

26-28/6/09

Corfu Greece

International Conference of Computer Ethics **Philosophical Enquiry**







Department of Archives and Library Science, Ionian University Department of Informatics, Ionian University International Society for Ethics and Information Technology (INSEIT)







PAPERS

8th International Computer Ethics: Philosophical Enquiry Conference

Ionian University: Department of Archive and Library Sciences-Department of Informatics with INSEIT

Ionian Academy
Ionian University, Corfu, Greece June 26-28, 2009

KEYNOTE ADDRESSES

Pr. Terrel Bynum: Weizenbaum Keynote Address: Philosophy and the Information Revolution

Pr. Urs Gasser: Past, Present, and Future of Copyright Law in the Digital Age

Pr. George Bokos: The Need For a New Mentality, Tools and Practices in the Modern Information Environment

PANELS

1. Aycock J., Ledin G., Marurushat A., Sullins J., Malware Ethics: the Beneficial Use of Forbidden Knowledge

2. Brey P., Briggle A., Bynum T., Johnson J., & Spence E., Information

Technology and the Good Life

3. A birthday panel: Capurro R., Britz J. & Nakada M., Changes of the Internet, Changing Information Ethics

4. Ferguson A., Thornley C., Weckert J. & Gibb F., The Ethics of the

Application of RFID in Libraries

5. Karachalios K., Garcia de los Santos L. & Partenheimer J., Paradoxes of contemporary knowledge: between invention, creation, information and

6. Taddeo M., Turilli M., Miller K., Elgesen D., E-trust: for and against

TABLE OF CONTENTS

1. An abstract of Terrel Bynum's Weizenbaum Address: Philosophy and the Information Revolution p. 5

2. Text from a Panel: Karachalios K., Garcia de los Santos L. & Partenheimer J., Paradoxes of contemporary knowledge: between invention, creation, information and control p. 5

3. Asai R., Classic yet Contemporary Gender Norm: is ICT an amplifier of

gender bias? p. 18

4. Alexiou A. & Vlamos P., Ethics at the Crossroads of Bioinformatics and Nanotechnology p. 26

5. Al-Fedaghi S., Modeling Aspects of Action Theory p. 40

6. Basart J. & Serra M., A Frame for Ethical Evaluation of (Information) Technologies p. 56

7. Briscolas G. & Roussos P.L., The Greek scale of attitudes towards unethical behaviors on the Internet p. 64

8. Briggle A., Computer-mediated Friendship: Illustrating Three Tasks for a Computer Ethics of the Good p. 76

9. Buchanan E. & Ess C., Internet Research Ethics: Report from US-Based Institutional Review Boards p. 90

10. Chopra S. & Dexter S., Free Software, Economic 'Realities,' and Information

11. Coeckelbergh M., Living in the Eye of the Artificial Other. Artificial Intelligence, Moral Imagination, and Disciplining p. 105

12. Collste G., The Ethics of Ambient Computing for Personal Health Monitoring

13. DeBeer F., Responsibility in the Age of the New Media: Are Cyborgs Responsible Beings? after p. 731, 1-10

14. De Laat P., Trusting Invisible Strangers in Open Source Communities: About the Assumption, Inference and Substitution of Trust p. 138

15. Foteinou G. & Pavlides G., Ethical aspects of e-Government: Social Actors, Politics and the Digital Divide p.181

16. Ganascia J.G., The Great Catopticon p. 199

17. Gotterbarn D. & Moor J., Virtual Decisions: Just Consequentialism, Video game ethics, and Ethics on the fly p.226

18. Grodzinsky F., Miller K. & Wolf M., Developing Artificial Agents Worthy of trust: "Would You Buy a Used Car from this Artificial Agent?" p. 239

19. Grodzinsky F. & Tavani H., Can the "Contextual Integrity" Model of Privacy Be Applied to Personal Blogs in the Blogosphere? p. 254

20. Healy M. & Fairweather B., ICT and financial services: learning lessons from the environmentalists p. 262

21. Horner D.S., Moral Luck and Computer Ethics: Gaugin at the keyboard p.

22. Kallinikou D., Papadopoulos M., Kaponi A. & Strakantouna V., Intellectual Property Issues for Digital Libraries in the Internet Networked Public Sphere p. 286

23. Konsta R. & Gerolimos M., A Brave New World for Libraries: the Sui Generis Right p.

24. Knox H., O'Doherty D., Vurdubakis T. & Westrup C., The Mirror and the Spell: Order and Organization in the Age of the Digital Reserve p. 312

25. Laaksoharju M. & Kavathatzopoulos, Computerized support for ethical analysis p.

26. Loukis E. & Papadakis J., Is the spammer evil? p. 334

27. Marturano A. Bellucci S.,, A debordian analysis of facebook p. 346

28. Mitrou L., The Commodification of the Individual in the Internet Era: Informational Self-determination or "Self-alienation"? p. 359

29. Molony J., The Joy of Excellence: Intellectual Property and Hackers' Virtue

30. Muller V., From Public Data to Private Information: the Case of the Super-Market p. 388

31. Makada N., Blogs and privacy in Seken as a Japanese Life-world Including Indigenous Moral Norms p. 403

32. Murata K. Information Ethics in Business Organisations p. 393

33. Orito Y., The Counter-Control Revolution: Silent Control over Individuals with Dataveillance Systems p.429

34. Reynolds C. J. & Casinelli A., Machine Self-Sacrifice p. 507

- 35. Pateli A., Ethical Issues in the Design and Use of Online Career Development Counseling Services p. 497
- 36. Papadopoulou M.D., Being a 'Step-Parent' to an 'Orphan' Work: Problemsa and Solutions Regarding the 'Adoption' of 'Orphans' p. 438

37. Patridge S., An Ethics of Video Game Play p. 465

38. Serbedzija N., Mixed Feelings Computing p. 517

- 39. Siemoneit O., Ethical Issues of Pervasive Computing in the Insurance Industry
- 40. Simon J. MyChoice & Traffic Lights of Trustworthiness: Where Epistemology Meets Ethics in Developing Tools for Empowerment and Reflexivity p. 563
- 41. Sinodinou T,. The Rights and Duties of Lawful User in Copyright Law p. 548
- 42. Spence E., Information, Knowledge and Wisdom: Groundwork for the Evaluation of Digital Information and Its Relation to the Good Life p. 563
- 43. Stahl B.C. & Rogerson S., Landscapes of Ethical Issues of Emerging ICT Applications in Europe p. 601
- 44. Stamatellos G., Security, Privacy and Technophobia in the School of the Future p. 619
- 45. Stamatoudi I., The Role of Internet Service Providers. Ethics, Reality and the Law: the Example of Promusicae v. Telefonica p. 628
- 46. Stiakakis E. & Alexandridou-Egytiadou E., The Digital Divide among Under-Age Individuals: An Economic and Legal Approach p. 661
- 47. Soraker J. H., Virtually Good What Can We Learn from the Argument from False Pleasures? p. 644
- 48. Spinello R., Remixing and Recoding: Revisiting the Copyright Wars p.651
- 49. Theoxaridou M. & Gritzalis D., Situational Crime Prevention and Insider Threat: Countermeasures and Ethical Considerations p. 667
- 50. Tuovinen L. & Roning J., Everybody Wins: Challenges and Promises of Knowledge Discovery through Volunteer Computing p.678
- 51. Yadin A., Information Systems and Future Ethics: a Reflection of Students' Current Behaviour p. 690
- 52. Yannopoulos G., CURIA NOVIT GOOGLE: The Quest for Law in the Era of Internet Search Engines = 731b-
- 53. Weber K., Constraints to the Application of ICT Implants: The Concept of Self-Ownership p. 699
- 54. Wong P.H., What Should We Share? Understanding the Aim of Intercultural Information Ethics
- 55. Zimmer M., "But the Data is Already Public": On the Ethics of Research in Facebook p. 718
- 56. CS deBeer, Responsibility in the Age of Cyborgs in the Age if New Media: Are Cyborgs Responsible Beings? p. 731a-
- 57. Konsta R. & Gerolimos M, A brave new world for libraries: the sui generis right p. 731c-

Living in the Eye of the Artificial Other Artificial Intelligence, Moral Imagination, and Disciplining

Mark Coeckelbergh1

¹Department of Philosophy, University of Twente

Abstract: Developments in robotics and artificial intelligence suggest a near-future scenario in which technology explicitly assumes a social function. In this paper, I develop a conceptual perspective for understanding living together with artificial others and draw some implications for ethics. In particular, I argue that we can understand some robots and some AI environments as constituting quasi-others, which, to the extent that they are social others, may become 'generalised others' or 'impartial observers' and therefore subject us to socialisation and moral disciplining. I conclude that ethical issues raised by such intelligent systems are not be limited to the protection of privacy and are better reconceptualized as problems we have as social beings trying to live good lives.

.....

Keywords: robots, artificial intelligence, ethics, quasi-other, moral imagination, disciplining, privacy.

1. Introduction

Developments in robotics and artificial intelligence (AI) suggest a near-future scenario in which technology explicitly assumes a social function. Robot designers expect that so-called personal robots or social robots will increasingly take part in our daily, personal lives. They put forward a vision of robots leaving the factories of industrial production and entering our households, our care institutions, and perhaps our personal relationships. Such a project raises ethical issues, such as 'Is it right to have sex with robots?' (Levy 2007) and 'Can robots contribute to the good life?' We also may want to invoke Orwell's image of 'Big Brother' in order to express our ethical concerns with AI systems that monitor us at home and respond to our (perceived) needs.

Usually the latter concern is phrased in terms of the 'privacy' principle: AI threatens our privacy. Privacy is a long-standing issue in ethics of information technology (see for example Van

den Hoven 2008; Tavani 1999; Nissenbaum 1998). The application of AI technology in the personal sphere seems particularly threatening with regard to privacy. Do we want to live in intelligent electronic environments full of AI systems and AI entities that watch us day and night?

In this paper, I take one step back and ask about the nature of the 'social' role of intelligent technology. What would it mean to understand robots and AI systems as social entities that live with us? What is there precise social role? And what are the implications for ethics? I develop a conceptual perspective for understanding living together with artificial others and draw some implications for ethics.

Engaging with various perspectives from philosophy of technology, I first distinguish three ways in which technology can be understood as social. Using Smith, and Mead I then argue that AI technologies cannot only become quasi-others but also social others that launch our moral imagination as social beings and discipline us. Finally, I discuss the implications for ethics and show how this approach re-conceptualizes the problem.

2. Technology and the social

There are at least three ways in which technology can be understood as 'social':

- 1. Some artefacts are designed to have a social function. For instance, a telephone is meant to connect people. A robot could be designed to be a companion or to play the role of a parent, teacher, or child. On this instrumental view of technology, a robot can be social if it is meant to be social, that is, if the designer had this aim in mind, if it is intended by the designer. For ethics, then, the task is to make sure that designers have goals that are morally good. For instance, engineers can design robots that assist education, but they can also design military robots that collaborate with human soldiers on the battlefield. Both kinds of robots are 'social', in the sense that they interact and communicate with humans, but the latter application is usually seen as morally problematic. Either way, the ethical analysis concerns the goal(s) of the technology, not the means (the technology itself). For example, if an AI technology is meant to assist elderly people, make their life easier, and allow them to stay longer in their own homes, then it can be argued that this is good since the aim is good.
- 2. In a broad (and perhaps more radical) sense technology has

always had a social function: contemporary philosophy of technology teaches us that artefacts mediate between people and do things (Verbeek 2005) Technology is not a 'mere instrument' that serves our pre-set aims and purposes, but does something more fundamental to the way we live together and changes our goals. For instance, a mobile phone changes how we organize our social lives – it changes that life itself. And from this perspective robots that are used in health care do not only serve the aims set by designers and health care professionals, but are also likely to change health care practices themselves. With the technology, the meaning of the practice changes. Computers became more than calculators and typewriters; they changed the way we work, live, and think.

My own version of this non-instrumental view of technology is inspired by Arendt's claim in *The Human Condition* that things are 'in-betweens'. The social life depends on a material world. By constituting a world, things relate people to one another. The table between us gather us, renders possible the social. Our social world is a world of people and things (Arendt 1958).

To live together in the world means essentially that a world of things is between those who have it in common, as a table is located between those who sit around it; the world, like every in-between, relates and separates men at the same time. (Arendt 1958, p. 52)

For an ethics based on this view, the task is to evaluate the social consequences of artefacts: what is their contribution to the social world? Do they relate people and how? For example, a baby robot in elderly care environments can relate people by making people collaborate to care for it, by making them talk about the robot, etc. (Kidd et al. 2006). A social world develops around the robot.

More generally, technology is not mere 'decoration', the background, or the stage on which then the social life develops. While it also provides part of the material conditions, it does more as well: it shapes the social life itself. It changes the script, the words, the scenes, and the actors. It does not play a marginal role; it co-constitutes the social by contributing to its conditions, its development, and its flourishing.

3. Viewed from these two perspectives, however, we cannot make sense of what makes the social role of artificially intelligent systems different from that of a mobile phone or a table. What is specific about the social dimension of this technology that warrants extra philosophical attention? In order to bring this to the foreground, let me explore an alternative conceptual route.

Apart from being designed for a particular purpose (social or other) and in addition to playing a more basic role as a condition and constituent of the social world, some AI systems tend to become what Ihde calls a 'quasi-other' (Ihde 1990). Ihde uses the term in *Technology and the Lifeworld* when discussing what he calls 'alterity relations': the technology is neither in the background nor embodied (part of me), but becomes an other to which I relate (Ihde 1990, p. 97-108).

Let me explain this. With many AI systems we may have a background relation, for instance an intelligent climate control system. Some might be embodied, for example glasses that highlight information about our environment. But when we have a conversation with a personal, artificially intelligent household robot, we are likely to engage in an alterity relation with that robot: it appears an other. Perhaps we give a name to the robot, for example. This is very plausible. We already treat some pet animals as if they were others. We neither experience them as part of us, nor are they merely 'part of the furniture'. They are quasi-others.

In order to further develop this perspective, I propose to understand these alterity relations as social relations. Consider how we interact with robots that resemble us. Although we are usually perfectly aware that a particular robot is not human, our social response to that robot is likely to resemble our social response to human others. For instance, Hiroshi Ishiguro has used humanoid robots to study (human) interaction (Ishiguro 2006; MacDorman and Ishiguro 2006). This is only possible since we tend to treat humanoid robots as social others. We do not treat such robots as 'mere things' and we are usually unaware of the social nature of all things in an Arendtian sense (things that gather us); instead, we tend to include them in our social world as we would do with fellow humans (and indeed some animals).

Note that to have a social or quasi-social relation, AI robots or other AI system need not be humanoid. Presently, we often treat animals and computers in a 'social' way. For example, Reeves and Nass have shown that interactions with computers are similar to social relationships with humans (Reeves and Nass 1996).

Note also that from this perspective, it makes no difference whether or not an AI system has consciousness or is sentient. Our treatment of (some) robots as social others is not based on an ontology of mental properties, but on social function. What counts is that the entity in question appears to us (humans) as a quasi-other, regardless of the properties the entity actually has.

What does this perspective imply for ethics? As I said in my introduction, one worry we might have is that the technology becomes 'Big Brother', that is, a quasi-other that watches us and perhaps restrains our behaviour. But does this happen, and if so, to what extent is that a problem? To better understand the social dimension of AI technologies as quasi-others, we must turn from philosophy of technology towards moral and social philosophy tailored to humans. The conceptual resources available to us in this area are vast; let me select and employ one perspective: a social ethics that recognizes the role of moral imagination.

.....

3. The individual, the social, and morality: Moral imagination and disciplining

There are at least two different and opposing views on the relation between the individual and the social (and the moral). According to the first view, there are 'first' individual, self-conscious individuals with a self, which raises the question how the social and the moral is possible. Hence we have theories such contractarianism that try to understand and justify the existence of society and political social institutions. According to the second view, it is the social that is more 'original'; self-consciousness, individuality, self, and morality emerge from the social rather than the other way around. It is the individual, not the social that stands in need of explanation. Here I explore a route that has more affinity with the second view and try to apply it to the 'Big Brother' issue with robots and AI systems.

I already argued that some robots and AI systems can be understood as quasi-others. Let me now say more on how they can be interpreted as *social* others and what this implies for morality.

Some traditions in moral philosophy highlight that we always live in the eye of the other. We are moral beings as social beings. Let me explain this by drawing on moral sentiment theory and pragmatism. In *The Theory of Moral Sentiments* (1759) Adam Smith shows that when judging our own conduct our moral

imagination is always directed to others. This takes on two dimensions. First, we judge our own conduct by placing ourselves in the situation of a particular other. We ask ourselves what that particular other would say of our conduct. Second, we also generalize the particular other. We ask what 'one' would say of our conduct. Smith uses the concept of 'the impartial spectator' to explain this:

We endeavour to examine our own conduct as we imagine any other fair and impartial spectator would examine it. If, upon placing ourselves in his situation, we thoroughly enter into all the passions and motives which influenced it, we approve of it, by sympathy with the approbation of this supposed equitable judge. If otherwise, we enter into his disapprobation and condemn it. (Smith 1759, pp. 109-110)

In a similar vein, Mead has argued in *Mind*, *Self*, and *Society* that humans are mainly social beings who develop themselves (morally) by taking the attitudes of our community as their values (Mead 1934). Mead explains this process by putting forward the concept of the 'generalized other':

He is putting himself in the place of the generalised other, which represents the organised responses of all the members of the group. It is that which guides conduct controlled by principles (Mead 1934, p. 162).

Thus, for Mead moral development is not so much a matter of gaining knowledge of abstract principles (and applying them) but of using one's moral imagination understood as social imagination. Principles embody the generalised other; they are not something that stands apart from the social. Both Smith and Mead share the view that morality is deeply social and imaginative: when we evaluate our conduct, we imagine how the other and 'one' would judge it.

If others are that important for our moral self-evaluations, this has implications for quasi-others as well. To the extent that we regard and treat AI robots and other AI systems as quasi-others, they can fulfil the following double role. First, they can act as particular social others. Humans are likely to evaluate their own conduct by asking questions such as 'What would the system say if I did that?' or 'How would the robot feel about that?' – even if the humans know that the robot is not a 'real' other. Second, when artefacts appear as social quasi-others they can come to be

imagined as embodying the values of the community. They can be seen as representing the impartial spectator, the generalized other. In this role, the technology confronts the humans involved with what they imagine as the evaluation of their conduct by the community.

In both roles, the quasi-others will have a disciplining effect. In Discipline and Punish (1975) and other works Foucault has pointed to the many subtle ways in which power is exercised over individuals by institutions. This also depends on artefacts. Many technologies already have disciplining effects in virtue of what they do to us: they facilitate some behaviour while discouraging or preventing other behaviour. Consider a speed bump: it constrains our conduct. This was one of the ways to understand the 'social' role of technology I mentioned above. However, the disciplining by the particular or generalised other (in the sense meant by Smith and Mead) is more subtle and follows a less material route: individuals will (not) choose a particular course of action because they imagine that the robot would (not) approve of it or because they will experience the 'eye' of the technology as the eye of the community that would (dis)approve of their action. They adapt their behaviour to these expectations and imagined judgments.

For ethics, then, the question is not if imagining the view of the generalised other and if disciplining as such is right or wrong. As social beings, we do so anyway. We imagine what others would say of our conduct and in this way we discipline ourselves. With respect to the social-imaginative process described above, the ethical question must concern (1) the rightness and quality of the imagined moral judgment (whatever the real robot or real humans would say) and (2) the rightness and value of the behaviour and life that result from the disciplining. Thus, with regard to AI technologies as quasi-others, the main evaluative question is not: 'Does Big Brother watch me and is that right?' but 'Is his judgment - as I imagine it - right and does he make me do things that are right and good? Does this technology - in its role as quasi-other that makes me imagine how others would evaluate my behaviour and that therefore disciplines me - facilitate a good life?' Instead of limiting ethical issues to what harm can be done to the individual by others or by society, this approach first accepts that we are social beings and asks from that starting point how individual conduct and individual lives are to be shaped. Instead of imagining the individual mainly as a fortress that needs to be defended against invasion, humans are understood as heavily dependent on others and on their community, as deeply connected to those others and that community both in their interactions and in their moral imagination. Others are not 'outside'; they are already in our minds. To the extent that AI technologies can appear as an other, they become part of that social-moral constellation and have to be studied and evaluated in those terms.

.....

.....

4. Conclusions

I have argued that we can understand some robots and some AI environments as social quasi-others. This implies that as particular others and as launching sites for imagining the 'general other' or the 'impartial observer' they subject us to moral disciplining. As such, they can have a two-fold moral-social significance similar to human others: they too can be considered as particular quasi-others we have to live with and as gateways to imagining the 'generalized other'.

Whether or not this moral disciplining is morally acceptable is not the right question since it is unavoidable as part of how we think as social beings. In general, moral disciplining is not necessarily bad in itself. It plays an important role in education and moral development, for example. Social institutions cannot do without it. Consider what car traffic would look like without disciplining. And perhaps good care or good education always violates the privacy of individuals. This does not mean that all AI technologies are always good. Particular technologies will certainly be morally problematic or even unacceptable because they contribute to wrong actions or low-quality lives. But I have argued that, with respect to technologies as social quasi-others, the ethical question concerns not so much the technology itself but the quality of our self-judgments and the behaviour and lives that result from these self-judgments. The importance of moral-social imagination does not imply that we are bound to follow social conventions or have to accept any technology. The technology as quasi-other launches our social imagination and renders possible self-judgment, but what matters morally speaking is the quality of that judgment and of our lives with that technology.

Note, finally, that both the social imagination and the disciplining effects it has may differ between cultures and societies. In some cultures, the social dimension of life is more important than in others. This has to be taken into account when evaluating robots and other AI systems as quasi-others.

I conclude that my re-conceptualization of the 'Big Brother' issue as a problem of quasi-others and social disciplining, changes the problem formulation. Usually the issue is formulated as a 'privacy' problem or in terms of other principles that have to do with protecting individuals from the state or from others. But rather than a problem of individuals threatened by technology, it is here viewed as a problem we have as social beings that use our imagination to judge ourselves. The ethical question is then about evaluating that judgment and the role of particular technologies that contribute to that judgment as quasi-others: what do they do to our moral-social imagination, our conduct, and our lives? This, I conclude, is my tentative answer to the main question I asked in the beginning of this paper: what it would mean to understand certain robots and AI systems as social entities assuming social functions, and what this implies for ethics.

.....

References

Arendt, H., (1958). *The Human Condition*. University of Chicago Press, Chicago.

Foucault, M., (1975). Discipline and Punish. (trans. A. Sheridan) Random House, New York, 1979.

Ihde, D., (1990). Technology and the Lifeworld. Indiana University Press, Bloomington/Minneapolis.

Ishiguro, H., (2006). Android science: conscious and subconscious recognition. *Connection Science*, 18 (4), December 2006: 319-332.

Kidd, C., Taggart, W., and Turkle, S. (2006). A Sociable Robot to Encourage Social Interaction among the Elderly. IEEE International Conference on Robotics and Automation (ICRA), Orlando, Florida, U.S.A., May 2006.

Levy, D., (2007). Love and sex with robots. Harper Collins, New York.

MacDorman, K.F. and Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3): 297-337.

Mead, G.H., (1934). Mind, Self, and Society. Ed. C.W. Morris. University of Chicago Press, Chicago.

Reeves, B., and C. I. Nass, (1996). The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places. Stanford: CSLI Publications; New York: Cambridge University Press.

Nissenbaum, H., (1998a). Protecting privacy in an information age: The problem of privacy in public. Law and Philosophy, 17: 559-596. Smith, A., (1759). The Theory of Moral Sentiments. (ed. D.D. Raphael and A.L. Macfie) Clarendon Press, Oxford, 1976. Tavani, H., (1999). Privacy On-Line. Computers and Society, 29(4): 11-19. van den Hoven, J., (2008) Information technology, privacy, and the protection of personal data. In J. van den Hoven and J. Weckert (eds.). Information Technology and Moral Philosophy. Cambridge: Cambridge University Press, 301-321. Verbeek, P.-P., (2005). What Things Do. Penn State University Press, Penn.