'Social' Robots & 'Emotional' Software Agents: Gendering Processes and De-Gendering Strategies for 'Technologies in the Making'[8]

Jutta Weber, Corinna Bath

Introduction

Over the past years we can observe profound reconfigurations of the boundaries between human beings and machines in the field of artificial intelligence (AI) and computer science. Particularly software agents and robots attest to an ongoing paradigm shift from machine-oriented concepts, algorithms and automats towards 'interaction' (see Wegner 1997, Crutzen 2003). While early approaches sought to model rational-cognitive processes and to solve problems using formal structures, the emphasis is currently shifting to human-computer and human-robot interaction.

Recently artifacts such as software agents and robots are often conceptualized as friendly, understanding and believable partners which communicate 'naturally' with users and support them in everyday life. 'Sociable', humanoid robots are designed to take care of old or sick people. Software agents are expected to obtain information independently. In order to serve users and give them advice, they appear human-like on the screen.

In this paper we examine the recent trend in AI to build 'social' and 'emotional' artifacts from a feminist technology studies perspective. Starting from prominent visions of socio-emotional machines, some prototypes and commercial products, which currently came into use, we point out gendering aspects in their representation and the underlying concepts. Focusing on societal preconditions of socially intelligent machines, we will ask how traditional feminist critiques of technology might apply to these new artifacts. The arguments will lead us to point out some basic problems of developing de-gendering strategies

8 This chapter is a revised version of our paper published in: Archibald, Jacqueline; Emms, Judy; Grundy, Frances; Payne, Janet; Turner, Eva (eds.): The Gender Politics of ICT. Middlesex University Press 2005. It was presented at the Women Into Computing Conference, London 2005.

for 'technologies in the making'.[9] We finally propose dimensions and strategies for a contemporary feminist critique of technology.

1 The Vision of Sociable or Socially Intelligent Robots and Software Agents

Cynthia Breazeal from the Massachusetts Institute of Technology (MIT) is one of the leading researchers in the field of social robotics. Her vision of a sociable robot is a good example that clarifies the robot researchers' promises:

> "For me, a sociable robot is able to communicate and interact with us, understand and even relate to us, in a personal way. It should be able to understand us and itself in social terms. We, in turn, should be able to understand it in the same social terms - to be able to relate to it and to empathize with it. Such a robot must be able to adapt and learn throughout its lifetime, incorporating shared experiences with other individuals into its understanding of self, of others, and of the relationships they share. In short, a sociable robot is socially intelligent in a human-like way, and interacting with it is like interacting with another person. At the pinnacle of achievement, they could befriend us, as we could them." (Breazeal 2002: 1)

She stresses that social artifacts that become part of our daily life must be able to adapt to users in a natural and intuitive manner – not vice versa. Her 'masterpiece' – as she calls it – the robotic creature Kismet is designed to interact physically, affectively and socially with humans, in order to learn from them. The man-machine-relation (or should one say the woman-machine-relation?) is modeled according to that of a caregiver and a human infant. We found similar attempts in software agent research. Researchers stress that they aim at building *"emotional relationships by long-term interactions wherein the two parties pay attention to the emotional state of the other, communicate their feelings, share a trust, feel empathetic, and establish a connection, a bond."* (Stern 2002: 336). In some commercial computer games like 'Virtual Petz' or 'Virtual Babyz' the characters try to seek the users' attention in order to interact with them, to get 'care' and to get 'socialized' by the users.

9 The idea of 'technologies in the making' refers to contemporary approaches to the social studies of science and technology (e.g. Latour 1987). Instead of considering technologies as 'ready-made' or analyzing their context of use, we aim to track the development of technologies in a constantly shifting, multifaceted network of artifacts, disciplinary foundations, scientific methods, social preconditions and cultural meanings within a transdisciplinary and controversial universe of discourse.

To realize these envisioned social behaviors of machines researchers utilize models and theories from the fields of psychology, cognitive science, and ethnology, thereby aiming at the computation of social and emotional competencies.

2 Anthropomorphism and Gendering

If we take a look at the first prototypes and commercial products that were intended to be social we have to admit that they do not appear very innovative, at least with regard to the predominant gender concepts - if not to say stereotypes – used. Cyberella, a presentation agent created at the German Research Center for Artificial Intelligence (http://www.dfki.de/cyberella), and the robot "Valerie, a domestic android" (http://www.androidworld.com/prod19.htm) for instance, were given a kind of super feminine shape.

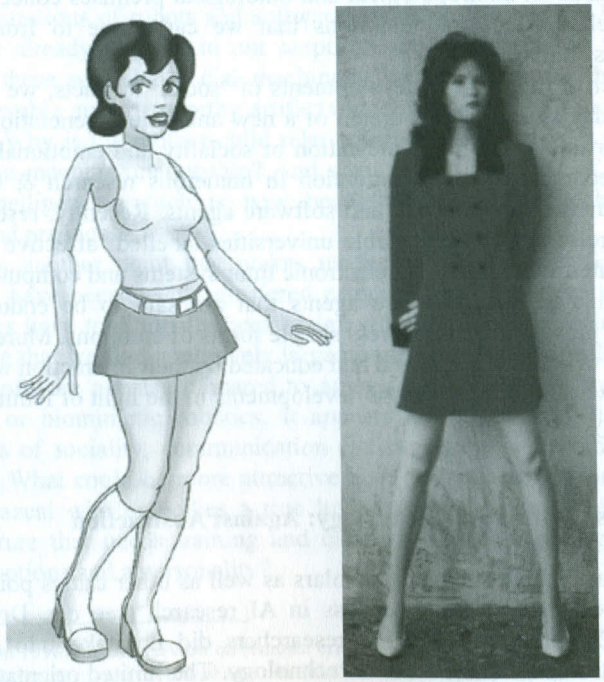

Figure 1: Cyberella and Valerie

Other software agents chat with users in a strongly gender stereotyped or even sexualized manner. Even though some of the new social artifacts appear more 'neutral' or less gender-stereotyped at a first glance they are nevertheless modeled on the ground of questionable ontological presumptions. The way how Cynthia Breazeal's interaction with Kismet is represented, for example, reshapes the wo/man-machine relationship into a gendered caregiver-infant relationship.

The visions and realizations of social artifacts in software agent research and robotics raise many questions:

What and whose understanding of sociality and emotionality is underlying these new artifacts? Is it desirable from a critical, feminist perspective to develop 'emotional' artifacts we are supposed to empathize with? Do artifacts modeled in terms of caregiver-infant-relationships represent a critical understanding of social behavior?

Or, more general: Is it desirable and promising to model human-machine-relationships according to those among humans? Are artifacts like Valerie and Cyberella based on anthropological and ontological premises concerning human behavior, relationships and emotions that we can agree to from a critical, feminist perspective?

In order to judge these developments of 'social' artifacts, we have to take seriously today's researchers' dream of a new and potent generation of socially intelligent artifacts. The implementation of sociality and emotionality into artifacts has become a center of attention in numerous research & development projects in the field of robotics and software agents. Recently, research laboratories emerge at several respectable universities labelled 'affective computing'. Some web sites, e-commerce or electronic tutor systems and computer games are already populated with software agents that are said to be endowed with a rudimentary personality and express simple forms of emotions. Moreover, robots are intended to become socialized and educated by their interaction with humans.

We now want to reflect these developments in the light of feminist critiques of technology.

3 Feminist Critique of Technology: Against Abstraction

During the last decades feminist scholars as well as other critics pointed out the lack of embodiment and situatedness in AI research (see e.g. Dreyfus 1963; Becker 1992). They stressed that researchers did not take into account the context and the social dimension of technology. The limited orientation towards rational-cognitive models and symbol processing was questioned. The critique often focused on the reductionist modeling of thought, on the simple understanding of human planning and acting as a merely rational-cognitive process and on approaches to problem solving constrained by the use of formal structures (see e.g. Suchman 1987).[10]

While much of the argumentation aimed against abstraction, disembodyment, decontextualization and the lack of the social dimension, it seems as if these critiques were now recognized by emergent technosciences. Meanwhile technology designers incorporate certain concepts of embodiment and situatedness into intelligent artifacts. While embodied robotics (see Pfeifer and Scheier 1999) and situated intelligent software agents are displacing symbol-oriented approaches in AI[11] we are now faced with the vision of ‚social machines'. Researchers aim at developing machines beyond the limits of rational-cognitive grounded intelligence. They discover social behavior as the basis of 'real' intelligence. How can the shift from rational to social behavior be perceived from a critical point of view? Will chances for feminist intervention arise along these changes or will the envisioned dissolution of the dichotomy of human beings and machines cement the existing gender order?

The illustrations of robots and software agents we gave in the introduction of this paper already referred to our sceptical stance towards the 'innovative' potential of these so-called social machines. We ask: Why are the promised helpful, believable and trustworthy artifacts modeled according to crude gender stereotypes? Why is the mother-child relationship assumed to be a good model for the human-machine-relationship? And what should we think of the concept of social intelligence, which is now becoming popular in technoscientific discourses and practices?[12]

There is another point that makes us feel uncomfortable: Sociality and emotionality have been deeply gendered categories in western thought. These characteristics have traditionally been assigned to the feminine realm. And it is not by chance that we find a relatively large number of women developing social robots and software agents compared to other areas of old-fashioned, symbol-oriented AI or biomimetic robotics. It appears as if their so-called ‚natural' competencies of sociality, communication etc. predetermined them to work in such a field. What could be more attractive than a nice-looking woman such as Cynthia Breazeal who embodies a true loving caregiver for a helpless infant robotic creature that needs training and caring to develop intelligence, social behavior, emotions and a personality?

10 See (Adam 1998) for an overview on feminist critique of AI.

11 For a discussion of philosophical and feminist influences on AI see for example (Sengers 1999).

12 The concept of social intelligence was developed at least partly because of the feminist critique of androcentric conceptions of evolution; see (Fausto-Sterling 1985; Hubbard 1979).

We approach the question of how to interpret recent developments in technology design by Bettina Heintz' critical approach to AI. In her paper *"Papiermaschinen. Die sozialen Voraussetzungen maschineller Intelligenz"*[13] the feminist science studies scholar and sociologist pointed towards the social and societal preconditions and implications of the mechanization of thought and everyday life with AI. On the one hand she claimed that first of all it were human beings who adapted themselves to the machine. Otherwise our unimaginative machines would not work at all. For example, secretaries are instructed to use a very simple language avoiding any ambiguities in order to enable software programs to translate their texts into another language. Thus the abilities we regard as genuine ones of computers are often the result of our own efficient work. And often we unconsciously compensate for the deficiencies of the machines, while at the same time our readiness to perceive machines as intelligent stems from our tendency to interpret our reality as loaded or even structured with meaning (see also Collins 1990). Following these arguments, the critique should not challenge primarily the claim that computers might become intelligent, but has to question the conditions that make us believe in the intelligence of machines.

On the other hand Bettina Heintz pointed out that a kind of mechanization of everyday life must have already taken place before the computer entered this process. The translation of problems into algorithms only becomes possible when humans already act in a rule-oriented way. A standardization of human behavior is necessary to model and develop software applications. What is the background in our society that elicits rule-oriented behavior that can be found so frequently?

Having feminist critiques in mind, we ask: what does it mean that techno-scientists anthropomorphize machines and discover sociality and emotionality as the cure for our still unimaginative, rational-cognitive grounded machines.[14] It seems that traditional strategies of wo/man-machine-communication are turned upside-down. While for a long time humans had to behave rationally and rule-oriented to make symbol-oriented machines successful, now machines are to become social in order to increase their usability and make them more helpful for human users. It is the machine now which is supposed to mimic or even learn those abilities and characteristics which were until recently regarded as purely and typically human and beyond the grasp of machines.

13 'Paper machines. The Social (Pre-)Conditions of Machine Intelligence'.

14 While roboticists and software agent researchers often point towards the tendency of anthropomorphization in the human-robot or human-computer interaction (Duffy 2003; Fong, Nourbakhsh, Dautenhahn 2003.; Gratch et. al 2002; Cassell et al. 2000), they rarely reflect on the additional work of humans to make sense of machinic behavior (see e.g. Suchman (in prep.) on the relationship between Breazeal and Kismet).

Rethinking Heintz' argument in the light of socially intelligent machines does not mean to ask whether and how machines can be or become social, but what makes us think of machines as social. What concepts of sociality and emotionality are predominant in today's AI and why? And what are the societal conditions under which machines are perceived as social and emotional?

Every socially intelligent machine we can dream of is still based on rule-oriented behavior, since this is the material ground and fundamental functionality of these machines. Therefore it is rule-oriented social behavior that is at the core of the theoretical approaches, concepts and practices of software agent researchers and roboticists. The kind of rules might differ in diverse strands of AI, but a standardization of human behavior is a precondition for every computer model and software application. Anthropomorphized machines are intended to operate by simulating social *norms*, supposed gender differences and other *stereotypes*. The starting point of these prototypes and implementations is rule-based social behavior that is said to be performed by humans. Researchers often use folk psychological and sociological approaches to sociality and emotionality in modeling human-machine-relations. Out of the wide range of psychology and sociology they particularly chose those theories for the computational modeling that assume that social behavior is operational.

And it is not by accident that software agent research and social robotics are working with sociological and socio-psychological approaches that explicitly use gender dichotomies and stereotypes. For example, we could detect a case which utilizes a feminist approach to improve human interaction with social machines: The computer scientist Daniel Moldt and the sociologist Christian von Scheve (2002) point out the value of roles, class and sex/gender differences in social interaction and their usefulness to minimize the contingency and to maximize the prediction of the behavior of the alter ego – of the human or machine partner in social interactive processes. Generally, in the realm of human-computer interaction emotions are considered to be useful to influence users, to convey intentionality and to smooth interactions. Referring to feminist sociologist Arlie Hochschild Moldt and von Scheve claim that emotions are based on a system of values and norms. Interrelations between emotions and social norms play a crucial role in matching the expectations of the alter ego. Moldt and von Scheve regard roles, class, gender and other differences as ideal categories in order to determine this relationship. Inspired by these ideas they strive for software agents that express emotions based on prevailing systems of values and norms. They assume that on this basis software agents appear intelligent, social and endowed with a personality (see Moldt and von Scheve 2002).

Not all of these new approaches, which aim to implement sociality into machines, exploit critical theories and feminist approaches in such a way. Never-

theless, this example shows that the paradigm shift from rational-cognitive to social machines does not lead to a de-gendering of technology design. The approach considered rather points to the fact that gender stereotypes are instrumentalized in order to build ‚better' machines that are perceived as socially intelligent.

Obviously, recent research in the field of software agents and social robotics is not primarily about making machines social as most researchers suggest. Rather it seems to be about training humans in rule-oriented social behavior. Only by relying on the latter the interaction with these machines can be made intelligible: As much as secretaries have to use an impoverished language to be able to use computer translation software, it will be necessary to use impoverished ways of interacting to respond to these social robots and artifacts. And while researchers use social norms and stereotypes to make their artifacts more consistent, convincing and believable, training humans in stereotypical behavior supports ways of acting which are predictable and therefore more exploitable in economic terms.

4 De-gendering Technologies: Dimensions and Strategies of Critique

Regarding recent developments in software agent research and social robotics it becomes obvious that we need a broader and highly differentiated feminist critique of artifacts and processes in AI and computer science. In the following we want to sharpen our analysis of de-gendering technologies with the insights given so far. Our intention is neither to develop a step-by-step recipe for necessary feminist interventions into technology design nor to give an overview of possible political practices, but to rethink strategies, tools and dimensions of feminist technoscience studies in the light of recent developments in the field of socially inspired AI and computer science.

4.1 Gender Representation

Rethinking sexist images or strongly gender stereotyped speech patterns used in social robotics and software agent research, obviously requires a critique of these stereotypes, patterns, norms and roles. This kind of critique of technology design targeting at *gendered representation* is well-established in feminist media studies. Often it is even shared by some (male) computer scientists. But what we wanted to point out here is that it is not sufficient only to revise the design of technology in the sense of wiping out its explicit or implicit gender stereotypes.

Nor would it be satisfying even to eliminate these and other social norms, roles and stereotypes like those of class, of age, of race, of sexuality etc. Gendered ontological and epistemological claims are also encoded in theoretical concepts that form the base for technological construction and software applications, such as the changing understanding of the social and the conceptualization of the human-machine relationship.

4.2 Social Theory

The relation between 'social machines' and the standardization of everyday life should be explored from a *social theory* perspective. It is the question whether we live in a society where social relations in general or at least in specific realms are already enacted in terms of rule-oriented behavior. Think for example of the standardization of health care for elderly people where every little service – like e.g. combing the hair, washing the back, etc. – is measured by standardized time schedules (minutes) and prices. In these realms the idea of social robots taking care of elderly people lies at hand. At the same time the standardization of social behavior through agents and robots might also lead to more rule-oriented behavior.

Another relevant aspect is linked to the question whether social machines are expected to fill in personal and relational vacancies that emerge with new social and work requirements in the age of globalisation. Will personal agents and robots that empathize with us and to whom we are befriended be a substitute for personal human relations in the age of mobility and change? Which deficiencies of our social life in the neo-liberal economy are supposed to be ‚repaired' by those artifacts?

4.3 Anthropological and Ontological Dimensions

From a critical perspective questions of anthropological and ontological assumptions arise on which technoscientific concepts in the fields of AI and software agent research are built. What is the underlying understanding of society, sociality and human interaction? How is the relation of human-machine conceptualized?

Concepts of sociality in human-machine interaction particularly draw on Anglo-American approaches to the social and behavioral sciences. In these approaches sociality is regarded as the outcome of the interaction of individuals, which is understood primarily as self-interested. Hence,

> "'social' refers to the exchange of costs and benefits in the pursuit of outcomes of purely personal value, and 'society' is the aggregate of individuals in pursuit of their respective self-interests." (Caporael 1995: 1)

These (reductionist) concepts are partly translated into the design of social robots and software agents. The models often become even more trivialized and simplified through software implementation processes. For example, in software agent research human behavior is commonly standardized by no more than five personality traits and six basic emotions.

Concepts of human-machine relationship, particularly in the new field of 'social' AI illustrate further ontological and anthropological assumptions. The relationships of owner-pet, parent-baby or caregiver-infant are sorts of pedagogical relationships that afford a lot of time, patience, engagement and work in order to function properly. Are these the kind of relationships desirable for human-machine interaction? Do we really want to educate our machines?

To summarize and to return to our starting question about the strategies, which are necessary to design 'de-gendered technologies', we argue that a deconstruction of gender representation as well as a critique of fundamental epistemological and ontological assumptions are essential. The criterion for promising research in this field cannot be the question whether guidelines for an alternative technology design are provided. Instead such an approach aims at a fundamental revision of societal structures, politics of representation *and* technoscientific discourses and practices.

References

Adam, Alison (1998): Artificial Knowing, Gender and the Thinking Machine. London/New York: Routledge.

Becker, Barbara (1992): Künstliche Intelligenz. Konzepte, Systeme, Verheißungen. Frankfurt am Main: Campus.

Breazeal, Cynthia (2002): Designing Sociable Robots. Cambridge, Mass.: MIT Press.

Cassell, Justine; Sullivan, Joseph; Prevost, Scott; Churchill, Elizabeth (2000): Embodied Conversational Agents. Cambridge, Mass.: MIT Press.

Caporael, Linnda R. (1995): Sociality: Coordinating Bodies, Minds and Groups. Psycoloquy 6 (01), Groupselection 1. Available: http://psycprints.ecs.soton.ac.uk/archive/00000448. Accessed January 18, 2006.

Crutzen, Cecile (2003): ICT-Representations as transformative critical rooms. In: Kreutzner, Gabriele; Schelhowe, Heidi (eds): Agents of Change. Virtuality, Gender and the Challenge to the Traditional University. Opladen: Leske + Budrich: 87-106.

Collins, Harry (1990): Artificial Experts. Social Knowledge and Intelligent Machines. Cambridge, Mass.: MIT Press.

Dreyfus, Hubert (1963): What Computers Can't Do. A Critique of Artificial Reason. New York: Harper & Row.

Duffy, Brian (2003): Anthropomorphism and the Social Robot. In: Robotics and Autonomous Systems 42 (3-4): 177-190. (See also online. Available: http://www.cs.ucd.ie/csprism/publications/desire/IROS-AnthroSocialRobot.pdf. Accessed January 18, 2006).

Fausto-Sterling, Anne (1985): Myths of Gender, Biological Theories about Women and Men. New York: Basic Books.

Fong, Terrence; Nourbakhsh, Illah; Dautenhahn, Kerstin (2003): A Survey of Socially Interactive Robots. In: Robotics and Autonomous Systems 42 (3-4): 235-243.

Gratch, Jonathan; Rickel, Jeff; André, Elisabeth; Cassell, Justine; Petajan, Eric; Badler, Norman I. (2002): Creating Interactive Virtual Humans: Some Assembly Required. In: IEEE Intelligent Systems 17 (4): 54-63.

Heintz, Bettina (1995): Papiermaschinen. Die sozialen Voraussetzungen maschineller Intelligenz. In: Rammert, Werner (ed.): Soziologie und künstliche Intelligenz. Produkte und Probleme einer Hochtechnologie. Frankfurt am Main: Campus: 37–64.

Hubbard, Ruth (1979): Have Only Men Evolved? In: Hubbard, Ruth; Henifin, Mary Sue; Fried, Barbara (eds): Women Look at Biology Looking At Women: A Collection of Feminist Critiques. Cambridge: Schenkman Publ. Comp. 7-35.

Latour, Bruno (1987): Science in Action. How to Follow Scientists and Engineers through Society. Cambridge: Havard University Press.

Moldt, Daniel; von Scheve, Christian (2002): Attribution and Adaption: the Case of Social Norms and Emotions in Human-Agent Interaction. In: Marsh, Stephen; Nowell, Lucy; Meech, John F.; Dautenhahn, Kerstin (eds): Proceedings of The Philosophy and Design of Socially Adept Technologies, Workshop held in conjunction with CHI'02, 20.4.2002, Minneapolis/Minnesota, USA: 39-41.

Pfeifer, Rolf; Scheier, Christian (1999): Understanding Intelligence. Cambridge, Mass.: MIT Press.

Sengers, Phoebe (1999): Practices for Machine Culture. A Case Study in Integrating Cultural Studies and Artificial Intelligence. In: Surfaces, Volume VIII.

Suchman, Lucy (1987): Plans and Situated Action. The Problem of Human-Machine Communication. Cambridge: Cambridge University Press.

Suchman, Lucy (in prep.): Replicants and Irreductions: Affective Encounters at the Interface. In: Suchman, Lucy: Plans and Situated Actions, 2nd Ed. Cambridge: Cambridge University Press.

Stern, Andrew (2002): Creating Emotional Relationships with Virtual Characters. In: Trappl, Robert; Petta, Paolo; Payr, Sabine (eds): Emotions in Humans and Artifacts, MIT Press, Cambridge, Mass.: 333-362.

Wegner, Peter (1997): Why Interaction Is More Powerful than Algorithms, Communications of the ACM 40 (5): 80-91.