also an interactive exhibit.

An interactive exhibit moitors, in hands-on interactive mode (meaning it can be held and manipulated by the user/visitor), the user's interactions with the exhibit. For example, our prototype exhibit is a car, and the sensors send information wirelessly about the angle the car is being held at, if the wheels are being spun, if the doors, bonnet or boot have been opened, if the interior seats have been touched etc. These interactions will flag for information to be given (in the form of text, video or sounds) on a digital plaque. As well as this information, there will also be a 3D computer graphics generated model of the exhibit which will mimic the interactions which are taking place. For example, if the passenger door of the car is opened on the actual exhibit, the passenger door of the CG model will also be opened. The model and exhibition space is illustrated in Figure 3.

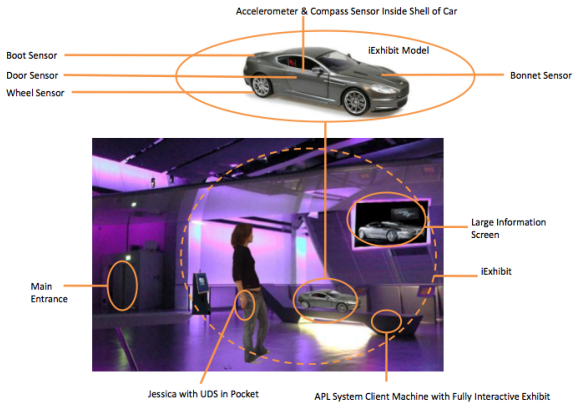

Figure 3: Interactive Exhibit

In browsing the museum, the visitor sees a large screen displaying some information about the car and an actual scaled replica model on a display stand. The visitor reads the information about the car on the large screen and then decides to pick up the iExhibit Model (iEM) car. The visitor now starts to browse and just as she touches the car basic instructions are provided of how to explore and interact with the iEM to obtain specific information about it.

In exploring the iEM hands-on, the visitor moves and rotates the iEM, whille the movement of the virtual model on the screen as she does this. She should realise that varying the position of the iEM has an effect on the information that is displayed. She holds the iEM so the driver door is facing her, and she receives information regarding, e.g. the central locking system on the screen.

The system currently exists as a stand-alone prototype. We intend to run some usability experiments with school-age children in a neutral setting, before running situated experiments in a real setting like a museum.

There are also three further avenues of investigation we are considering to follow. Firstly, we can examine how user-specified policies can be used to guide the user and configure the information being displayed on the screen. Secondly, we are considering how the system could dynamically adapt to the user and even propose new manipulations that could be interesting (for example, based on actions and profiles of similar previous users). These proposals could possibly be made by some kind of avatar, representing an electronic docent (a museum helper). Thirdly, we propose to examine how affective computing (in terms of inferring emotive state from physiological signals – see also section 5) can be used to condition the interaction and enhance the overall quality of experience.

In this context, we here note the potentially constructive relationship with the work of [16], which is (also) concerned with the use of fuzzy systems and affective computing to determine and feedback stress level of people in their ambient environment. Their paper presents the development of a (preliminary) body sensor network combining accelerometers and sensors for galvanic skin response,

which can be used to estimate activity and affective state respectively. This could be useful in the museum context where interaction with the exhibit has potentially few states.

4 Agent-Agent Interaction

The use of socio-cognitive concepts like trust and reputation to secure fine-grained, intelligent, experiential control over interactions with (initially unknown) second- and third-parties has also been well-established in multi-agent systems research. The idea behind *digital blush* [12] was to complement such mechanisms with additional emotive concepts like shame and embarrassment.

The principal idea was to use Castelfranchi's [3] model of cognitive evaluation (as opposed to affective appraisal) concerning shame, and operationalise this model in the context of an appropriate agent architecture and corresponding communication protocol. Actions out of turn in the protocol, and an awareness that these actions would be poorly evaluated by others, would lead to a 'feeling' of shame or embarrassment by the offending agent.

Castelfranchi argues that there are two aspects of emotions that need to be represented in computational models and uses of emotions. The first is that human emotions are not simply reactive mechanisms: they have particular cognitive components, are characterised by rich and complex mental states, and are activated by 'evaluations' of those cognitive components in certain mental states. The second is that human emotions are also felt: furthermore, that feeling should be the source for drives to be satisfied, e.g. to return internal states to some homeostatic equilibrium through intended action, and the emotional internal states should work as positive or negative internal reinforcements for learning.

Castelfranchi then suggests that an agent i should be ashamed in front of agent j for some proposition ϕ (where ϕ is something like (**has** i p), (**did** i p), (**is_a** i p), etc.), if i has the following beliefs, that:

- ϕ is negatively evaluated (i.e. by i itself);
- j has a negative evaluation of ϕ (so i and j have shared values and evaluation of ϕ);
- that j believes that i saw to it that ϕ;
- j has a negative evaluation of i (for doing/seeing to it ϕ);

so i's goal of positive social image is threatened by its awareness of low peer esteem.

Space constraints preclude a full exposition of the operational model, but in [12] we used the BDI (Belief-Desire-Intention) agent architecture [14], together with Searle's speech act theory of communication [15], in particular the idea that in certain contexts particular agents are empowered to establish conventional facts by the performance of designated actions.

As far as the architecture is concerned, we have a belief database and a desire (goal) database: we write $\Delta_a \vdash \phi$ to denote that agent a's current program state proves ϕ, and $\Delta_a \rightsquigarrow \phi$ to denote that a's current program state has, as a persistent goal, that in some future program state $\Delta_a \vdash \phi$. Then we introduce the modal operator [**eval** a ψ]$\pm$ for the act of evaluation with either a positive or negative outcome, to which we will give this a subjunctive reading: "*were* agent a to perform an evaluation of ψ, then it *would* be positive/negative". Then $\mathbf{eval}(a, \psi, +)$ and $\mathbf{eval}(a, \psi, -)$ represent the fact that agent a has performed an evaluation of ψ and evaluated it positively or negatively respectively. (Note that ψ could be a proposition or an agent.) Then 'shame' in our cognitive agent architecture was given by the axiom:

$$\Delta_a \vdash \mathbf{ashamed}(a, b) \leftrightarrow \Delta_a \rightsquigarrow [\mathbf{eval}\ b\ a]+ \ \wedge\ \Delta_a \vdash \mathbf{eval}(b, a, -)$$

In other words, agent a has the goal that were b to evaluate a, it would do so positively, but the contradictory belief.

In the communication model (here represented using an axiom from the Event Calculus [10]), multiple infractions (each of which is sanctioned) in, for example, a file sharing protocol could lead to the revocation of trader status by an empowered agent (the trusted

third party, ttp):

$$
\begin{aligned}
&\mathit{revoke}(I,J)\ \text{initiates}\\
&\quad \mathit{role\_of}(J,\mathit{trader}) = \textbf{false}\ \text{at}\ T \leftarrow\\
&\quad \textbf{pow}(I,\mathit{revoke}(I,J)) = \mathit{true}\ \ \text{holdsat}\ \ T\\
&\textbf{pow}(I,\mathit{revoke}(I,J)) = \mathit{true}\ \ \text{holdsat}\ \ T \leftarrow\\
&\quad \mathit{sanctions}(J) = 2\ \ \text{holdsat}\ \ T\ \ \wedge\\
&\quad \mathit{role\_of}(I,\mathit{ttp}) = \mathit{true}\ \ \text{holdsat}\ \ T
\end{aligned}
$$

The loss of trader role, in conjunction with the 'axiom of shame' above, would lead to a frequently infringing agent experiencing 'shame', i.e. $\Delta_a \vdash \textbf{ashamed}(a,b)$.

One possible application of this work is in file sharing digital media. Rather than increasing emphasis on ultimately unreliable Digital Rights Management systems, we advocate an alternative approach, based on, for example, so-called disruptive technologies, like agents, and innovative business models, for example voluntary payments. So we can envisage a peer-to-peer file sharing network where the interaction between peers is 'agentified' in the manner described in this section, and there is a voluntary payment scheme for digital content. Then 'shame' is 'experienced' if the reciprocal payment is not made for downloaded content. This shame could also be visualised in the interface of the p2p client in some appropriate way.

5 AffectiveWare Platform

A potential enhancement to the use of emotions in computer-mediated human-human interaction and human-agent interaction is the possibility of directly detecting, and interpreting, physiological signals, in order to infer emotional state.

Affective computing [11] can be loosely interpreted as developing computer interfaces to sense and react to a user's emotions, to use emotions in automated reasoning and decision-making, and possibly even to display emotions. To this end, we have developed the AffectiveWare platform, based on the premises that, firstly, physiological signals are an indicator of psychological (emotive) state [4]; and secondly, that signal processing techniques can be used, in conjunction with a theory of emotion (e.g. [13]) to infer that emotional state from recorded signals, in soft real time (i.e. fast enough to usable in a human-computer interface).

The Imperial AffectiveWare platform consists of software and hardware components (a similar platform is described in [16]). In hardware, we have implemented sensors which measure physiological signals, such as the AffectiveRings and the AffectiveMouse (an ordinary mouse covered in conductive paint), which are used to measure galvanic skin response. The AffectiveRings are illustrated in Figure 4.

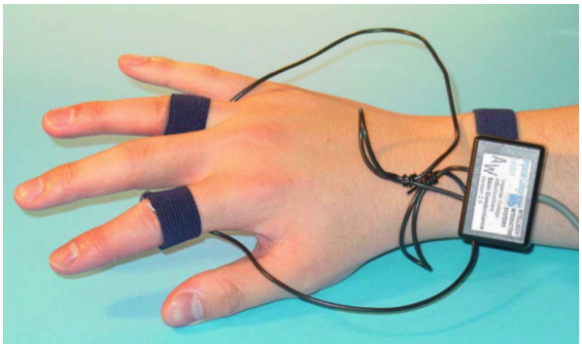

Figure 4: AffectiveRings

The crucial software component is the AffectiveAnalyser which performs the computation of emotional state, and is implemented using Fuzzy Logic. For the classification to work effectively, it requires reliable training data. The AffectiveWare platform enables an operator to classify data samples previously stored on an SQL database to a known emotion, and to set the window size of each emotional state.

After training, to perform classification, the Affective Analyser retrieves n number of samples for each emotion, where n is the window size of its corresponding emotion, recorded after the date and time set in the user interface. The mean, standard deviation, mean absolute first difference error and mean absolute second difference error parameters are then calculated for those n samples and fed into the Fuzzy Logic classifier that computes the degree of match those n samples data has to the

trained data sets for each emotion based on the four parameters. The output of this process is the emotion that has the highest level of confidence.

The AffectiveWare platform has been used in an experiment [6] to reveal and measure the emotive response of teenage users to global fashion and sport brands, given particular social, environmental, economic and technological information. One group of three female 17 to 19 year olds participated, each of whom had a white, middle class socio-economic background. The session was held at HP-Research Labs UK, and lasted 1.5 hours with a break in the middle.

The experiment was divided into two parts. In the first part, the users looked at and held each of four pieces of iconic branded fashion garments whilst reading the company PR from the official web site. The galvanic skin response was measured. In the second part, the users were shown the same pieces with a brief explanation of a sustainable issue in the public domain connected to each item relative to clothing supply chains consumption and disposal. The galvanic skin response was again measured.

This experiment, such as it is, has clear scientific limitations. The sample is, of course, far too small to derive any statistically significant conclusions. Furthermore, there is no concrete link between the physiological signals and the emotive descriptors of the Plutchik model [13]. However, the intention of this first experiment was only to confirm if there was a detectable reaction.

The findings confirmed this intuition and showed, firstly, that there was a measurable physiological response to iconic clothing items; and secondly, that the response is different according to reading company PR as opposed to factual information about sustainable issues. This suggests that the integration of affective computing, computer intelligence and digital communications can reinvigorate the debate on sustainable consumerism, both changing behaviour and empowering users within a product supply and disposal chain.

6 Summary and Conclusions

We define an open agent society as a distributed computer system or network where the relationships and dependencies between components is a microcosm of a human society, and therefore includes aspects of communication, conventional rules, memory, and emotions as mechanisms to deal with the inevitability of sub-ideal operation between unpredictable, heterogeneously designed components. The research presented here has indicated how the presentation and interpretation of emotions may be used to support autonomic functionality for ubiquitous computing applications where uncertainty concerning interactions needs to be managed.

We have based the idea of the open agent society on the premise that it is embedded or situated in human institutions or society. In particular, we place a strong emphasis on the psychological and physiological aspects of the relationship between humans and computer inhabitants of virtual worlds.

A potentially interesting avenue of investigation is then offered by considering emotions as a particular manifestation of *presence*, and we are increasingly concerned with making this bi-directional. Bi-directional emotional presence may be used to facilitate interactions between agents and humans, since we observe that hybrid networks of people/organizations and agents, who require different degrees and dimensions of robustness for autonomic communications to deliver an optimal outcome.

Then, we can imagine, firstly, the projection of a human personality (from the material world) into an electronic society of software agents (a digital world); and secondly, the projection of an electronic personality into a human society of ordinary people (i.e. digital world to material world projection). This is, to some extent, simply a logical extension of the fact that digital 'goods', e.g. from Second Life, can be sold for money (which can be used to purchase goods and/or services in the material world). However, it would also

represent an almost complete convergence of Web 2.0 with the real world.

Acknowledgements

Thanks to Petar Goulev, whose thesis research is reported in Sections 5; also thanks to the two anonymous reviewers for their very helpful comments. Dan Henrick created the avatar illustrations and animations of Figure 1.

References

[1] A. Artikis, J. Pitt and M. Sergot. Animated specifications of computational societies. In *Proceedings AAMAS 2002*: pp1053–1061, 2002.

[2] A. Bhusate, L. Kamara and J. Pitt. Enhancing the Quality of Experience in Cultural Heritage Settings. *Proceedings First European Workshop on Intelligent Technologies for Cultural Heritage Exploitation*, Italy, 2006.

[3] C. Castelfranchi. Affective appraisal versus cognitive evaluation. In A. Paiva (ed.): *Affective interactions: towards a new generation of computer interfaces*, LNCS1814, pp76–106, Springer, 2001.

[4] J. Cacioppo and L. Tassinary. Inferring psychological significance from physiological signals. *American Psychologist*, 1(45): 16-28, 1990.

[5] M. Dodge and R. Kitchin. *The Atlas of Cyberspace*. Addison Wesley, 2001.

[6] J. Farrer and P. Goulev. Conscience Clothing Research Project: The Emotional Episode. *Sustainable Innovation '06: Global Challenges, Issues and Solutions*, Chicago, 2006.

[7] Humaine Network of Excellence. http://emotion-research.net/

[8] Intelligent Information Interfaces. http://www.i3net.org/.

[9] D. Keltner and B. Buswell. Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin*, 122(3), 250-270, 1997.

[10] R. Kowalski and M. Sergot. A Logic-Based Calculus of Events. *New Generation Computing*, vol. 4 pp.67–95, 1966.

[11] R. Picard. *Affective Computing*. The MIT press, 1997.

[12] J. Pitt. Digital blush: towards shame and embarrassment in multi-agent information trading applications. *Cognition, Technology & Work*, 6(1), pp23–36, 2004.

[13] R. Plutchik. The emotions: Facts, theories, and a new model. New York: Random House, 1962.

[14] A. Rao and M. Georgeff. BDI-agents: from theory to practice. In V. Lesser (ed.): *Proceedings of ICMAS'95*, San Francisco, p312–319,1995.

[15] J. Searle. *Speech Acts*. Cambridge University Press, 1969.

[16] G. Trivino and A. van der Heide. *Linguistic summarization of the human activity using skin conductivity and accelerometors*. Proceedings IPMU'08 (this volume).

[17] A. Vasalou, A. Hopfensitz and J. Pitt. In praise of forgiveness: ways to repair trust breakdowns in one-off interactions. *International Journal of Human-Computer Studies*, (in press).

[18] A. Vasalou, A. Joinson and J. Pitt. Constructing my online self: avatars that increase self-focused attention. In *Proceedings of Conference on Human Factors in Computing Systems*, pp445-448; San Jose, USA, 2007.

[19] A. Vasalou and J. Pitt. Reinventing Forgiveness: A Formal Investigation of Moral Facilitation. In P. Herrmann, V. Issarny and S. Shiu (eds): *iTrust 2005*, LNCS3477, pp146–160, Springer, 2005.

[20] M. Weiser. Hot Topics: Ubiquitous Computing. *IEEE Computer*, October 1993.