

Is Ethics Computable, Or What Other than *Can* Does *Ought* Imply?

Anthony F. Beavers, Ph.D.
The University of Evansville
<http://faculty.evansville.edu/tb2/>

In 2007, Anderson and Anderson wrote, “As Daniel Dennett (2006) recently stated, AI ‘makes philosophy honest.’ Ethics must be made computable in order to make it clear exactly how agents ought to behave in ethical dilemmas” (16). To rephrase, a computable system or theory of ethics makes ethics honest. But at what cost? Might Turing’s 1950 prophecy that “at the end of the century the use of words ... will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted” (1950, 442) soon take on normative dimensions due to research in artificial morality. Will attempts to make ethics computable lead us to redefine the term “moral” to fit the case of machines and thus change its meaning for humans also? I call this the threat of “moral nihilism ... the doctrine that states that morality needs no internal sanctions, that ethics can get by without moral “weight,” i.e., without some type of psychological force that restrains the satisfaction of our desire and that makes us care about our moral condition in the first place” (Beavers, 2011a).

Analyzing this possibility requires inspection of the meaning of the term “ought” and what it implies. In 2009, I argued that, following Kant, *ought* not only implies *can*, but also *might not*, in which case it would be morally wrong to create artificial Kantian agents, since doing so would require designing them in such a way that they *could* act immorally, but would not do so. Only on such a condition would it make sense to hold a machine responsible for its actions and praise or blame it for its behavior. In 2011, I argued that if *ought* implies *can*, then it also implies *implementability*. If a machine or human *can* act morally, this can only be because the mechanisms (whether in software or wetware) have the requisite components to allow for it. Thus, any theory of morality must be implementable in real working agents to qualify as a viable moral theory. Given the conclusions of 2009, I argued in 2011 that designing machines in such a way that they behaved morally but were not able to act immorally would require redefining the term “morality” in such a way that full moral agency with internal sanctions was not intrinsic to ethics, but “merely a sufficient, and no longer necessary, condition for being ethical.” In this case, internal states such as conscience, responsibility (as felt affective weight) and thus moral accountability are, *ex hypothesi*, not necessary for ethics either. Thus, if we build machines capable of being described by the term “moral” we can only do so by redefining the term. So, if a time is coming when we can speak of a machine as moral without expecting to be contradicted, we will have succeeded in turning ethics into a strictly extrinsic, behavioral affair in which internals are irrelevant.

Since on the surface, an ethics without an *ought* is as empty as *thinking* without *insight* or *wisdom*, it is necessary to explore what else *ought* implies in order to form an adequate conception of a metaphysics of morals that will fit the information age. While other research for a working conception of ethics has already been done (e.g., Floridi and Sanders, 2004), a careful exploration of this foundational concept still appears lacking. I hope to fill this gap to explore whether ethics can get by without its cherished *ought* and, if so, what that implies for ethics more generally. The concern guiding this talk is

whether the information age is issuing in a post-ethical age or whether it is leading to a redefinition of ethics that is both long overdue and needed.

References

- Anderson, M., & Anderson, S. (2007). Machine ethics: Creating an ethical intelligent agent. *AI Magazine*, 28(4): 15-26.
- Beavers, A. (2011). Moral machines and the threat of ethical nihilism. In P. Lin, G. Bekey & K. Abney (Eds.), *Robot ethics: The ethical and social implication of robotics*. Cambridge, MA: MIT Press, forthcoming.
- Beavers, A. (2009, March). Between angels and animals: The question of robot ethics, or is Kantian moral agency desirable. The Eighteenth Annual Meeting of the Association for Practical and Professional Ethics, Cincinnati, Ohio.
- Dennett, D. (2006, May). Computers as prostheses for the imagination. The International Computers and Philosophy Conference, Laval, France.
- Floridi, L., & Sanders, J. (2004). On the morality of artificial agents. *Minds and Machines* 14(3): 349-379.
- Turing, A. (1950). Computing machinery and intelligence. *Mind* 59: 433-460.