Wired for Warmth:
Robotics as Moral Philosophy

Alan E. Singer, Appalachian State University, USA

ABSTRACT

An aspect of the relationship between philosophy and computer engineering is considered, with particular emphasis upon the design of artificial moral agents. Top-down vs. bottom-up approaches to ethical behavior are discussed, followed by an overview of some of the ways in which traditional ethics has informed robotics. Two macro-trends are then identified, one involving the evolution of moral consciousness in man and machine, the other involving the fading away of the boundary between the real and the virtual.

Keywords: *Artificial Moral Agents, Computer Engineering, Ethical Behavior, Morality, Philosophy, Robotics*

1. INTRODUCTION

Traditionally, engineering and philosophy have been regarded as completely separate academic "disciplines" and duly accommodated in different academic faculties. However, in the last few decades it has become commonplace to advocate interdisciplinary studies and to point to ways in which subfields like computer-engineering, control-theory and robotics have the potential to inform fundamental philosophical questions (e.g., Boden, 2006; Wallach & Allen, 2009). Accordingly, the present paper discusses how robotics can inform ethics or moral philosophy and *vice versa* (i.e., the notion that robots can be ethical, under various meanings of the latter). The discussion also refers to the related notion of virtual worlds and the increasingly problematic distinction between the virtual and the real.

DOI: 10.4018/ijsodit.2012070102

The paper starts with a brief discussion of the idea that robotic informs ethics because engineers are "doing" philosophy. Building mainly upon the work of Wallach and Allen (2009), it offers an account of some specific ways in which the design of artificial moral agents (AMA's) has informed philosophy. Then, several areas where AMA design and moral philosophy seem to parallel each other are considered (Section 4) followed by an overview of some of the ways in which moral philosophy has informed robotics. Two long term trends are then identified (Section 6) one involving the evolution of conscience in man and machine, the other involving the fading of the boundary between the real and the virtual.

2. THE ENGINEERING OF PHILOSOPHY

According to Dennett (1997) "you don't really know how something works if you can't build

it," so that "robotocists are doing philosophy, whether or not they think this is so." A decade later this seems increasingly to be the case. Yet this is a distinctive "experimental and constructive computational philosophy" (Wallach & Allen, 2009), or a "philosophy plugged" (e.g., Singer, 2010) that also fits well with some independently developed epistemological and ontological notions such as:

i. Knowledge as coordination-of-action (Zeleny, 2005);
ii. Information as "*in*-formation"; that is, codes co-creating physical form as in robotic manufacturing contexts (Zeleny, 2005); and
iii. The convergence and unity of the physical and mental worlds, somewhat in line with Spinoza's 17th century writings, discussed subsequently.

The task of constructing AMA's has repeatedly spun-off sharply-framed questions that are both philosophical and technological in nature, but that also carry significant implications for policy (i.e., *macro*-ethics, to use terminology from business-ethics). Indeed, when ethics is plugged-in, so to speak, it looks and feels quite different from the penned works of Kant, Mill, Bentham, or the Bible. In part this is due to the fact that, as correctly predicted by Alvin Toffler (e.g., Toffler & Toffler, 1990) the development of robots and AMA's is almost entirely a project of the military-industrial complex, being "done" outside the public gaze and far away from the desk of the traditional philosopher. For example, one military project involves installing (or instilling) a "functional morality" into a robot machine gun. The design-objective in this case was to re-program the robot guns with a form of ethics so they would stop killing friendlies or "innocent" civilians and concentrate all their firepower on the bad guys[1]. Eventually, as Singer has noted, AMA's "might be endowed with a conscience that would…make them more humane (as) soldiers than humans" (2009, p. 425).

3. HOW ROBOTICS INFORMS ETHICS

An overview of the wider emerging literature on ethical robots and artificial general intelligence (e.g., Goertzel, 2002 Wallach & Allen, 2009) suggests that there are at least four areas in which computer engineering has already substantially informed moral philosophy (Figure 1). These are with respect to: ethical-incrementalism (i.e., achieving ethical outcomes by means of frequent small steps, rather than occasional major decisions), evolutionary ethics, the notion of the "difficulty" of moral values and the vexed question of moral-agency, as follows:

3.1. Ethical Incremental-ism & LIDA

According to information readily available on the web, the Learning Intelligent Distribution Agent (LIDA) is an autonomous general intelligent system (AGI) built by the US Navy to make human resource related decisions. Here, ethical decision making (EDM) is reduced to a series of selections of internal and external micro-actions, rather than one-off choices

Figure 1. Areas where AMA design strongly informs philosophy

amongst given projects or courses of action. Inside LIDA, lines of software known as codelets scan a virtual workspace in which all inputs are represented. As noted by Snow (2009), these codelets are quite similar to the "demons" in a 1970's cognitive model called Pandemonium, but also the "agents" in Minsky's Society of Mind (Minsky, 1988). The codelets scan a virtual workspace (cf. Baars, 1997) for informational inputs that should be brought to the attention of the wider system, or brain. In a competition for attention that lasts about 0.1 second, a "winning" piece of information emerges, which is then broadcast throughout the system.

The next step within each EDM cycle is to (i) act, or (ii) reflect more, or (iii) add something to a mental model that is always under construction in semantic memory. So, even though LIDA does not execute any programs of top-down moral reasoning (e.g., utilitarian cost-benefit analysis) it detects "morally-relevant inputs" and acts on them.

The task of programming a machine to identify moral relevance within a general perceptual space, or some constructed view of the world, now represents an engineering challenge as well as an opportunity to inform traditional moral philosophy. An example of such a challenge is the recent MITI ruling that robots (and no doubt the driverless cars of the future) "must have sensors to prevent collision with humans" (Singer, 2009, p. 423). This is a specific example of a general moral imperative, identified much earlier by a British philosopher, Iris Murdoch, to "gaze" at the world before acting ethically within it. Another moral philosopher, David Hume, also stressed the need to understand all the relevant (available) facts before striving to consider them from a "general" point of view, whilst Amory Lovins (an ecologist) also claimed more recently that "the single most important thing … is to pay attention" (i.e., to all aspects of the relevant situation). In sum, the notion of identifying and attending to *morally-relevant inputs*, as already expressed by LIDA and MITI, appears to merit more sustained attention from anyone involved in applied ethics.

3.2. Evolutionary Ethics & Robot Competitions

Evolutionary ethics includes any account of moral behavior (e.g., benevolence, integrity, restraint, etc.) that involves adaptation and the fitness of various entities. In Robert Axelrod's path-breaking "evolution of cooperation" work (e.g., Axelrod, 1984), computer programs that played the iterated Prisoner's Dilemma Game were pitted against each other, with winning programs duly selected for the next round (or generation) in a primitive virtual world. This was the first demonstration of how computational power could go beyond intuition and "penned" mathematics; not only in game theory, but also in philosophy and evolutionary biology. More recently, it has been has pointed out (e.g., Hall-Stores, 2007) that a similar approach might be extended to sets of physical robots, each programmed with a variety of moral behaviors. The extent to which any given robot then obeys specified moral laws (such as Asimov's laws of Robotics, or Kant's categorical imperative) might then be used as a fitness criterion in a new kind of evolutionary competition. Such experiments with groups of differently-programmed robots have the potential to put flesh (or nuts) on the concept of "survival of the most moral" as well as the sustainability of particular moral rules. This, in turn, might help humans to better understand the effect of specific ethical behaviors on their own survival and sustainability.

3.3. Collective and Artificial Moral Agency

There has been a philosophical debate for at least the last forty years about the notions of corporate and collective moral agency (CMA). In philosophy, the claim that only individuals can be moral-agents (e.g., Friedman 1970) has confronted numerous arguments to the contrary (e.g., Danley, 1984; French 1984; Gilbert, 1986; Singer, 1994) together with recent steps towards a pragmatic resolution (e.g., Buchholz & Rosenthal, 2006). In law, the agency-principle holds a corporation vicariously liable for the

acts of various individuals; the identification-principle holds that a layer of senior officers is the (responsible) mind or brain of the firm, and the systems-principle holds that the existence of an internal decision making structure is deemed sufficient to confer corporate liability. More generally, the concept of the corporation or any collectivity as a moral agent is supported by an almost limitless supply of metaphors between individual and corporate/collective behavior and cognition (e.g., internal vs. external analysis in business strategy is like a reflective individual looking inward and outwards, etc.). The AMA project has the already informed this debate and it has the potential to transform it. For example, even in the present article so far, several new metaphors have arisen, such as:

i. The idea of survival as an "easy" value, which seems also to apply to corporations.
ii. The idea of ethical-incrementalism, akin to logical-incrementalism in organizational behavior and strategy.
iii. The surprisingly high performance of insect-like robots in which each insect-"leg" takes its cue mainly from its other legs, which inspired a push for "subsumptive organizational architectures" whereby department heads talk directly to each other, with no headquarters.
iv. The general principle that autonomy precedes "sensitivity" in the development of AMA's (e.g., the "ethical" robot gun) which reinforces the notion that consciousness develops into conscience.
v. The trend from artificial intelligence towards artificial morality which corresponds to the viewpoint that business strategy is becoming more ethical (e.g., emerging standards for corporate social responsibility, etc.) and the wider notion of human moral progress.

The potential for a transformational impact of the AMA project on the CMA debate was expressed recently by Hall-Stores (2007, p. 313), as follows:

"can we say anything about the rights and duties of corporations if…the AI's (artificial intelligences) probably will be running them within the next few decades?"

Taking that challenge at face value now seems to be a serious philosophical project. As a preliminary observation, one might first note the potential advantages of having AI's "running" corporations, as these could be more careful (or less negligent) that human managers. More abstractly, the philosophical project of linking the CMA and AMA debates can be briefly summarized as follows: (i) there exists a set of *pro*-CMA and a set of *anti*-CMA arguments, with another distinctive but overlapping set of *pro*-AMA and *anti*- AMA arguments. (ii) These have the potential to mutually refine and augment each other. Yet as the quote from Hall points out: (iii) the two sets of arguments become identified directly with each other when one introduce the notion of the virtual firm, or a corporation run (or owned) by robots (Figure 2 and Figure 3).

Computer-engineering now also the potential to transform the more fundamental idea of individual human moral agency (Figure 3) because the *pro* and *anti* AMA arguments have to take into account the fading boundaries between the real and the artificial (a trend discussed subsequently, in Section 6). For example, according to the "robust view of ethics" any moral agent must have "real" feelings or sentience, like an individual human being. It must experience *qualia,* the qualitative aspect of emotion. By that account, only individual humans (and perhaps some animals) can be "real" moral agents. Also, according to theological-ethics, human virtues include religious faith with a belief in the soul (of humans, at least). An AMA or robot can express these beliefs and act *as if* it holds them; it might also be sensitive to others who hold these beliefs; but the question remains as to whether it can actually *have* faith and a soul? This question remains important to many, even though trying to answer it may be a "Monkish pursuit" as Wallach and Allen put it (2009, p. 215).

Figure 2. Collective and artificial moral agency arguments inform each other

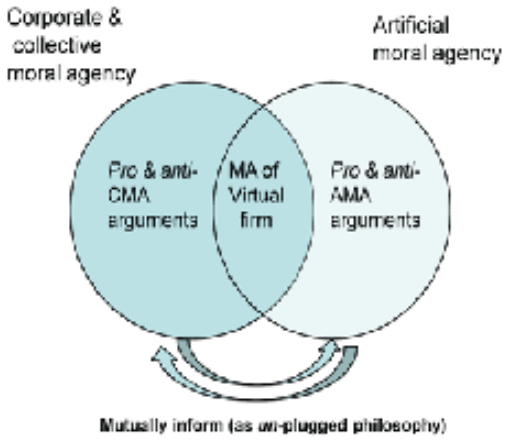
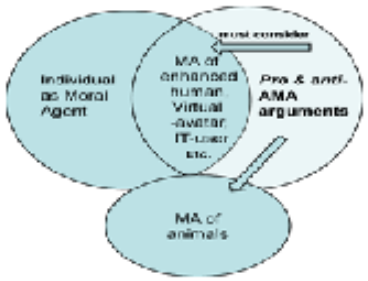


Figure 3. The effect of boundary-blurring on moral agency arguments



Perhaps a more urgent philosophical question involves the moral-agency status of hybrid types of "agent" or entity. What can be said about the moral agency of enhanced humans (or animals) with 'wet AI' brain implants; or creatures equipped with neuro-prosthetics that might also be controlled exogenously? What about the moral agency of human-like avatars in ever more sophisticated virtual worlds? In this context, Julian Savelescu (at the Uehiro Centre for Practical Ethics, Oxford) has argued that:

"if other beings possess rationality and the ability to cooperate and to empathise… then we should treat them no differently than other human beings." (as cited in Snow, 2009)

But this component of the moral agency debate remains controversial. There is less doubt, however, that the fading and blurring of boundaries between once-stable categories has complicated and added a new dimension the moral agency debate, even as it potentially sheds new light on some parts of it.

3.4. The Difficulty of Values

According to Goertzel (2002), any autonomous system (e.g., robot, individual or possibly a corporation) needs to have "survival" as a basic value, in line with the evolutionary ethics discussed earlier. He identified a set of "easy basic values" for autonomous systems. "Easy" in this context means "easy to program" into an AMA and so we can see from the outset that there is potential to inform traditional ideas about human value-priorities. The easy values are: (i) keep yourself healthy, (ii) preserve patterns that have been valuable, and (iii) create diversity. These are also obviously good for humans, although, significantly, the last

two differ from the classical Platonic human goods (i.e., health, friendship, justice, wealth and aesthetics). In contrast, Goertzel's "hard values," or harder-to-program values, include (i) preserving other life and (ii) making others happy. These are not so obvious, nor natural, and they normally have to be taught by a society or imposed by an authority (e.g., "Thou shall not kill"). The latter distinction in turn points to a wider arena in which engineering and philosophy have mutually informed each other, that is, with respect to the relationship between Top-down vs. Bottom-up ethical behavior.

4. TOP-DOWN & BOTTOM-UP ETHICS

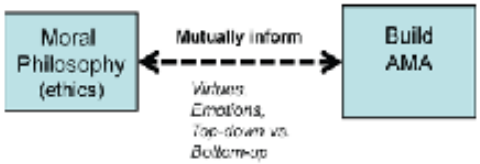
The question of top-down vs. bottom up influences on behavior and morality (e.g., following given rules vs. independent experiential learning) reveals a more mutually informative relationship between engineering and philosophy (Figure 4). For example, Genghis is an insect-like robot (also described in Wallach & Allen, 2009). It does not have much of a brain, but it certainly appears to know what it's doing. Each leg takes its cue from the other legs, with a few local features. Genghis' "knowledge" is thus fully expressed as coordination-of-action, as there is not much else. Yet, Genghis moves around better than its more brainy competitor bots. In some ways, practical ethics and moral judgment do seem to be more like Ghengis than Kant (or God). For example, socially adept responses and authentic social skills constitute an important part of ethics in applied social contexts. Perhaps, therefore, applied ethicists should also pay a more attention to these aspects of human behavior.

4.1. Virtues

Engineers have also been steered towards classical virtue ethics, where Plato pointed to secular virtues such as wisdom, courage, moderation and justice. Aristotle drew a distinction between intellectual virtues such as loyalty and moderation that can be taught (or programmed in a top-down sense) vs. virtues such as humor and politeness (i.e., the mentioned "social skills," typically acquired "bottom-up" through practice and habit). Wallach and Allen (2009) noted that a future AMA might be able to emulate some aspects of these virtues. Indeed, a robotic or virtual AMA has the potential to be more moral than humans in this respect, because human virtues are often merely apparent, or unstable, or temporary.

"Bottom up ethics" broadly encompasses neural networks, connectionist psychology and particularist ethics (e.g., Dancy, 1998) in which the focus is upon a moral agent learning how to articulate moral reasons for actions, properly rooted in any given context, episode, or narrative. Indeed, according to Wallach and Allen (2009, p. 130) any fully-functioning moral agent has to be able to "represent the reasons that might be applied" to justify a course of action. Bottom-up ethics also involves the above-mentioned "easy values," the learned social skills and some of the virtues. In contrast (Figure 5) top-down ethics involves the rules and guides found in the traditional ethical theories and more specifically in ethical principlism (e.g., Beauchamp & Childress, 1994). Here, an external social system or authority becomes the source of ethics, as well as the kinds of explanations and justifications produced by the AMA. In general, the building of AMA's seems to be

Figure 4. Areas where AMA design and moral philosophy inform each other



demonstrating that the relationship between top-down and bottom-up ethics is recursive and complementary, involving not only behavior in real-time, but also the meta-level (philosophical) understanding of the nature of that behavior.

4.2. Emotions

Many aspects of emotion (or emotional intelligence) are also programmable and capable of being learned by an AMA. These include (i) the ability to detect and respond intelligently and expressively to others' facial expressions or body posture, and (ii) interpreting other's intentions in context, or responding sympathetically and appropriately to others' predicaments[2]. A tougher challenge, also discussed by Wallach and Allen (2009), involves making a robot behave as if it was experiencing or anticipating its own quasi-emotions. For example, an emotionally-intelligent robot gun might avoid friendly fire if it were able to anticipate the pain this would cause to itself. This "pain" would only have to be some internal state with "valence" (a +/- parameter), such as opposing the robot's "easy" values, or interfering with its goal-attainment, or slowing it down. In Metzinger's Phenomenal-Self Model (Metzinger, 2004), an entity is able to "see" its own somatic (cellular, bodily) responses. Accordingly, if the peripheral components of a robot somehow responded directly to emotionally-relevant (and morally-relevant) inputs, it might be able to compute quasi-emotions and adjust its behavior accordingly. This would emulate the autonomous nervous system and kinaesthetic memory in animals and humans, but also fits well with LIDA and Ghengis.

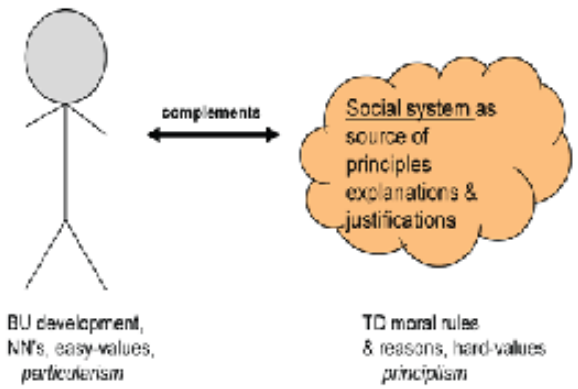
5. HOW ETHICS HAS INFORMED ROBOTICS

Several writers have claimed that that traditional moral philosophy has distracted engineers and programmers, or otherwise questioned its importance. For example, according to Wallach and Allen (2009, p. 214) "it is not possible to see a clear way to implement a (traditional) ethical theory as a computer-program" so "one might wonder whether these play a guiding role for human action." Nonetheless, some basic themes such as consequences, logic, rationality, virtue and emotion do appear to have guided the AMA project, broadly as follows.

5.1. Consequentialism

Since computer simulations enable better identification and forecasting of moral effects and consequences (e.g., traffic delays, pollution, etc.) it should be possible to build a powerful consequentialist AMA. This might be egoist (self-interested, perhaps with the "easy" values) or utilitarian (weighing the interests of all stakeholders). In either case, such an agent would be capable of a more informed moral judgment than an unaided human egoist or

Figure 5. Bottom-up development vs. top down moral rules



utilitarian, respectively. That is, it would be able to pass a moral Turing Test.

5.2. Deontology

Wallach and Allen also noted (2009, p. 95) that "a very powerful computer might be able to determine whether its current goal would be blocked if all other agents were to operate with the same motive or maxim." That is, it could execute a version of the Golden Rule (a kind of Kant-plugged). This also holds out the prospect of super-moral AMAs, because, as discussed in the previous section, humans have to be cajoled by authorities into following such rules, or extrinsically rewarded for following them. On the other hand, a Kantian AMA would immediately and permanently shut down that military bot-gun. Another Kant-inspired (logic-based) line of contemporary research involves the use of theorem-proving software to assess the adequacy of a block of software code for creating its intended outcome.

5.3. Contractarianism

A final area in which philosophy might guide AMA development involves contractarian moral-political theory (i.e., *macro*-ethics) (Figure 6). It might be possible to use virtual worlds (and eventually populations of robots) to simulate and test the kinds of social principles and policies that are associated with this philosophy. Contractarian theory (e.g., Rawls, 1972) holds that the core of ethics lies in agreements reached amongst free and independent persons. This might be updated to include general intelligences and "sentient beings." The "free" parties reflecting on social policies are then held to be in an "original position of equality," or under a "*Veil of ignorance*" about their actual identity in the society (Figure 7). That is, the designers of the social system are "ignorant" of their own position in it. Under such hypothetical conditions, according to Rawls, one can deduce the following principles:

i. Maximise liberty, provided that there is similar liberty for all, and
ii. Inequalities (e.g., In wealth, or the possession of other human goods) are ok, provided that they can be expected to work out to advantage of all.

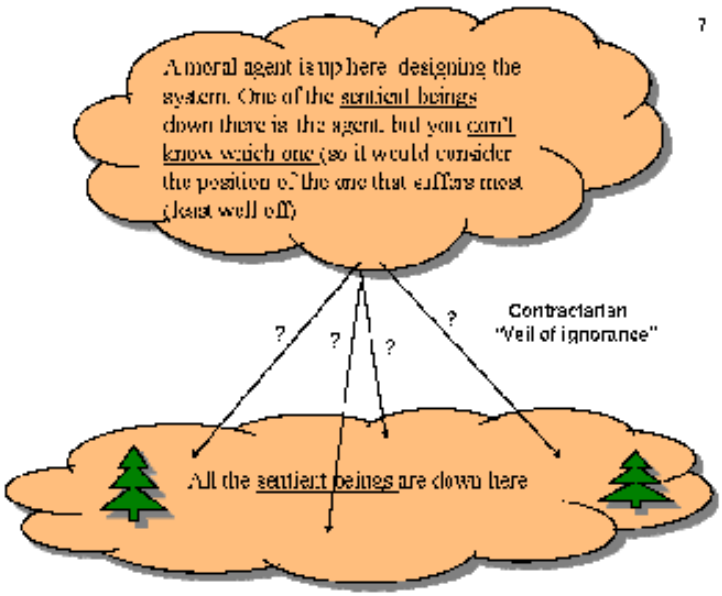
In future, virtual worlds might be deployed to refine and test these kinds of principles, as they might apply to diverse moral agents. For example, what balance of positive and negative freedoms works best with respect to the first principle? What types of behaviour or program at the micro-level (i.e., for each moral agent) would create a society in which that "expectation" of good has the highest chance of being realised? One can thus envision the potential for using virtual worlds or groups of robots to test and refine these kinds of principles as well as the social and legal policies derived from them.

6. MACRO-TRENDS

The discussion so far has indicated at least two macro-trends, each having substantial implications for both philosophy and robotics. The first is the idea that intelligence and rationality develops over time into "general intelligence" and ethics. The second involves the fading or blurring of the boundary between the real and the virtual (Figure 8). With regard to the "de-

Figure 6. Areas where philosophy informs AMA design

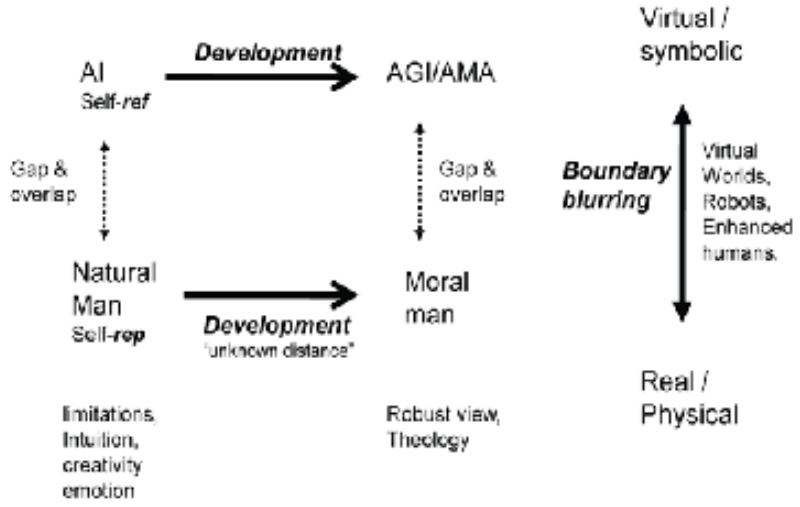
Figure 7. Testing contractarian macro-ethics in virtual worlds



velopment" of ethics, Kenneth Goodpaster (at Harvard Business School) remarked long ago (on video) that there was an observable "evolution of moral consciousness in the executive suite." Hunt and Mehta (2006) claimed more recently that "a new sense of responsibility is evolving" in connection with advances in *nano*-technology. Also on this optimistic side, Wallach and Allen (2009) have speculated about an *invisible hand of system interactions* whereby the operation of many self-sustaining easy-value holding AGIs/ AMA's might lead to the overall (macro-level) good, even if those agents individually lack the hard values, such as helping others.

All such speculations evoke the philosophical notion of moral progress, according to which "natural-man" (cf. Boden, 1987) de-

Figure 8. The development of conscience and the blurring of boundaries



velops into moral man, thus reducing what Engelberg once described as an "unknown distance" (cf. Singer, 1984). Rationality and artificial intelligence develop accordingly into a rational-morality and a general intelligence. At the early "rational/intelligent" stage, there is a gap separating AI from natural man due the cognitive limitations, creativity, intuitions and emotions of humans. At the more advanced AGI/moral man stage, there remains another gap due *inter alia* to the "robust" (humanistic) and theological views of what "ethics" means.

6.1. Boundaries

At the same time, the boundary between the symbolic or virtual worlds and the real or physical worlds is fading and potentially vanishing completely. Indeed, arguably the most profound way in which computer-engineering informs philosophy is by explicating (demonstrating, making real) the idea that the physical and mental worlds are ultimately the same thing, or the same "substance." They are undoubtedly becoming that way. This idea is not new, even though its widespread philosophical and public acceptance would be[3]. In the 17th century, Baruch Spinoza considered, in essence, that the physical and mental worlds were one and the same, a position that became generally known as neutral monism. According to the latter philosophy "there can exist certain substances (like persons) that are intrinsically neither material nor mental" (plato.stanford.edu/neutral monism). Throughout the industrial era, that idea was largely set aside in favor of the Cartesian separation of (physical) body and mind. That Cartesian position has since come under attack from many directions, not only from a re-evaluated monism, but also from an "immanent" classical pragmatism (e.g., Webb, 2007) and its associated ecological understandings.

7. CONCLUSION

In line with Dennett's claim, there do appear to be many ways in which the AMA project has informed moral philosophy, and *vice versa*. Some examples have been mentioned in this paper, where they have been classified according to the general direction of the "informing." The AMA project has particularly advanced notions of ethical-incremental-ism, evolutionary-ethics, moral-agency and the difficulty-of-values. In other areas, such as virtue ethics, emotions and top-down *vs*. bottom- up explanations, the contributions from philosophy and robotics appear to be more mutual. Finally, the traditional grand ethical theories have so far provided only a modest input to AMA development.

According to Boden (2006) "machines that compute and communicate…provide fruitful metaphors that help us to understand the mind." Several such "metaphors" have indeed been identified in this paper. However, in light of the blurring and fading of boundaries, Boden's reference to "us" merits further comment. The vexed question "*Who is 'us'?*" was first posed by Robert Reich, about 20 years ago, in connection with the paradoxes of international trade. It has great moral significance given the intuitively obvious notion that persons everywhere deserve to be treated ethically. The question has now taken on new meanings and importance, as hybrid AMA's are (recursively) being co-produced. These new types of agent are developing to the point where they become self-programming (i.e., the so-called "hard-takeoff point") which strongly suggests that problems involving identity and ethics are both likely to proliferate.

Although rapid (even exponential) moral progress involving super-moral machines remains a possibility; there is also the prospect of rapid moral-regression, because robot weapons (and viruses) might be re-programmed by malicious agents. On a more uplifting note, Wallach and Allen (2009, p. 215) also pointed out that "aircraft and birds fly in different ways." Accordingly, even if a morally-perfect AMA does "take-off" in future, so to speak, ordinary human morality might still remain a mystery to "us," as the robust and theological views can also persist and spread. One thing that we can be more confident about is that concepts such as "faith" and "soul" will become more

sharply defined and understood in the future, by all types of moral agent.

REFERENCES

Axelrod, R. (1984). *Evolution of cooperation*. New York, NY: Basic Books.

Baars, B. (1997). *In the theatre of consciousness: The workspace of the mind*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780195102659.001.1

Beauchamp, T. L., & Childress, J. (1994). *Principles of biomedical ethics* (4th ed.). New York, NY: Oxford University Press.

Boden, M. (1987). *Artificial intelligence and natural man*. New York, NY: Basic Books.

Boden, M. (2006). *Mind as machine*. Oxford, UK: Oxford University Press.

Buchholz, R. A., & Rosenthal, S. B. (2006). Integrating ethics all the way through: The issue of moral agency reconsidered. *Journal of Business Ethics*, *66*, 233–245. doi:10.1007/s10551-005-5588-9

Dancy, J. (1998). Can a particularist learn the difference between right and wrong? In *Proceedings of the 20th World Congress of Philosophy*, Boston, MA.

Danley, J. R. (1984). Corporate moral agency: The case for anthropological bigotry. In Hoffman, W. M., Frederick, R., & Schwartz, M. (Eds.), *Business ethics: Readings and cases* (pp. 172–179). New York, NY: McGraw-Hill.

Dennett, D. C. (1997). Cog as a thought experiment. *Robotics and Autonomous Systems*, *20*(2-4), 251–256. doi:10.1016/S0921-8890(97)80709-9

French, P. (1984). *Collective and corporate responsibility*. New York, NY: Columbia University Press.

Friedman, M. (1970, September 13). The social responsibility of business is to increase its profits. *The New York Times*.

Gilbert, D. R. (1986). Corporate strategy and ethics. *Journal of Business Ethics*, *5*, 137–150. doi:10.1007/BF00382755

Goertzel, B. (2002). *Thoughts on AI morality*. Dynamic Psychology.

Hall-Stores, J. (2007). *Beyond AI: Creating the conscience of the machine*. New York, NY: Prometheus.

Hunt, G., & Mehta, M. (Eds.). (2006). *Nanotechnology: Risk, ethics & law (Science in Society Series)*. London, UK: Earthscan.

Metzinger, T. (2004). *Being no-one: The self-model theory of subjectivity*. Cambridge, MA: MIT Press.

Minsky, M. (1988). *The society of mind*. New York, NY: Simon and Schuster.

Rawls, J. (1972). *A theory of justice*. Oxford, UK: Clarendon.

Singer, A. E. (1984). Planning, consciousness and conscience. *Journal of Business Ethics*, *3*, 113–117. doi:10.1007/BF02388812

Singer, A. E. (1994). Strategy as moral philosophy. *Strategic Management Journal*, *15*, 191–213. doi:10.1002/smj.4250150302

Singer, A. E. (2010). Philosophy plugged: How robotics informs ethics. *Human Systems Management*, *29*(1), 92–96.

Singer, P. W. (2009). *Wired for war: The robotics revolution and conflict in the 21st century*. New York, NY: Penguin.

Snow, P. (2009). Woe, Superman? *Oxford Today*, *22*(1), 13–15.

Toffler, A., & Toffler, H. (1990). *War and anti-war: Survival at the dawn of the 21st century*. New York, NY: Little Brown.

Wallach, W., & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. Oxford, UK: Oxford University Press.

Webb, J. L. (2007). Pragmatisms (plural) part 1: Classical pragmatism and some implications for empirical inquiry. *Journal of Economic Issues*, *41*(1), 1063–1086.

Zeleny, M. (2005). *Human systems management: Integrating knowledge management and systems*. Singapore: World Scientific. doi:10.1142/9789812703538

ENDNOTES

1. As one reviewer pointed out, it would presumably be ethical to "hack" a robot that was programmed to kill a human.
2. This notion of "appropriate response" can be fruitful. For example, in the movie *I-Robot*, a robot "saved the main character in a car that had

crashed, because his chances of survival were greater," instead of saving a little girl, "whereas a human would have saved the girl." Several principles and heuristics would support the Robot's choice. These can be compared with the flawed heuristics that humans often use when making decisions under pressure.

3. It is likely that AI will be as revolutionary as the internal combustion engine. Singer (2003, p. 430) observed that contemporary terminology such as "artificial intelligence" and "unmanned vehicle" are still calling things "by what they are not," just as cars were once called "horseless carriages"…a description would seem ridiculous today.

Alan E. Singer PhD is the James Holshouser Distinguished Professor of Ethics at Appalachian State University, North Carolina. Previously he was the John Aram Professor of Business Ethics at Gonzaga University, Spokane, Washington and Reader in Strategic Management at the University of Canterbury, New Zealand. He holds degrees in mathematics (Oxford), psychology (London), and Management (Canterbury, NZ). He is the author of Integrating Ethics & Strategy *(World Scientific) and* Strategy as Rationality *(Avebury series in philosophy), co-editor with Pat Werhane of* Business Ethics in Theory and Practice *(Kluwer) and editor of* Business Ethics & Strategy *(Ashgate series in public & private ethics). He has published widely in journal such as* Strategic Management Journal, Business Ethics Quarterly, Business Ethics: A European Review, Journal of Business Ethics, Omega, Human Systems Management, Systems Practice, Decision Sciences, Small Business Economics, *etc. He is the worldwide book review editor for* Human Systems Management *and a member of the ASU sustainability council.*