Ethics by Design:
A Conceptual Approach to Personal and Service Robot Systems

H. F. Machiel Van der Loos, PhD, *Member, IEEE*

Abstract—**The design of robots is driven largely by application area. Industrial robots have no need to be lightweight, soft or compliant, and, being inherently unsafe to humans, they are kept physically out of reach of factory workers. This approach is not tenable for personal and service robots, since, in addition to having an informational interface, people will expect human-robot interaction to be familiar and as safe as daily human-human interaction. This position paper describes how approaching both the software design and hardware design to mimic human attributes in behavior, physical aspects and motion quality will embed ethics in a deep and fundamental fashion, very different from artificially constructing a control scheme on top of an intrinsically human-unsafe structure.**

I. Introduction

UNTIL this decade, the ethics of the deployment of robots has mainly focused on societal issues related to displacing human workers in industrial contexts. It has only been in science fiction, from the provocative works of Capek [1] and Asimov [2] to recent motion pictures such as I, Robot [3] and Spider-Man 2 [4], that people have been exposed to cyborg and robot scenarios and issues likely to become real at some point in the future, such as moral agency, free will, identity and the coexistence of humans and robots in the same environment. While we can examine science fiction's thought-provoking exercises to spark discussion, and while we can applaud their creators for developing scenarios that portray believable human values of the future, they do not provide a path to guide us today in developing the desirable and avoiding the undesirable technologies and societal constructs depicted. This task is one Society must do by itself, and for it to unfold responsibly, the roboticists of today need the philosophy, psychology, law and other professional communities to develop the ethics scaffolding to guide the physical, electronics, control and software design.

Robotics is an implicit member of the group of technologies collectively known as the Nano-Bio-Info-Cogno (NBIC) Convergence [5]. As these domains mature and combine into technologies of the future, a common set of ethical issues [6] will emerge to study and (re)solve. Whether or not the ethics of robotics will in the future remain a separate identity or will be subsumed under the NBIC collective, the fields have much to share when discussing ethical issues. Initiatives such as the NSF Report on Ethics [5], the ICRA 2005 Roboethics Technical Committee meeting [7] and Roboethics Atelier that spawned the EURON Roboethics Roadmap [8] have begun this work.

Building on these first substantial steps, this paper will describe an overall biomimetic strategy to help inform the design process. This strategy is not meant as a template but rather as an optic for designers to consider when making decisions on robot form and function.

II. Safety and Coexistence

A. Robot Safety

Industrial robot safety has for decades been the subject to regulation and standards, but the end result is always that humans and robots are mandated to be kept physically separate except in very precisely constructed scenarios, such as on-site repair and programming. Although ISO suggests that parts of the current standard, ISO 10218-2006 [9], may be useful in non-industrial robotics applications, there are no ISO regulations specifically for service and personal robots, much less cyborgs. In their understandable absence given the state of robot technology and intelligence software today, engineers' professional codes of ethics are the fall-back, such as the IEEE code of ethics, whose ten rules start with the commitment of its members to "… accept responsibility in making decisions consistent with the safety, health and welfare of the public…" [10].

B. Rehabilitation Robotics

The absence of standards has not stopped human subjects studies and commercial product development of robots occupying the same environments as people. Notably, in rehabilitation applications, in which arguably the user experiences the closest physical relationship seen in the personal and service robot sector, industrial robots (e.g., the PUMA [11],[12]) and specially-designed mechanisms (e.g., the MIT-MANUS [13] and Lokomat treadmill robot [14]) have been used with a notable lack of reported accidents. The reasons for this result are careful design and redundant layers of physical, electrical and software security, plus a constant vigilance on the part of caregivers and researchers

Manuscript received January 20, 2007. This work was supported in part by the "Ethical Dimensions in Neuroscience" Program of the Stanford University BioX Interdisciplinary Initiative and the Stanford Center for Design Research.

H. F. Machiel Van der Loos is with the Dept. of Veterans Affairs Palo Alto Rehabilitation R&D Center and with Stanford University in the Dept. of Mechanical Engineering and the Center for Biomedical Ethics (phone: 650-493-5000 #65971; fax: 650-725-8475; e-mail: vdl@stanford.edu).

[15]. However, as rehabilitation and personal care robotics products move into the home and office where the consumer's behavior is the only real human safeguard, these provisions are not sufficient, especially as the robots become human-scale, which we can expect to happen in the future. The task-selection interface, which with today's software is more reminiscent of a bank teller machine, will gradually evolve in functionality to the point where ethical consideration of tasks the human requests it to perform becomes mandatory.

C. *Companion and Pet Robotics*

Outside of the rehabilitation arena, there is a neighboring product sector in companion robots. Several small pet robots designed not as toys but as therapeutic aids have been developed. Paro [16] is a small, white robotic seal, approximately the size, weight and feel of a small cat, with sensors and actuators and adaptive interface software running its on-board controller. It is not mobile, but can move its body segments in response to input from its user. T. Shibata, the inventor, purposely chose a seal metaphor and not a more common pet species, since people are not as familiar with "seal-ness" as with "cat-ness". Had he attempted to design a realistic robotic cat, people's high familiarity with the species would have put a great onus on the designer to get every nuance of "cat-ness" right to avoid the user ending up in the depths of the "Uncanny Valley" (fig. 1) [17] and rejecting the pet for its companion value. A seal, on the other hand, allows users to consider it just a small furry mammal, and it is much easier to design generic interactions and responses representative of "mammal-ness", thereby placing users in a different mindset for its acceptance. These design considerations represent the type of reflection necessary for any subsequent generation of animal-like and humanoid robots to properly frame user expectations.

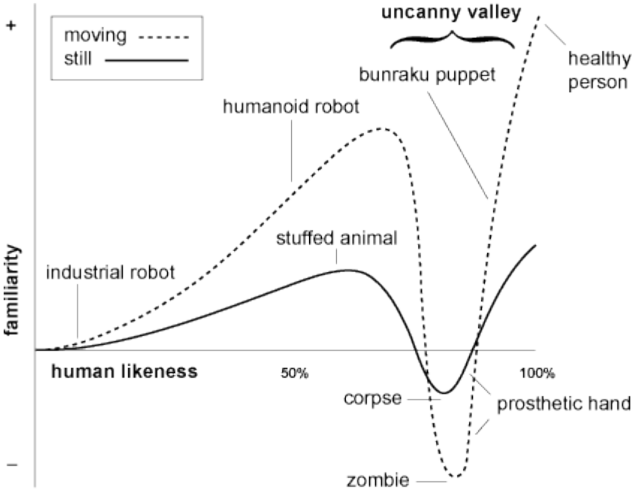

Figure 1. The Uncanny Valley [17]. Familiarity and acceptance rise as a robot becomes more human-like; however, at a certain point, the behavior is too uncannily human-like but still not perfect, and the imperfections reverse the attraction and trust. Only when the behavior becomes much more natural will normal patterns of interactions be possible and familiarity rise once more.

The recent announcement of Ugobe's Pleo [18], a small mobile dinosaur robot, underscores that it is being marketed as a pet, not a toy, with software that makes it "lifelike". However, it is also clear that since no one of course can really know the behavior and movement patterns of the real Camarasaurus species, Pleo's programmed behaviors are more generically lifelike based on this era's animal species, which is the most appropriate design strategy to gain acceptance by today's consumers.

D. *Companion and Humanoid Robots*

Human-size robots designed as companions or servants to people have been designed in fact and fiction for centuries. Actual humanoid robots, however, have not yet made it to the right side of the Uncanny Valley, or put another way, none has yet passed a Turing Test for human-ness, with all its non-perfectness. Several designs have made partial inroads. For example, the Actroid full-body (but not mobile) female humanoid robot [19] is very realistic in terms of speech, facial and gestural expressions, in fact, *too* perfect to be a true human. Other companion robots, such as the Wakamaru, have made no pretense of human-ness, preferring to convey an unmistakably robot identity [20] yet building in human-like communication behaviors, such as holding its hand to its forehead when retrieving information wirelessly from the Internet for its user, which on a computer would be a twirling hourglass or other "I'm thinking about it" symbol. The Cog Project, including the Kismet face robot [21] at MIT is another that seeks to mimic human behavior and motion with the goal of better understanding the relationship between the two. The Sony SDR and Honda ASIMO humanoid robotics projects have created numerous iterations of increasingly functional, mobile, dynamically walking and running robots. While not made to be human-like, per se, encased in hard plastic with backpacks and helmet-like heads, they move like humans. Their experimental nature becomes evident when pushed to the boundaries of control, with several recent examples of full system failure and painful (to watch) falls when ascending and descending stairs. They are, of course, allowing new and important future explorations in more robust control strategies of dozens of degrees of freedom simultaneously [e.g., 22] to accomplish real-world tasks.

People's reactions to these exemplars of humanoid robotics carry with them numerous conflicts. While we are fascinated with the inventiveness of their creators, we also note with trepidation the semblance of a measure of independence that accompanies their movements and behaviors interacting with humans. When the Paro seal behaves cuddly on being caressed, or when the Wakamaru robot moves around the house as a sentry, what amount of behavior do we have to observe to feel safe and not have to be constantly vigilant? In addition, are we identifying with their developers or with the robot itself? What sort and level of robot adaptability of behavior, in other words, deviations from earlier observations in similar situations, will we tolerate? How are plans and actions conceptualized [23], robot "brains" structured [24],[25] and application areas approached [26]? These are design questions for the developers, and they will become harder as robot behavior become more reliant on intelligence software and not just sensory and motion-control modules and layers.

III. Ethics and a Biomimetic Robot Design Strategy

The discussions above focus on robot design and our reactions to observed robot behavior. The overlap with ethics resides in the concept of transparency: for true biomimesis, hardware developers and programmers of interface and intelligence software must pair each new layer of complexity in robot behavior with a corresponding layer of explicit attention paid to conveying those behaviors to the surrounding people (and robots, too). It is not necessary that a robot be fully human-like in physical capability, but for the actions that it is capable of exhibiting, it must be capable of communicating the intention of doing them through, for example, gestures, voice and context. In this sense, if a Turing Test could be developed, it would not examine how lifelike a robot acts, but rather how human-like it is capable of communicating its intention to act (see, e.g., [27] on mimicry and imitation in computing machines).

The imitation of human behavior can increase layer by layer, but at each layer, the robot's designed-in thought processes and adaptability must be brought out. José Galvan argues that a robot will never have free will [28] since it will always be a product of our technological creation, whereas Ray Kurzweil argues that compute power will in less than a few decades make it possible to create software that is indeed smarter than humans [29]. Whether one or the other is right (or both may end up being right or wrong), it is still incumbent on designers to keep the concept of conveying the interaction between action and communication in the forefront.

IV. Conclusion

At this point in history, professional ethics is guiding humanoid robot design, in computer science and engineering. Implementing ethics in a robot is in its infancy [30]. The implementation of any theoretical philosophical framework, whether based on Kant, utilitarianism, consequentialism, virtue ethics, casuistry, or another, becomes all the more difficult when the theory has to be reduced to algorithms and rules [31]. On top of that, cultural differences need to be be handled as well [32],[33]. Indeed, even if one theory were implemented, would it be the right one for robots? Who will decide?

References

[1] K. Capek. RUR: *Rossum's Universal Robots*, 1920. Translation in English: Washington Square Press edition, Simon and Schuster, 1973.

[2] I. Asimov . *I, Robot.* Ballantine Press, NY, 1950.

[3] *I, Robot*, Fox Filmed Entertainment, Inc., 2004.

[4] *Spider-Man 2*, Sony Pictures Digital, Inc., 2004.

[5] M. Roco, W. Bainbridge (eds.), *Converging Technologies for Improving Human Performance – Nanotechnology, Biotechnology, Information Technology and Cognitive Science*, NSF Report, Arlington, VA, 2002.

[6] G. Veruggio and F. Operto. First International Symposium on Roboethics: The ethics, social, humanitarian and ecological aspects of robotics, January 30-31, 2004, Villa Nobel, Sanremo, Italy.

[7] P. Dario, K. Tanie, R. Arkin, IEEE Technical Committee on Roboethics. accessed December 1, 2005 http://www-arts.sssup.it/IEEE_TC_RoboEthics/.

[8] G. Veruggio (ed.), EURON Roboethics Roadmap, *Proceedings EURON Roboethics Atelier*, Genoa, 2/27-3/3, 2006.

[9] ISO Robot Safety Standard 10218-1:2006, International Organization for Standardization, June, 2006.

[10] IEEE Professional Code of Ethics, accessed January 19, 2007: http://www.ieee.org/web/membership/ethics/code_ethics.html

[11] H. F. M. Van der Loos, VA/Stanford Rehabilitation robotics research and development program: Lessons learned in the application of robotics technology to the field of rehabilitation. *IEEE Trans. Rehabilitation Engineering*, Vol. 3, No. 1, March, 1995, pp. 46-55.

[12] R. M. Mahoney, H. F. M. Van der Loos, P. S. Lum, C. G. Burgar, Robotic stroke therapy assistant, *Robotica,* Volume 21, Issue 01. January 2003. pp. 33-44.

[13] B. T. Volpe, H. I. Krebs, N. Hogan. Robot-aided sensorimotor training in stroke rehabilitation. *Adv Neurol*. 2003;92:429–33.

[14] Lokomat treadmill robot, Hocoma, Inc., accessed January 19, 2007: http://www.hocoma.ch/.

[15] H. F. M. Van der Loos, D. S. Lees, L. J. Leifer, Safety considerations for rehabilitative and human-service robot systems. *Proc. RESNA 15th Annual Conference*, Toronto, Canada, June, 1992. pp. 321-324.

[16] T. Shibata, K. Wada, K. Tanie,. Tabulation and analysis of questionnaire results of subjective evaluation of seal robot in Japan, U.K., Sweden and Italy. *Proceedings. IEEE International Conference on Robotics and Automation*, Volume 2, 2004, pp. 1387-1392.

[17] M. Masahiro, The uncanny valley (translated by K. F. MacDorman and T. Minato), *Energy, 7*(4), 1970, pp. 33-35. Figure copied under the "GNU Free Documentation License" from http://en.wikipedia.org/wiki/Image:Moriuncannyvalley.gif

[18] Pleo announcement, Ugobe, Inc., accessed January 19, 2007: http://www.ugobe.com/

[19] Actroid, Kokoro Company, Ltd., accessed January 20, 2007: http://www.kokoro-dreams.co.jp/english/robot/act/index.html

[20] Wakamaru, Mitsubishi Heavy Industries, accessed January 20, 2007: http://www.mhi.co.jp/kobe/wakamaru/english/about/index.html

[21] R. A. Brooks and L. A. Stein. Building brains for bodies. *Autonomous Robots*, 1: 7–25, 1994.

[22] L. Sentis, O. Khatib, Synthesis of whole-body behaviors through hierarchical control of behavioral primitives, *International Journal of Humanoid Robotics*, Vol. 2, No. 4, 2005, pp. 505-518.

[23] L. A. Suchman. *Plans and Situated Actions: The Problem of Human/Machine Communication*. Cambridge University Press, Cambridge U.K., 1987.

[24] B. Scassellati. Theory of mind for a humanoid robot. *First IEEE/RSJ International Conference on Humanoid Robotics*, September, 2000.

[25] M. Kawato, Robotics and BMI/BCI interface research in Japan, 1st International Workshop on Neuroethics in JAPAN: Dialog on Brain, Society, and Ethics, 2006, Tokyo, Japan.

[26] D. Moreno. *Mind Wars: Brain Research and National Defense*. Dana Press, Washington, DC, 2006.

[27] B. P. Bloomfield and T. Vurdubakis, Imitation games: Turing, Menard, Van Meegeren, *Ethics and Info. Tech*. 5: 27–38, 2003.

[28] J. M. Galvan, On technoethics, *IEEE-RAS Magazine*, 10 (2003/4), pp. 58-63.

[29] R. Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*. Penguin Group, New York, NY, 2005.

[30] B. M. McLaren, Computational models of ethical reasoning: Challenges, initial steps, and future directions, *IEEE Intelligent Systems*, vol. 21, no. 4, pp. 29-37, July/August, 2006.

[31] J. Gips, Towards the ethical robot, in: *Android Epistemology*, K. Ford, C. Glymour, P. J. Hayes (eds.). AAAI Press, 1995, pp. 243-252.

[32] J. J. Wagner, D. M. Cannon, H. F. M. Van der Loos. Cross-cultural considerations in establishing roboethics for neuro-robot applications, *Proc. ICORR'2005*, Chicago, IL, USA, June 28-July 1, 2005.

[33] N. Kitano, A comparative analysis: Social acceptance of robots between the West and Japan, *EURON Atelier on Roboethics*, 2006.