Editorial

Anthony F. Beavers

Published online: 13 July 2010
© Springer Science+Business Media B.V. 2010

It has already become something of a mantra among machine ethicists that one benefit of their research is that it can help us better understand ethics in the case of human beings. Sometimes this expression appears as an afterthought, looking *as if* authors say it merely to justify the field, but this is not the case. At bottom is what we must know about ethics in general to build machines that operate within normative parameters. Fuzzy intuitions will not do where the specifics of engineering and computational clarity are required. So, machine ethicists are forced head on to engage in moral philosophy. Their effort, of course, hangs on a careful analysis of ethical theories, the role of affect in making moral decisions, relationships between agents and patients, and so forth, including the specifics of any concrete case. But there is more here to the human story.

Successfully building a moral machine, however, we might do so, is no proof of how human beings behave ethically. At best, a working machine could stand as an existence proof of one way humans could go about things. But in a very real and salient sense, research in machine morality provides a test bed for theories and assumptions that human beings (including ethicists) often make about moral behavior. If these cannot be translated into specifications and implemented over time in a working machine, then we have strong reason to believe that they are false or, in more pragmatic terms, unworkable. In other words, robot ethics forces us to consider human moral behavior on the basis of what is actually implementable in practice. It is a perspective that has been absent from moral philosophy since its inception.

Wendell Wallach's paper, "Robot Minds and Human Ethics: The Need for a Comprehensive Model of Moral Decision Making," shows how machine ethics is engaged in the assessments mentioned above. He calls for a comprehensive model because current attempts at making moral machines show how fragmentary existing moral frameworks really are. Even though moral machines need not function precisely in the way that moral humans do, we stand to learn a great deal about the mechanisms for moral behavior in general by looking at the fine details of implementation.

In "Moral Appearances: Emotions, Robots and Human Morality," Mark Coeckelbergh looks at the role of affect in moral behavior, noting that because robots cannot have emotions, they cannot be full moral agents. Furthermore, we risk a danger in building them without affects, since rule-following by itself can run awry. Morally semblant robots are nevertheless possible and non-problematic, since we understand human morality on the basis of appearances as well. Also from Mark Coeckelbergh, this issue continues with "Robot Rights? Toward a Social-Relational Justification of Moral Consideration," in which the author presents an ecological framework that, when it comes to extending moral regard, fits both the cases of intelligent robots and human beings.

A hotbed for debate concerning moral machines surrounds military applications that are used in warfare. Because machines, now armed and proliferating, engage action on the battlefield, careful attention to a variety of ethical issues is immediately demanded. In "RoboWarfare: Can Robots Be More Ethical than Humans on the Battlefield," John Sullins compares the ethical behavior of human soldiers to the moral claims made by roboticists with the goal of assessing the role robots may play in fighting wars more justly. The next paper, "The Cubicle

A. F. Beavers (✉)
The University of Evansville, Evansville, IN, USA
e-mail: tb2@evansville.edu

Warrior: The Marionette of Digitized Warfare,” by Lamber Royakkers, follows Sullins nicely. In it, Royakkers argues that fighting wars by remote control dehumanizes enemies thereby resituating the ethical playing field. He worries (rightly so, perhaps) that in the end, we will be unable to hold these “cubical warriors” responsible for their decisions.

Yvette Pearson and Jason Borenstein turn our attention from war to care in “Robot Caregivers: Harbingers of Expanded Freedom for All?”. They discuss an approach to governing strategies for robots that seeks to expand freedom and promote human flourishing. They argue that this “capabilities approach” will prove more successful than standard ethical theories when it comes to gauging the ethical impact of robot caregivers.

The next paper, “Implications and Consequences of Robots with Biological Brains,” by Kevin Warwick, examines the prospects of using living neural tissue, itself a computational platform, in robots. Imagining the possibility that a robot with a brain might be conscious, Warwick considers the case for giving it moral rights. Finally, in the closing paper, “Designing a Machine for Learning about the Ethics of Robotics: The N-Reasons Platform,” Peter

Danielson discusses a concrete implementation of machine ethics that learns from human responses to experimental surveys and that, on that basis, “evolve[s] an ethically relevant data structure.” Because it relies on human subjects, the N-Reasons platform is also instructive for understanding how humans make ethical choices.

The collection appropriately ends with three reviews of a new book cited in six of the eight papers included here: *Moral Machines: Teaching Robots Right from Wrong* (Oxford, 2009) by Wendell Wallach and Colin Allen. The reviews are presented here from shortest to longest, the first offering an overview of the text, the last an extended critique. I had the privilege of reading this book in draft. In this connection, it is fitting to note that it provided the impetus to compile this special issue of ETIN in the first place.

Additionally, this edition could not have come together without the assistance of Managing Editor Noëmi Manders-Huits, who took care of all the fine details. All of the papers were put through an iterative review process following the standard practice of the journal. I, therefore, wish to acknowledge the several anonymous reviewers, who offered careful commentary, and the authors, who were genuinely receptive to hearing it.