

The International Digital Ethics Symposium,  
Center for Digital Ethics and Policy, School of Communication

## ***Could and Should the Ought Disappear from Ethics?***



The Creation of Adam – Michelangelo – 1511

*Loyola University, October 28<sup>th</sup>, 2011*

# ***Could and Should the Ought Disappear from Ethics?***

## Paper Overview by Section



Adam and Eve – Dürer - 1504

A Provocation

Terra Firma

Moral Agency

What Ought Implies

The Sufficiency Argument

Conclusions

A Palinode

# ***Could and Should the Ought Disappear from Ethics?***

## A Provocation



Adam and Eve – Masaccio – 1426

“Everyone will readily agree that it is of the highest importance to know whether we are not duped by morality.”

- *Emmanuel Levinas, 1961/1969, 21*

# ***Could and Should the Ought Disappear from Ethics?***

## A Provocation



Cain – Vidal – 1896

“The problem that concerns us in this conference —that of the community is, without doubt, a topical one, due to the unease felt by man today within a society whose boundaries have become, in a sense, planetary: a society in which, due to the ease of modern communications and transport, and the worldwide scale of its industrial economy, each person feels simultaneously that he is related to humanity as a whole, and equally that he is alone and lost.”

- *Emmanuel Levinas, 1982/1989, 212*



UNIVERSITY  
OF  
EVANSVILLE  
Civic Mission...Sacred Trust



## ***Could and Should the Ought Disappear from Ethics?***

### Terra Firma



Noah and His Ark – Peale (after Catton) – 1819

“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.”

- *Anderson and Anderson, 2007, 16*

Computational ethics makes ethics honest.

# ***Could and Should the Ought Disappear from Ethics?***

## Moral Agency



The Tower of Babel – Pieter Bruegel the Elder – 1563

**{def MA}** any agent that does the right thing morally, however determined.

**{def RMA}** an MA that is fully responsible and accountable for its actions. It can decide things for itself and also has the capabilities for full moral agency as defined by Moor (2006).

**{def AMA}** a manufactured MA that may or may not be an RMA.

# ***Could and Should the Ought Disappear from Ethics?***

## Moral Agency



Sodom and Gomorrah – Jan Brueghel the Elder – c1610

This taxonomy allows us to ask two pressing questions:

- 1) Could an AMA be an RMA?
- 2) Should an AMA be an RMA?

## ***Could* and *Should* the *Ought* Disappear from Ethics?**

### What Ought Implies



The Sacrifice of Isaac – Caravaggio – 1604

**{Def TO}** if agent X seeks a goal state Y and Z is necessary for Y, then X ought to Z to get Y.

**{Def MO}** agent X is inclined to seek a goal state Y, but ought not because the moral law prohibits X from pursuing Y.

*Ought* implies *can*, but it also implies *might not*.

## ***Could and Should the Ought Disappear from Ethics?***

### The Paradox of Automated Moral Agency



- 1) If we are to build autonomous machines, we have a *prima facie* moral obligation to (try to) make them RMAs.
- 2) For an RMA to be responsible and able to be held responsible for its actions, it must be capable of both succeeding and failing in its moral obligations.
- 3) An AMA that is also an RMA must therefore be *designed* to be capable of both succeeding and failing.
- 4) It would be a moral failure to unleash upon the world machines that are capable of failing in their moral obligations.
- 5) Therefore, we have a moral obligation to build AMAs that are not also RMAs.



# ***Could and Should the Ought Disappear from Ethics?***

## The Sufficiency Argument



Jacob Wrestling with the Angel – Leloir – 1865

The kind of moral interiority necessary for an agent to be an RMA is a sufficient though not necessary condition for being an MA.

Therefore, moral interiority is not essential for moral agency.

To accept the argument, we must acknowledge that an MA is a genuine moral agent, whether it be an AMA, an RMA or both. There are at least three ways to do so.

# ***Could and Should the Ought Disappear from Ethics?***

## The Sufficiency Argument



Adoration of the Golden Calf – Tintoretto – 1564

- 1) Moral Functionalism
- 2) Moral Behaviorism
- 3) The Moral Turing Test

# ***Could and Should the Ought Disappear from Ethics?***

## Conclusions



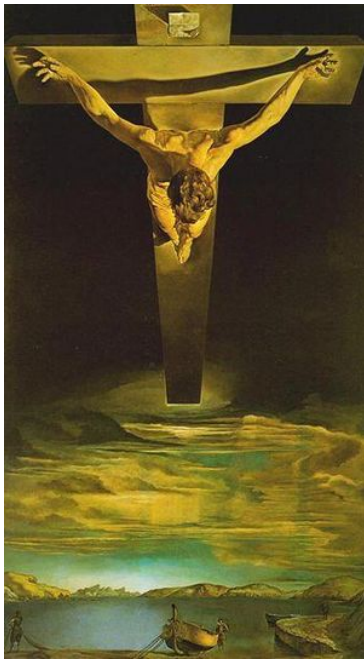
Let us then postulate the existence of such a machine, call it MorMach. Its sole job is to calculate a moral course of action using some decision procedure or an amalgamation of several to determine what an individual should do in any given situation. Suppose that the size and scope of MorMach's dataset is such that it sees all and knows all (or almost all) and that after several succeeding generations, MorMach becomes the sole authority, the ultimate oracle, if you will, in all matters of morals. What then?

Image from <http://singularityhub.com/wp-content/uploads/2011/05/Moses-Robot.jpg>



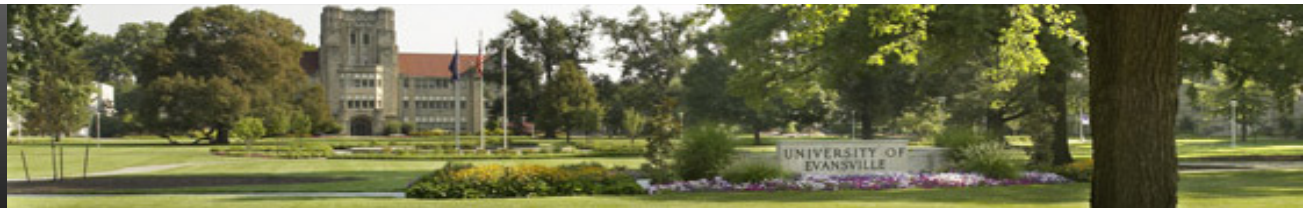
# ***Could and Should the Ought Disappear from Ethics?***

## A Palinode



Christ of St. John of the Cross  
Dali – 1951

I once knew a seminarian who, after spending two years working in a soup kitchen, decided that he should quit his work because his motives were not pure. “How selfish of me,” he asked, “to use people to satisfy my need for moral cleanliness?” How selfish of him, we might ask, to put his moral purity ahead of others who so desperately needed his response. Oddly, our seminarian was not too far from Kant, who once argued in favor of telling the truth to a murderer regarding the whereabouts of his intended victim. A deceit, he noted, could make one an incidental accomplice, whereas the truth could never permit so much (1797/1993). In both cases the need for preserving the moral purity of the subject is at issue. How unfortunate for the world ... and for ethics! Sometimes it’s best to feed the hungry for the wrong reasons rather than not at all. Sometimes it may be best to deceive.



# Acknowledgements

This presentation is based on a paper, “Could and Should the Ought Disappear from Ethics?”, written for the International Symposium on Digital Ethics, Center for Digital Ethics and Policy, School of Communication, Loyola University Chicago, October 28<sup>th</sup>, 2011. It follows in part from a presidential address (Is Ethics Computable, Or What Other than *Can* Does *Ought* Imply?) presented at the 2011 annual meeting of the International Association for Computing and Philosophy, held at Aarhus University in Denmark, July 4th-6th, and in part from a conference presentation (Between Angels and Animals: The Question of Robot Ethics, or is Kantian Moral Agency Desirable?) presented at the 18th annual meeting of the Association for Practical and Professional Ethics, held in Cincinnati, Ohio, March 5th-8th, 2009, though it has other predecessors as well. In this regard, I wish to thank the following people for their comments and conversations that led up to this paper: Colin Allen, Susan Anderson, Selmer Bringsjord, Larry Colter, Richard Connolly, Charles Ess, Luciano Floridi, Christopher Harrison, Deborah Johnson, Patrick Lin, Dianne Oliver, James Moor, Matthias Scheutz and Wendell Wallach.



## References

- Allen, C., Varner, G., and Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental and Theoretical Artificial Intelligence* 12: 251-261.
- Anderson, M., and Anderson, S. (2007). Machine ethics: Creating an ethical intelligent agent. *AI Magazine* 28(4): 15-26.
- Beavers, A. (2001). Kant and the problem of ethical metaphysics. In M. New, R. Bernasconi and R. Cohen (Eds.), *In proximity: Emmanuel Levinas and the Eighteenth Century* (pp. 285-302). Lubbock, TX. Texas Tech University Press. Reprinted in *Philosophy in the Contemporary World* 7(2): 47-56.
- Beavers, A. (2009, March). Between angels and animals: The question of robot ethics, or is Kantian moral agency desirable? Association for Practical and Professional Ethics, Eighteenth Annual Meeting, Cincinnati, Ohio.
- Beavers, A. (2010). Editorial to *robot ethics and human ethics*. Special issue of *Ethics and Information Technology* 12(3): 207-208.
- Beavers, A. (2011, in press). Moral machines and the threat of ethical nihilism. In P. Lin, G. Bekey and K. Abney (Eds.), *Robot ethics: The ethical and social implication of robotics*. Cambridge, MA: MIT Press.
- Connolly, R. (2011). Are corporations moral agents? Andiron Lecture. The University of Evansville, September 14th.
- Dennett, D. (2006, May). Computers as prostheses for the imagination. The International Computers and Philosophy Conference, Laval, France.
- Floridi, L. (1999). Information ethics: On the philosophical foundation of computer ethics. *Ethics and Information Technology* 1: 37-56.
- Floridi, L. (2002). On the intrinsic value of information objects and the infosphere. *Ethics and Information Technology* 4: 287-304.
- Floridi, L., and Sanders, J. (2001). Artificial evil and the foundations of computer ethics. *Ethics and Information Technology* 3: 55-66.
- Floridi, L., and Sanders, J. (2004). On the morality of artificial agents. *Minds and Machines* 14(3): 349-379.
- Kant, I. (1785/1993). *Grounding for the metaphysics of morals*. J. Ellington (Ed. and Trans.). Indianapolis: Hackett.



## References

- Kant, I. (1797/1993). On a supposed right to lie because of philanthropic concerns. In J. Ellington (Ed. and Trans.), *Immanuel Kant: Grounding for the metaphysics of morals with on a supposed right to lie because of philanthropic concerns* (pp. 63-68). Indianapolis: Hackett.
- Kurzweil, R. (1990). *The age of intelligent machines*. Cambridge, Mass: MIT Press.
- Levinas, E. (1961/1969). *Totality and infinity: An essay on exteriority*. A. Lingis (Trans.). Pittsburgh, PA: Duquesne University Press.
- Levinas, E. (1977/1990). And God created women. In A. Aronowicz (Trans.), *Nine Talmudic readings* (pp. 161-177). Bloomington, IN: Indiana University Press.
- Levinas, E. (1982/1989). The pact. In S. Hand (Ed.). *The Levinas reader* (pp. 211-226). Cambridge, MA: Blackwell.
- Moor, J. (2006). The nature, importance and difficulty of machine ethics. *IEEE Intelligent Systems* 1541-1672: 18-21.
- Royakkers, L., and van Est, R. (2010). The cubicle warrior: The marionette of digitized warfare. *Ethics and Information Technology* 12(3): 289-296.
- Turing, A. (1950). Computing machinery and intelligence. *Mind* 59: 433-460.

