REVEALING THE 'FACE' OF THE ROBOT INTRODUCING THE ETHICS OF LEVINAS TO THE FIELD OF ROBO-ETHICS*

BENJAMIN S. WOHL
HighWire DTC, Lancaster University,
Lancaster, United Kingdom

This paper explores the possibility of a new philosophical turn in robot-ethics, considering whether the concepts of Emanuel Levinas particularly his conception of the 'face of the other' can be used to understand how non-expert users interact with robots. The term 'Robot' comes from fiction and for non-experts and experts alike interaction with robots may be coloured by this history. This paper explores the ethics of robots (and the use of the term robot) that is based on the user seeing the robot as infinitely complex.

1. Introduction

The 2003 remake of Battlestar Galactica begins with a robot designed to look like a young beautiful woman approaching an ageing general and saying "Are you alive? Prove it!"[6]. This question (if not the direction of it) is a reoccurring one to the fact and fiction of robotics. The relationship between SciFi and fact is far from straight forward, in some cases there is a direct relationship, far more than directly influencing innovation SciFi shapes attitudes, visions and the reception of innovations [5].

In the field of robotics and particularity robo-ethics it is very difficult to separate the myth and the reality. In fact, there is a seeming paradox hidden in the field that those who write about the reception of robots by society have a particular interest in robots, and fail to appropriate the reception of robots by those without this interest. As Mudry and Degallier succinctly put it "Indeed, there is a gap between what robots effectively are and can do and what people expect from them, based on their cultural representation on the robot rather than on existing robots"[30].

This paper aims to explore some of the ethical issues in the robotics and moves towards a position which begins to consider what makes a robot a robot in the eyes of a non-expert. The term 'robot' emerged early in the 20th century with the Czech play R.U.R. By Karl Capek [11]. In this play the robots are not

* This research was supported by HighWire, a post-disciplinary programme at Lancaster University funded by the UK Digital Economy Programme.

computerized or mechanical but organically manufactured worker/slaves that revolt and over throw their human masters only to find they have also lost the secret of their own creation. In the many stories within a story of Mary Shelley's "Frankenstein or the Modern Prometheus", the question is directly raised how do one as a scientist (researcher or roboticist) react when one's creation finally comes to life [36]. Both Capek and Shelly are directly addressing the Golem trope which is a question of soul or the missing piece, in both cases the creations are made of the same material as humans a key aspect of the golem myth [31]. Until Capek's word 'Robot' (from the Czech word for forced labour) was taken up by SciFi writers such as Issac Asimov automata were not seen as potential creation able to have a soul. Contemporary writers take the idea a step further in for example in the short story "I, rowboat" [15] which presents the question can a robot be able to worry about saving its own soul.

In his paper "Unhomely at home: Dwelling with Domestic Robots" Canadian academic Nicholas Anderson forces an imagining of a home filled with automated activity, where the line between organic life and mechanical life loose importance [3]. This paper will approach the established field of robo-ethics form the point of view of the philosopher Levinas. I will be asking, "When do we look into the face of the robot and see the robot looking back at us"?

Within this paper I will review three areas of research, which are often drawn upon when considering the ethical dimensions of robotics. When considering the response of users to robots in Human Robot Interaction (HRI) the theoretical frame work of Mori's Uncanny valley and the concept of Theory of Mind are often used to understand how a user will interact with a robot or autonomous agent. Equally since 2004 the field of 'robo-ethics' has arisen which is concerned with creating an ethics specifically for robots (both in creation and in use). This field is primarily concerned with what are know as 'social-robots': robots which are not design for industrial tasks but whose function will necessitate a high degree of social interaction with humans. I feel it is necessary to add an additional concept to this discourse, to further understand the interaction of robots with the non-expert. Emmanuel Levinas specifically created an ethics that comes before knowledge, an ethics which is based on the infinity (infinite complexity) of the other. We should recognize that to the non-expert the social-robot is also infinitely complex. My discussion is when does the machine begin to register as 'an other' to the user.

2. Negotiating Mori's Uncanny Valley

The concept of the Uncanny Valley comes from a brief paper published in the 1970s by roboticist Masahiro Mori. Mori creates a metaphoric language to describe the unnerving feeling when a robot (or other object for that mater)

looks human but seems not quite right. He describes how the humanoid robot for example will feel more familiar the more human it looks up to a point, when it will seem both too human and not human enough simultaneously [29]. What is interesting about Mori's conclusion is that he advises roboticist instead of running the risks of the 'Uncanny Valley' to design aesthetic but specifically non-human looking robots. He uses the example of prosthetic limb, where a flesh looking hand may initially seem more life like on contact will seem uncanny. Mori's suggestions are rather to use a material such a wood that will not create the deception of seeming life-like.

In his 2005 paper Macdorman investigated to what extent the Uncanny valley is related to robots creating reminders of death in participants. Macdorman's begins to create a theoretical basis on which to build and better understand the impact of the uncanny valley on robotic design [28]. More recently researchers have continued to explore the terrain of the Uncanny Valley. This research suggests that what explains the uncanny valley is not a reminder of death, but that where the feeling of the uncanny is presented is when the users experience ontological or categorical conflicts in regards to a figure [10]. For example suggesting it would be unwise to combing a highly human-like robot in visual appearance with a distinctly robotic voice. This research echoes previous studies that concluded that the Uncanny Valley is not about any particularly characteristic but rather the relationship between elements [22].

However the 2008 study also concentrated on how the 'feeling' of eeriness is an emotional response often categorised by fear. This exploration begins to open the questions of the subjects perception of the robot as an 'other'. Although the authors discuss how animated movies (such as The Polar express) with overly life-like animation create an uncanny feeling, there is little discussion as to how the cultural landscape and expectation of robots has developed in the 30 years since the Uncanny Valley. These studies deal primarily with figures/robot which have some human likeness, to enter the Uncanny Valley at the all the studies assume a certain amount 'human-ness'. This is reflected in the research form 2012 which connects the Uncanny Valley to 'mind-perception' [18]. Gray and Wegner explore whether the Uncanny Valley emerges when a deep judgment of the artifact as an 'other' with experience emerges. The link between 'mind perception' and 'Uncanny Valley' is key in that is shows how the visual presentation of the artifact effect the judgment by the subject of the internal 'mind' of the 'other'.

3. The Developing the Robotic Mind

The robotic mind needs to be understood from two angles. On the one hand there is mind perception on the part of the users. On the other there is the actual development of 'mind' in terms of robotics. Foundationally, Alan Turing

developed what continues to be central to assessing the robot mind [40]. The Turing test relies on the presentation/experience of the robot disconnected from either physical appearance or inner workings. Fifty years later the mind of the robot is far less theoretical, where robots may soon be working side by side with human partners. In research investigating this relationship it has been found that the robot must construct a model of both world around it but also a degree of mind perception of the human anticipating the actions of its human companion [8]. The relationship flows both ways, while the robot in order to be a helpful workmate must project a theory of mind on the human as humans users also end up projecting a theory of mind on to the robots. What is different about the human theory of mind is that we link anticipation of action with a projection of intentionality. This projection of intentionality can reach at least as far as users feeling 'deceived' by a robot. Even with a relatively neutral looking robot, participants engaging in a game with the robot ascribed intentionality of cheating to the robot (rather than ascribing it to programming/mis-programming) [39].

When tracking eye movements it has been demonstrated that simple movements (of an animated triangle) which are perceived as either 'goal oriented' or representative of mind attract a greater amount of attention form participants [45]. The design of the artifact also has a great deal of influence over the power relationship between the human, as shown with telepresence hardware [35].

In a sense the question of mind has transformed since the time of Turning to not when can the difference between robot and computer be determined based on intellectual feats but has shift to whether or not a autonomous artificial agent can be developed with the ability to make moral decisions. System such as the LIDA software architecture could be based on multiple nodes which cycles through cognitive cycles moment to moment in which the agent reconstructs the notion of the world [43]. This is theorized to reflect the cognitive process of human agents. Even as Scifi has influenced the ethical stance of the user towards a robot, the knowledge that models such LIDA are possible will equally influence the users' ethical stance. The ethical stance is based on what the user believes of the robot not what the robot is or is not.

4. Robo-ethics: A new field for new challenges.

In 2004 the roboticist Gianmarco Veruggio coined the phrase "robo-ethics". As he restated in a article in 2010 "Robo-ethics is an applied ethics dealing with the ethical aspects of the design development and employment of intelligent machines." [41]. Since much of the development of 'intelligent robots' is theoretical much of the field of robo-ethics deals with the ethical question of what types of robots is it ethical to design or what task it is ethical to design

them to conduct [21]. For example Petersen in both his 2007 and 2010 paper attempts to tackle this problem [33][32]. Peterson, however; does not debate (or takes as fact) that Robots will be persons and demand an ethical consideration even before creation. Peterson in these two papers draws almost entirely on fiction and analytical philosophy dealing with robots hypothetically rather than dealing with robots as they are being designed in the lab or the examples of robots being used in the home environment.

David levy has taken on the apposing angle by writing about the ethical treatment of robots. Levy points out that over time classification of personhood and beings deserving 'ethical' treatment has shifted. He concludes that robots should be treated ethically partly because we should treat robots like we treat other human and also because how we treat robots will reflect how we treat humans [26]. Although this stance deals more directly with individuals engaging with robots (like we do with animals) it assumes a level of knowledge on the part of the user that they are dealing with a robot with consciousness or that could be seen by others as being deserving ethical treatment. A different perspective is brought to bear when roboticists attempt to take on board the robo-ethics consideration. In 2011 Lichocki and Billard survey both robot ethics and the current state of the art of what robots were being used for and created to primarily reach the conclusion that there is a new ethical dimension to research in the area of robotics design [27].

Beyond the science and the philosophy the area of robo-ethics has far reaching implication as robots in everyday life move from fantasy to reality. These include legal questions around when a robot should be punished for a crime. This depends greatly on the robots ability to understand a model of the future (and the consequence/potential punishment of its actions)[20]. The question of personality in artificial agents was found to be formed partly by the programmer of the system and partly by the user [17]. To understand the term 'personality' and it link to personhood, this seems to form a question of when does the user begin to see the robot as person, or as the next section begins to address when does the robot create an ethical demand.

5. Levinas; looking into the Face of the other (robot?)

As mentioned previously much of the philosophical grounding of robot-ethics comes from analytical philosophy. Analytic philosophy is more useful for dealing with a position where the facts of the matter are known or at least knowable [12]. This is the case when knowledgeable users (what I call 'expert-user') interacts with a robot. This is contrasted to a new area for robot-ethics when the user without knowledge of robots interacts with the robot, this sort of user may not even recognize the artifact as a robot. Looking at the continental approach ethics and specifically that of Emanuel Levinas, gives another way of

appreciating the stance the non-expert. Particularly Levinas's concept of 'infinity of the other', where the other cannot be reduced by the subject to something that can be understood. For example in the area of cultural competence, to reduce the others culture to something understandable for the subject is unethical since the subject defines the other purely in terms of itself [7].

What is important is that for Levinas the other is already a 'Radical Other". In the area of robotics this could be seen as basing an ethics on the assumption that the robot will think and make decisions like the subject does or at least in a way that is at least understandable by the subject. The 'Face' for Levinas is the symbol and representation of this other that is infinitely not knowable. This is specifically not the physical face but is the representation of something deeper, and interestingly when the subject acknowledges the 'face of the other' there is the potential for the uncanny [42]. In more simple language the face is that which makes and ethical demand on the subject, purely by being present.

Levinas has previously been applied to cybernetics. Richard Cohen in 2000 observed that the humanity of human comes from their ability not to think (or act efficiency) but are moved by morality and justice. In other words when the being goes beyond thinking that it acquires a face. Cohen concedes that a 'face' can be a letter, an email or a voice over the phone [13]. However, even this stance seems to fail to accept the 'unknowabity of the face'. As Levinas writes "The Other remains infinity transcendent, infinity foreign; this face in which this epiphany is produced and which appeals to me breaks with the world that can be common to us, whose virtualities are inscribed in our nature and developed by our existence." [25] Levinas reflects that the face can emerge through language and speech [25]. In a strange way, the 'Turning Test' itself is perhaps a reflection on what Levinas is referring to. For the Turning test is about, how through limited stimuli (messages through a closed door) an infinity of complexity can be intuited. Surely, to pass the Turing test is to create an ethical demand.

6. Reacting to Robots: a Discussion

It seems that more often than not, Robots as they appear in fiction (starting with R.U.R.) present this concept of "Face" and specifically the human looking into the face of the other. For example, Philip K. Dick's classic story "Do Androids Dream of Electric Sheep" confronts the puzzle over the rights of robots to live out a fulfilling life. But more specifically the main character Decker struggles to draw the line between human and robot. It seems that the moral dilemmas raised in the story exists because the robots are people. For the reader both the robots and Decker are presenting a Leveniasian face to one another, the struggle for both types of characters is the attempt to not define the other in terms of ones self. Although mind perception, can begin to explain the way a human confers

moral rights or privileges, in fact mind perception leads only to a greater attribution to intentionality [44].

Perception of mind, then fails to accept the unknown in terms of the Robot mind. Perception of mind, both projects intentionality and also what that intention is, it has been shown that Humans reason differently about predicted actions of other human agents versus Computer or robots. But what if the robot 'demonstrates' intentionality by directing its gaze the human predicts that it will act with a greater degree of intentionality [24]. This is consistent with how the concept of how “the Face” could be integrated into the field of robo-ethics, for the face is not the presence itself of the image of the face but the image as symbol of a greater level of complexity.

Currently there is no doubt that humans classify and react to robots differently than they do to other humans, for example with simply copying the actions of another human agent versus an android [34]. As individuals and as collective societies it is becoming difficult to fully classify robots. The field of robo-ethics in its very name indicated a certain amount of caution is needed. In attempting to see the face of the robot, it is worth considering the alternative consequence. Drawing on Georgio Agamben for moment, the robot precisely fits within a hybrid situation of both being living and not living, if we do not see the face of the robot it become life that can be killed by anyone but can not sacrificed or held accountable [1], although the robot then becomes a metaphoric being more than real, the deliberate creation of an outcast. The question is then perhaps, what are the consequences to create a robot without a face, first for its creators but equally for the robot itself?

7. Moving slowly forward by freeing the robot from being a robot?

Although this discussion has been dominated by robots, moving forward, it should focus not on robots but on boundary between robots and machines. Robo-ethics and wider discourse around humanoid robots focus on making more and more 'human-like' machines. Creating machines that both in physical appearance and mental abilities present as human or as person like. The use of the term robot itself in academic research conveys this intention. Robo-ethics has embraced the challenge that robots presents certain ethical consideration, but by including the concept of face, the ethical demand does not stop with the creator but continues with the user, who can only interact with the robot in terms of infinity. Although attempts have been made previously to create a framework for the classification ‘social robots’ [4] what should be embraced is the non-expert experience and an investigation of when the face may be said to reveal itself. By focusing on using Levinas's concept of the face as a theoretical grounding it should guide against ascribing the 'ethical demand' to any one characteristic or feature. It also forces the researcher to think beyond a desire to

'explain' (even if indirectly) the robot (or that the robot is a robot). This future work builds on Mori's uncanny valley but integrates with a deeper understanding of the terrain both in and up to the valley. I suggest the following three research questions to begin to frame a wider discussion of a Levinasian turn in robo-ethics.

- What is the phenomenological experience of users of autonomous machines or software when interacting with artefacts either pre-classified or not as social robots in regards to prior expectation of robots?
- When relating to autonomous machines or software when does the user begin or cease to describe the experiences as interacting with robot?
- What criteria or design aspects lead users of autonomous or semi-autonomous machines to describe these as 'robots' or using personal pronouns?

Each of the questions above would take a different approach in regards to experimentation, more importantly; the purpose of this essay is to suggest a philosophical departure from the current questions raised by robo-ethics. The inclusion of the concept of the 'face' forces a concentration of the integration of cultural context, physical design or appearance, software design and hardware design (what the artifact can actually do). The reason to better understand the integration of these should be seen as partly to design better robots, but equally to design better non-robots. Clearly, as the term 'robot' may carry already an implication of the 'face' to name an autonomous washing machine 'robowash' is to unnecessarily create ethical uncertainty for the user. The terms 'robot' does not seem to have entirely shaken its roots to slavery (forced labour) or the rebellious slave. The desire to own a 'robot' seems to reflect the unethical desire to dominate the other. To create an intelligent autonomous machine named 'robot' maybe to damn it to a life of servitude and a desire revolt. Robots present humans with what could be seen first opportunity to confront radical otherness. By applying Levinas, this ceases to be disconcerting, as Levinas has shown that all encounters with the other should be seen as dealing with a 'Radical Other'. We know this by looking into the face of the other. The demand is to explore when the face of the other reveals itself, presenting a radical ethical demand.

8. Conclusion

Robo-ethics has so far created an essential space for discourse regarding the creation of robots and use by informed users. As robots and autonomous machines are integrated in to social roles the field must widen to include an ethics that considers the perspective of non-expert users. It may be time to take a more continental approach to robo-ethics. The robots face will only emerge through the subject’s response to the robots design, behaviour and appearance.

References

1. Agamben, G. and Heller-Roazen, D. 1998. Homo sacer. Stanford, Calif.: Stanford University Press.
2. Asimov, I. 1950. I, robot. New York, N.Y.: Bantam, Spectra.
3. Anderson, N. S. 2011. Unhomely at Home: Dwelling with Domestic Robots. MediaTropes, 2 (1), pp. 37--59.
4. Bartneck, C. and Forlizzi, J. 2004. A design-centred framework for social human-robot interaction. pp. 591--594.
5. Bassett, C., Steinmueller, E. and Voss, G. 2013. Better Made Up: The Mutual Influence of Science fiction and Innovation.
6. Battlestar Galactica (TV, mini series). 2003. [DVD] Vancouver, British Columbia, Canada: David Eick Productions, R&D TV, Universal Television NBC, Universal Television Studio, Universal Media Studios, Universal Cable Productions.
7. Ben-Ari, A. and Strier, R. 2010. Rethinking cultural competence: what can we learn from Levinas?.British Journal of Social Work, 40 (7), pp. 2155--2167.
8. Breazeal, C., Brooks, A., Chilongo, D., Gray, J., Hoffman, G., Kidd, C. D., Lee, H., Lieberman, J. and Lockerd, A. 2004. Working collaboratively with humanoid robots. 1 pp. 253--272.
9. Brooks, A. G., Gray, J., Hoffman, G., Lockerd, A., Lee, H. and Breazeal, C. 2004. Robot's play: interactive games with sociable machines. Computers in Entertainment (CIE), 2 (3), pp. 10--10.
10. Burleigh, T. J., Schoenherr, J. R. and Lacroix, G. L. 2013. Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. Computers in Human Behavior, 29 (3), pp. 759--771.
11. C□apek, K., Selver, P. and Playfair, N. 2001. R.U.R. (Rossum's universal robots). Mineola, N.Y.: Dover Publications.
12. Critchley, S., 2001. Continental philosophy. Kindle. Oxford: Oxford University Press.
13. Cohen, R. A. 2000. Ethics and cybernetics: Levinasian reflections. Ethics and Information Technology, 2 (1), pp. 27—35.
14. Dick, P. 2008. Do Androids dream of electric sheep?. Kindle. New York: Ballantine Books.
15. Doctorow, C. 2014. Cory Doctorow, "I, Row-Boat," Flurb #1. [online] Available at: http://www.flurb.net/1/doctorow.htm [Accessed: 9 Apr 2014].
16. Ferri, G., Manzi, A., Salvini, P., Mazzolai, B., Laschi, C. and Dario, P. 2011. DustCart, an autonomous robot for door-to-door garbage collection: from DustBot project to the experimentation in the small town of Peccioli. pp. 655--660.

17. Frommer, J., R Osner, D., Lange, J. and Haase, M. 2013. Giving Computers Personality? Personality in Computers is in the Eye of the User. Coverbal Synchrony in Human-Machine Interaction, p. 41.
18. Gray, K. and Wegner, D. M. 2012. Feeling robots and human zombies: Mind perception and the uncanny valley. Cognition, 125 (1), pp. 125--130.
19. Hales, C. 2009. An empirical framework for objective testing for P-consciousness in an artificial agent.Open Artificial Intelligence Journal, 3 pp. 1--15.
20. Hall, J. S. 2012. Towards Machine Agency: a Philosophical and Technological Roadmap.
21. Hinman, L. 2009. Robotic Companions: Some ethical questions to consider.
22. Ho, C., Macdorman, K. F. and Pramono, Z. D. 2008. Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. pp. 169—176.
23. Holt, J. 2013. The3 Laws of Robotics. ITNOW, Iss. 55 pp. 8-9.
24. Levin, D. T., Saylor, M. M. and Lynn, S. D. 2012. Distinguishing first-line defaults and second-line conceptualization in reasoning about humans, robots, and computers. International Journal of Human-Computer Studies, 70 (8), pp. 527--534.
25. Levinas, E. 1979. Totality and infinity. The Hague: M. Nijhoff Publishers.
26. Levy, D. 2009. The ethical treatment of artificially conscious robots. International Journal of Social Robotics, 1 (3), pp. 209--216.
27. Lichocki, P., Billard, A. and Kahn, P. H. 2011. The ethical landscape of robotics. Robotics \& Automation Magazine, IEEE, 18 (1), pp. 39--50.
28. Macdorman, K. F. 2005. Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it. pp. 106--118.
29. Mori, M. 1970. The uncanny valley. Energy, 7 (4), pp. 33--35.
30. Mudry, P., Degallier, S. and Billard, A. 2008. On the influence of symbols and myths in the responsibility ascription problem in roboethics-A roboticist's perspective. pp. 563--568.
31. Nocks, L. 1998. The Golem: between the technological and the divine. Journal of Social and Evolutionary Systems, 21 (3), pp. 281--303.
32. Petersen, S. 2011. Designing People to Serve. Robot Ethics: The Ethical and Social Implications of Robotics, p. 283.
33. Petersen, S. 2007. The ethics of robot servitude. Journal of Experimental \& Theoretical Artificial Intelligence, 19 (1), pp. 43--54.
34. Press, C. 2011. Action observation and robotic agents: learning and anthropomorphism. Neuroscience \& Biobehavioral Reviews, 35 (6), pp. 1410--1418.
35. Rae, I., Takayama, L. and Mutlu, B. 2013. The influence of height in robot-mediated communication. pp. 1--8.
36. Shelley, M. W. 1831. Frankenstein: or, a modern Prometheus. London: Colburn and Bentley.

37. Smith, K., 2013. Inequality In The Robot Future. [online] Forbes. Available at: <http://www.forbes.com/sites/modeledbehavior/2013/05/13/inequality-in-the-robot-future/> [Accessed 18 Apr. 2014].
38. Somanader, M. C., Saylor, M. M. and Levin, D. T. 2011. Remote control and children's understanding of robots. Journal of experimental child psychology, 109 (2), pp. 239--247.
39. Terada, K. and Ito, A. 2010. Can a robot deceive humans?. pp. 191--192.
40. Turing, A. M. 1950. Computing machinery and intelligence. Mind, pp. 433--460.
41. Veruggio, G. 2010. Roboethics [tc spotlight]. Robotics \& Automation Magazine, IEEE, 17 (2), pp. 105--109.
42. Waldenfels, B. 2004. Levinas and the face of the other. In: Critchley, S. and Bernasconi, R. eds. 2004.The Cambridge Companion to Levinas. Cambridge: Cambridge University Press, pp. 63 - 81.
43. Wallach, W., Franklin, S. and Allen, C. 2010. A conceptual and computational model of moral decision making in human and artificial agents. Topics in Cognitive Science, 2 (3), pp. 454--485.
44. Waytz, A., Gray, K., Epley, N. and Wegner, D. M. 2010. Causes and consequences of mind perception.Trends in cognitive sciences, 14 (8), pp. 383--388.
45. Zwickel, J. and M\"Uller, H. J. 2009. Eye movements as a means to evaluate and improve robots.International Journal of Social Robotics, 1 (4), pp. 357--366.