AUTONOMOUS WEAPONS AND THE NATURE OF LAW AND MORALITY: HOW RULE-OF-LAW-VALUES REQUIRE AUTOMATION OF THE RULE OF LAW

*Duncan MacIntosh**

I. INTRODUCTION

This article is a part of a larger project in which I try to see what can be said for the use of autonomous weapons systems (AWS) in conflicts with both State and non-State combatants.[1] In this article, I focus on the defensibility of delegating the formulation, administration, and enforcement of law—and by implication the law of war—to autonomously acting machines.

While AWS have obvious military advantages, there are prima facie moral objections to using them. In Part II of this article, however, I point out similarities between the structure of law and morality on the one hand and of automata on the other. I argue that these, plus the fact that automata can be designed to lack the biases and other failings of humans, require us to automate the formulation, administration, and enforcement of law as much as possible. Ethically speaking, deploying a robot is not much different from deploying a more or less well-armed, vulnerable, obedient, or morally discerning soldier into battle, a police officer on patrol, or a lawyer into a trial. All feature automaticity in the sense of deputation to an agent we do not then directly control. Such relations are well understood and well-regulated in morality and law; so there is not much challenging philosophically in having robots be some of these agents—excepting the implications of the limits of robot technology at a given time for responsible deputation.[2]

---

* For helpful discussion my thanks to fellow participants at a workshop on autonomous legal reasoning at the Institute for International Law and Public Policy at Temple University's Beasley School of Law, co-sponsored by the Washington Delegation of the International Committee of the Red Cross (ICRC), October 2015 (hereafter the ICRC conference). Earlier versions of this material were part of a larger paper mooted at a conference on the ethics of autonomous weapons systems held by the Center for Ethics and the Rule of Law (CERL) at the University of Pennsylvania Law School in November 2014 (hereafter the CERL conference), and I thank the other participants for useful feedback. My thanks also to the students in my classes and at lectures I gave at Dalhousie University, St. Mary's University, and at some high schools in Nova Scotia, as well as to Sheldon Wein, Greg Scherkoske, Darren Abramson, LW, and SD.

1. Two other papers in this project, both growing out of the paper for the CERL conference are: Duncan MacIntosh, *Fire and Forget: A Defense of the Use of Autonomous Weapons Systems in War and Peace* (Oct. 2015) (unpublished paper) (on file with author); and Duncan MacIntosh, *Autonomous Weapons and the Proper Character of War and Conflict* (Or: *Three Objections to Autonomous Weapons Mooted—They'll Destabilize Democracy, They'll Make Killing Too Easy, They'll Make War Fighting Unfair*) (Oct. 2015) (unpublished paper) (on file with author).

2. Jens David Ohlin, *The Combatant's Stance: Autonomous Weapons on the Battlefield*, 92 INT'L. LEGAL STUD. U.S. NAVAL WAR COLL. 1 (2016). Ohlin argues that law already has the means to deal with the question of who is responsible for the actions of autonomous weapons

In Part III, I consider this proposal in light of the differences between two conceptions of law. These are distinguished by whether each conception sees law as unambiguous rules inherently uncontroversial in each application; and I consider the prospects for robotizing law on each. Finally, in Part IV, I identify certain elements of law and morality, discussed by philosopher Immanuel Kant, which robots can participate in only upon being able to set ends and emotionally invest in their attainment.

II. SIMILARITIES BETWEEN MORALITY/LAW/MORAL AGENTS AND AUTOMATA; HOW RULE-OF-LAW-VALUES REQUIRE AUTOMATION OF THE RULE OF LAW

The debate around autonomous robot weapons is an occasion to reflect on the nature of human-discerned morality and human-created and administered law. And an awareness of the vast similarity between human and robot in this respect argues against objecting to robotic weapons for their somehow lacking a crucial feature of human morality. Not only would transferring adjudication and enforcement of law and morality to robots be consistent with human law and morality, but the highest ideals of both require this.

Foundational to my case is that, as legal and moral beings, much of the most important part of our lives is the attempt to discipline ourselves to rules.[3] These are things that could easily be the programming foundation for automated expert systems in law, policing, and morality.[4] Here, I go beyond those who think that human rights laws of engagement (international humanitarian law) can be programmed into robots, or that robots can be programmed to respect distinctness, necessity, and proportionality—that is, respectively, be programmed to protect civilians over soldiers, to only inflict harms when necessary to important military objectives, and to not let the weight of harms exceed the importance of the objectives. I say robots represent the possibility of the de-virtualization of the rule of law, the logical extreme of such rule. For robots are algorithm-driven; and so, in its fullest fruition, is law.

In fact, ironically, the hardest things to automate are the low-level perceptual skills we share with animals, things like situational awareness and identifying items in our environments.[5] The "higher" functions sustaining morality and law are easiest. Relatedly, it is relatively easy to program an expert system to make medical diagnoses and prescribe treatments once the system is given descriptions

---

systems, on the model of holding accountable the master-minds of systems—whether the systems are comprised of agents or are merely instruments—who use the systems in initiating harmful sequences of events. General Charles J. Dunlap, Jr. powerfully advances related points in a recent article. *See generally* Charles J. Dunlap, Jr., *Accountability and Autonomous Weapons: Much Ado About Nothing?*, 30 TEMPLE INT. & COMP. L. J. (forthcoming Spring 2016).

3. *See generally* JOHN STUART MILL, ON LIBERTY (1859).

4. Merel Noorman, *Computing and Moral Responsibility*, STAN. ENCYCLOPEDIA OF PHIL. (July 18, 2012), http://plato.stanford.edu/entries/computing-responsibility/.

5. *See* Eugene Wei, *Moravec's Paradox and Self-Driving Cars*, REMAINS OF THE DAY (Oct. 22, 2014), http://www.eugenewei.com/blog/2014/10/13/moravecs-paradox-and-self-driving-cars (discussing the difficulty of automating the low-level sensorimotor skills common to all animals).

of patient symptoms—think of WebMD.[6] But it is relatively hard to teach a machine to do such things as visually recognize signs of pain in a patient.[7] And it is easy to design an expert system to identify incorrect income tax payments once financial data are input to the machine—think of on-line tax calculation programs like Simple Tax.[8] But it is very hard to get a machine to recognize a signature on a check.[9] Of course even robotizing the higher functions is relatively hard.[10]

Fortunately, there are other features of morality that make it amenable to robotizing. One is that moral labour is divided, its several parts apportioned to those most competent at each respective part. Ordinary citizens are entitled and obliged to make decisions affecting their own welfare in ways not affecting each other.[11] Police must decide whom to arrest and how for prima facie and well-evidenced violations of the law.[12] Judges decide guilt and penalty.[13] Legislators decide laws.[14] Commanders-in-chief decide when and how to make war.[15] And there are similar stratifications in armed forces roles—the responsibilities of Privates are different from those of Corporals, and so on.[16] Thus, no one is obliged to master all of morality or law or the full ambit of combat. And so no robot in the automation of morality and law need be designed to do all of that either. This leaves only the easier task of robotizing any given part.

As a corollary, it is not required of any agent within a given part of our

---

6. WEBMD, http://www.webmd.com/ (last visited Apr. 15, 2016).

7. *See Affective Computing*, MIT, http://affect.media.mit.edu/ (last visited Apr. 15, 2016) (describing the field of creating computer algorithms that can interpret and respond to human emotional cues).

8. *See e.g.*, *Professional Tax Software*, Simple Tax Services, http://simpletax.co/ (last visited July 28, 2016).

9. *See* Lee Dye, *Computer Tests Prove Handwriting Analysis Legitimate*, ABC NEWS (June 5, 2002) (discussing how the U.S. Postal system uses automated handwriting analysis that is 75% accurate in order to route mail).

10. *See* Sebastian Anthony, *A new (computer) chess champion is crowned, and the continued demise of the human Grandmasters*, EXTREME TECH (Dec. 30, 2014), http://www.extremetech.com/extreme/196554-a-new-computer-chess-champion-is-crowned-and-the-continued-demise-of-human-grandmasters (discussing the history of programming computers to compete with humans at the game of chess).

11. *See* MILL, *supra* note 3, at Ch. IV (discussing the limits to the authority of society over the individual).

12. *See generally How to become a Police Officer,* CRIMINAL JUSTICE USA, http://www.criminaljusticeusa.com/police-officer/ (last visited Apr. 20, 2016).

13. *See generally* U.S. CONST. art. III.

14. *See generally* U.S. CONST. art. I.

15. *See generally* U.S. CONST. art. II.

16. *See Duties, Responsibilities and Authority explained*, ARMY STUDY GUIDE, http://www.armystudyguide.com/content/army_board_study_guide_topics/nco_duties/duties-responsibilities-authority-of-nco.shtml (last visited Apr. 11, 2016) ("A noncommissioned officer's duties are numerous and must be taken seriously . . . . Corporals and sergeants . . . must know and understand their soldiers well enough to train them as individuals and teams to operate proficiently.").

system of morality that it in effect pass a God level moral Turing test.[17] That is, it need not reliably be able to do things indistinguishable from what an omnisciently morally wise and impeccably self-conducting agent competent in any and every circumstance could do.[18] It is required at most only that the system of which the agent is a part, pass. Consider that element of our morality which is the Justice System. We hope that system will convict only the guilty; but we do not expect any given agent in the system to know how to perform every job that plays a role in bringing that about. Those agents who are expert at gathering evidence, say, do not also need to be expert at jury selection. This in turn means that what is required for a given agent to pass its part of the moral Turing Test—whether the agent is human or machine—is delimited by and specific to the agent's situation and particular role.

Moreover, we accept that both the whole system and the agents who are its parts are in some degree fallible. We aspire to have the system and the agents within it operate without error, but we do not trash the whole system when it occasionally makes mistakes, and we do not necessarily fire any given agent in the system for being imperfect. It would seem to follow, therefore, that it would be acceptable to have these functions performed by autonomous machines provided they could do the jobs of the agents within the moral system as well or better than fallible humans could, and provided the system of agents they would then collectively comprise could attain a similar standard.[19]

All of this connects with another feature of morality and law suiting them to robotizing: even now the institutions of law and morality are so stratified that those higher in the strata make decisions that those lower implement as if automatically or robotically relative to the higher.[20] Take law: legislators decide the law, judges

---

17. *See* Alex Hern, *What is the Turing test? And are we all doomed now?*, THE GUARDIAN (June 9, 2014), https://www.theguardian.com/technology/2014/jun/09/what-is-the-alan-turing-test ("Coined by computing pioneer Alan Turing in 1950, the Turing test was designed to be a rudimentary way of determining whether or not a computer counts as 'intelligent.' The test, as Turing designed it, is carried out as a sort of imitation game. On one side of a computer screen sits a human judge, whose job is to chat to some mysterious interlocutors on the other side. Most of those interlocutors will be humans; one will be a chatbot, created for the sole purpose of tricking the judge into thinking that it is the real human.").

18. *See* Aubrey de Grey, *When is a Minion not a Minion?*, EDGE, https://www.edge.org/responses/q2015 (last visited Apr. 11, 2016) ("Since objective moral judgements build on agreed norms, which themselves arise from inspection of what we would want for ourselves, it seems impossible even in principle to form such judgements concerning entities that differ far more from us than animals do from each other, so I say we should not put ourselves in the position of needing to try.").

19. *See* Jean-Baptise Jeangène Vilmer, *Terminator Ethics: Should We Ban "Killer Robots"?*, ETHICS & INT'L AFF.(Mar. 23, 2015), http://www.ethicsandinternationalaffairs.org/2015/terminator-ethics-ban-killer-robots/#fnref-9002-20 (asserting AWS should not be required to be infallible, but rather limitedly fallible or fallible to the extent that humans are; the system should be required to pass an adaptation of the Turing test in which a robot satisfies the legal and moral requirements—and can consequently be deployed—when it can be demonstrated that it can respect the laws of war *as well as or better* than a human in similar circumstances). Thanks also to Kate Gotziaman for discussion of these points.

20. *See* Craig T. Palmer and Kathryn Coe, *From Morality to Law: The Role of Kinship,*

may only interpret it, and police may only arrest people for prima facie violations of it.[21] The job of the police is relatively independent of that of legislators—the police may not legislate.[22] As for morality, we, in effect, task some members of our culture—for example, university philosophy professors—with reflection on higher-level moral problems.[23] These persons produce theories for discussion by the rest of us and do the heavy lifting of constructing and evaluating arguments for the theories.[24] An extreme version of this stratification is the decisional hierarchy of the Catholic Church: the Pope is presumed to be the moral expert and his flock follows his decisions about moral matters.[25]

It is often true of course that people playing moral and legal roles at lower levels will resist decisions made at higher levels. But they do so only in the way of data testing theory: police can say a law is not working; judges can contest whether a piece of legislation is constitutional. However, the whole system is so structured that persons at each level play limited roles—the police can agitate for changes to the law, but cannot change it themselves.[26]

A final point about what suits law to being robotized is that the purport of rule of law is to have people live by principles impartially to the benefit of all.[27] But the problem with having law human administered and enforced is that humans are inevitably partial, biased, weak willed, exhaustible, unable to fully work out all of the consequences of the principles they are to follow, and so on.[28] We have to take

---

Tradition and Politics, POL. & CULTURE (Apr. 29, 2010), https://politicsandculture.org/2010/04/29/from-morality-to-law-the-role-of-kinship-tradition-and-politics/ (explaining the development of law from morality and the outsized influence the more powerful members of society have on that development).

21. *See Separation of Powers – An Overview*, NCSL, http://www.ncsl.org/research/about-state-legislatures/separation-of-powers-an-overview.aspx (last visited Apr. 11, 2016) (discussing the individual powers of each branch of government and the impermissibility of one branch exercising the powers of another).

22. *See id.* ("The legislative branch is responsible for enacting the laws of the state . . . . The executive branch is responsible for implementing and administering the public policy enacted and funded by the legislative branch.").

23. *See, e.g.*, Dan. W. Brock, *Public Moral Discourse, in* SOCIETY'S CHOICES: SOCIAL AND ETHICAL DECISION MAKING IN BIOMEDICINE 215, 222 (R.E. Bulger, Bobby E. Meyer & H.V. Fineberg eds., 1995).

24. *See Becoming a Philosopher: Careers, Salary Info & Job Description*, LEARNINGPATH.ORG, http://learningpath.org/articles/Philosopher_Career_Info.html (last visited Apr. 11, 2016) (describing the roles and responsibilities of philosophers).

25. *See FAQs on Moral Authority of the Pope*, GLOBAL CATHOLIC CLIMATE MOVEMENT, http://catholicclimatemovement.global/faqs-on-moral-authority-of-the-pope/ (last visited Apr. 11, 2016) (asserting that dogma is an example of papal teaching that includes the most basic aspects of Christian morality).

26. *See* NCSL, *supra* note 21 (discussing the individual powers of each branch of government and the impermissibility of one branch exercising the powers of another).

27. *See generally What is the Rule of Law?*, A.B.A., https://www.americanbar.org/content/dam/aba/migrated/publiced/features/Part1DialogueROL.authcheckdam.pdf (last visited Apr. 11, 2016).

28. *See* Kevin Kelly, *Better Than Human: Why Robots Will – and Must – Take Our Jobs*,

people as we find them, biases and all. Robots, however, we could make into perfect administrators and enforcers of law, unbiased and tireless engines of legal purpose.[29] This is why so deploying them is the perfection of the rule of law and so required by rule of law values.

True, it is fallible people who will design autonomous legal engines. But the hope is that they will do the designing under conditions less likely to feature human failings. One can program at one's leisure, in abstraction from the felt urgencies of real situations, with fellow programmers (and citizens, legislators, judges, etc.) checking one's work, over-seers examining it for bias, interest groups with different perspectives being consulted, and so on. Hopefully, one then produces a device immune to human frailties when actually making moral and legal decisions under the pressure of real events.

III. TWO CONCEPTIONS OF LAW AND THE PROSPECTS FOR ROBOTS ON EACH

I began by suggesting that rule of law values require that we automate it. This is especially true in situations where the use of an AWS would be more in the spirit of the rule of law rather than the rule of a man, or of the rule of humans generally. Recall that delegating the operation of law to automata can result in administration of law in a way that is above human susceptibility to moral failings of specious partiality, impulse, exhaustion, weakness of will, and emotional over-reaction.[30] So the use of automata may be required by the very commitment to the rule of law. For example, suppose we are tempted to arm one side in a civil war. But we worry that, due to past religious conflicts, they would use our weapons not only to win the war, but also to take revenge against and oppress the other side. Then, we might instead think that using AWS programmed by us would be more legally just and temperate, spare more life, and have a greater chance of producing a lasting peace, for it would interrupt the violence cycle of endless revenge.

But my view may presuppose a false conception of what law is. There are at least two conceptions of law. On both it is the determining of what should be done in each situation that people face, whether what was done in a given situation was right, and, if not, what should be done in consequence. But one conception sees this as the unbiased and unexceptional literal application of unambiguous rules. Call this the rote conception. The other sees all this as occurring by argumentation based on proposing and interpreting inherently ambiguous rules and negotiating their inherently contestable applicability in given cases. Call this the deliberation conception.

The rote conception deploys the metaphor of law as pre-given rules of a machine. The deliberation conception sees the institution of law as a kind of ongoing debate about behaviour-regulating norms that is conducted in terms of discussion about human-chosen principles, rather than discussion in terms of, say,

---

WIRED (Dec. 24, 2012, 6:30 AM), http://www.wired.com/2012/12/ff-robots-will-take-our-jobs/ (describing the ways in which robots are superior to human beings due to technological capabilities that surpass human ability and fallibility).

29. *Id.*

30. *Id.*

character traits, best outcomes, divine commands, or the impulses of a powerful person. We ask questions like: What ought to be the rules? What conduct is consistent with these rules? Which sub-rules would express the spirit of previously enacted rules? Which decisions about a given case would be consistent with more detailed rules themselves consistent with previous more general rules? And so on. In the end it may look like the following of rules. But in fact, at every stage something more complicated is in play, something requiring judgments themselves not obviously understandable as algorithm-driven. The whole process is more improvised than rule-guided. Indeed, on this second view, the behavior to be regulated could probably instead have been improvised without need of prior meditation upon principles. Invoking laws is just a handy conversational trope. We could instead have asked, for example, what would a virtuous person do? And the fact that the whole thing could have been done without reference to any laws or principles implies that it is not inherently a matter of rule-following at all.

Obviously the rote conception of what law is seems to have law as something more fully automatable.[31] But even on the deliberation conception some aspects of law—those that would still verge on being exhausted by rules of an unambiguous sort—might be robotized, e.g., parking law enforcement. Or, for example, income tax collection—most people's tax forms are never seen by a human reviewer, only by a computer that decides according to its program the appropriateness of the tax paid given declarations and independent reports of income.[32] And even on the rote conception of what law is, namely, clear rules that are supposed to be followed, a robot advanced enough to deliberate about and justify its actions just as a person can might be needed in order to do something else that is key to the very concept of law, namely, decide whether to comply with a given law, just as people must decide. Knowing what law is means in part knowing that sometimes you should break it. True, the breaking of some laws is itself something whose permissibility is sometimes mandated by other laws.[33] But it is doubtful whether, for any circumstance in which some law should be broken, there is already some other law that requires or permits this. The decision whether to be legal is not and should not always be a legal question. And robotic "blind" obedience to law no matter what would show a failing to fully understand what law is, and so would be a failure of the enterprise of robotizing the institution of law. It would appear, then, that to

31. Kate Gotziaman argues that a machine operating the law on the first conception of law would not really be autonomous, only automated. Kate Gotziaman, *Automated and Autonomous Robots: Drawing the Distinction* (2015) (unpublished student paper, Dalhousie University) (on file with author). I am attracted to her suggestion, although I am not sure how much the distinction matters at this point in my argument. It will matter later, but at that point, I think my position is consistent with Gotziaman's. It is contested what the distinction in question amounts to, but one way of drawing it might be this: behaving automatically is doing as one is told or acting according to a rule, while behaving autonomously is deciding for one's self what to do, perhaps by writing a rule for one's self and then following it. *Id.*

32. *See The Examination (Audit) Process*, INTERNAL REVENUE SERV. (Jan. 10, 2006), https://www.irs.gov/uac/The-Examination-(Audit)-Process (noting returns are selected for audit based on scores calculated by a computer).

33. *See, e.g.*, Ray v. Wal-Mart Stores, Inc., 359 P.3d 614 (Utah 2015).

make a machine able to fully participate in the legal realm we would also have to make it capable of participating in the moral realm, a realm that considers even more things than does the law in seeking to answer the question what ought to be done.[34]

The difficulty of producing a robot capable of operating within either conception of law is vastly overestimated. Suppose we try to produce a machine that can substitute for a human judge in a court of law. Suppose the machine we produce can only do things like follow a pre-given algorithm for co-varying mercy in sentencing with a degree of remorse; and suppose it can only do that if the degree of remorse is discovered by a human interviewer and given to the machine on a rating scale of one to ten. Maybe the machine is programmed to prescribe light sentences given high remorse, heavier sentences for medium remorse, and so on. On first hearing we might think that the program is too crude. But on reflection we might come to think that, because there is a lot of latitude about what morality and law require, this is actually good enough—good enough that we would be comfortable operating under this standard. In other words, maybe we learn what level of detail our legal and moral systems have and what an acceptable version of those systems is, from what we could make a machine follow. Perhaps our conceptions of morality and law should evolve with our attempt to reproduce them in machines.

We learn a similar lesson when we see how difficult it can be to educate people from different moral cultures into our culture, the one we are tempted to think of as uniquely right. The difficulty of such a project can make us re-think our expectations of what is reasonable to require of people. Should everyone live by Christian morality, for example? It is easy to think so until one encounters other cultures with different and yet still defensible conceptions of spiritual thriving. We also see something similar in our own personal lives: as each of us acquires more life experience we realize just how messy and varied living a moral life is and how much tolerance and forgiveness it requires—think of the norms differing from person to person around homosexuality, or monogamy.

But what are the prospects for machines being able to disambiguate inherently ambiguous rules, or negotiate their proper application in a given case? Lawyers will know about this more than I do, but I gather that in dealing in inherently ambiguous rules, and arguing for one disambiguation or another, the law is all about interpretation on the facts given book law and case law.[35] A reading of a

34. Relatedly, it is sometimes thought that one can make people in a given profession certain to behave in morally correct ways simply by having them internalize the code of ethics specific to that profession—codes of business ethics, engineering ethics, military ethics, medical ethics, and so on. But this fails to recognize that for any given code there can be conditions where morality might require its violation. Thus moral expertise involves more than just code mastery. Likewise, then, for legal expertise. *See* Duncan MacIntosh, *The Sniper and the Psychopath: A Parable in Defense of the Weapons Industry*, UNIV. OF PA. (Apr. 9, 2015), https://www.law.upenn.edu/live/files/4391-thesniperandthepsychopathpapermacintoshpdf (discussing different codes to regulate the weapons industry).

35. I am much indebted to the lawyer, Christian Weisenburger, for conversation on the ideas in this paragraph.

given case will strike a judge as right. Then she will justify the reading by citing elements of book and case law. For each legal question there are many interpretations one could impose. And book and case law can be used to support any of these interpretations. (If they cannot be so used, it is definitely a wrong interpretation.) Is there a uniquely right understanding for each question? Arguably not. Instead there are many plausible interpretations.

Suppose a given machine is programmed to randomly pick among the interpretations presented to it by the different parties in a given legal case, then to seek utterances from digitized case law that entail the verdict entailed by the interpretation in question, and then present these utterances as premises in a syllogistic argument whose conclusion is the verdict. Much of this is already technically possible since computers can now understand syntax and semantics, and have long been able to detect and produce syllogistic reasoning.[36] The machine keeps working through interpretations until it finds one able to be syllogistically so defended, and offers it as the verdict. (And let us assume that, often, the machine could have found more than one defensible interpretation if it had kept going.) If we did not know that a given verdict had occurred by a machine "flipping a coin" amongst the possibilities and then compiling an argument from book and case law for it, we might have found the result perfectly legally plausible, and might have found it an example of an entirely livable and prima facie just system of law. So, it is easier than one might have thought to have machine-run law, because the level of resolution in relation to some imagined truth that the law operates at is coarser than one might have thought. Machines can produce good enough law, in part because good enough is all that law ever is.[37]

It may seem that moral theorizing cannot be automated. However, one reason people doubt that we could make fully moral robots is that, even if we feel we have a kind of intuitive moral competence ourselves, we do not feel we could describe it,[38] and put it into programmable instructions.[39] And on the assumption that a robot will acquire morality only if we program morality into it, since we do not know how to do that, a robot cannot acquire it.[40]

But we know better than we think how to describe the content of our moral

---

36. *See* Monica Bucciarelli & P.N. Johnson-Laird, *Strategies in Syllogistic Reasoning*, 23 COGNITIVE SCI. 247, 251 (July 1999) (discussing syllogistic reasoning in computer models).

37. Actually we might need to add in permission for judicial activism: it may be that where book and case law are indeterminate on the question at issue, the right judgment would be the one with the best social consequences. These factors too could be programmed in. But it might be that even facts about this still fail to force a unique decision. What then? Once again, randomization to the rescue!

38. *See* Wendell Wallach, *Implementing Moral Decision Making Faculties in Computers and Robots*, 22 AI & SOC'Y 463 (Mar. 20, 2007) (discussing the approaches researchers are using to implement moral decision making in computers).

39. *See id.* at 469 (noting the challenges in moral judgment and how this affects the prospects for Automated Moral Agents).

40. *See* Ryan Tonkens, *A Challenge for Machine Ethics*, 19 MINDS & MACHINES 421 (July 31, 2009) (arguing it is unclear how to program ethics into machines).

competence. The totality of the body of law in North American countries is a good first approximation: these countries are well enough on the moral track that their laws tend to reflect and codify moral truth.[41]

Second, even if we could not directly program morality into a robot, we might be able to teach it to be moral by another means, namely, by exposing it to our unsystematised moral reflexes and letting it learn to copy them. Based on certain theories of how machine learning is possible this could result in the robot acquiring moral expertise even if none of us can adduce an algorithm for what we are doing.

Third, we do not have to give a robot all of morality to responsibly use a robot in some morally important context; we only need to give it enough to make it function like a moral Turing Machine for that context and that task.

Next, while it may be true that we have moral knowledge that we do not know how to express in rules yet, it is possible instead that the knowledge does not exist only because morality is in part made up as we go along—there is nothing there yet to know. And perhaps one day that process too could be automated.

Further, a large part of instruction in moral judgment is instruction in rules of thumb whose following will get you a long way—think of laws of the land, professional ethical codes for medicine, engineering and other professions, laws of war, international human rights laws, military rules of engagement, and so on. In each of these cases we have some rules which, while they do not exhaustively describe our moral duties, give us morally plausible guidelines the following of which will *usually* have us do morally defensible actions. It is imaginable that there will one day be legal and moral analogues of programs like WebMd—WebLaw, WebEthics—programs that will request input about legal or moral issues with sequenced questions, and that will produce legal or moral judgments as outputs.[42] One can imagine systems like this being used in combat; for example, "WebGeneral" might ask platoon leaders questions about proposed missions—risk to soldiers, risk to civilians, importance of target, etc.—and then output judgements and authorizations that would respect distinctness, necessity, and proportionality. In fact, arguably it would be easier to produce an autonomous machine general than an autonomous machine soldier; for the latter would require mechanization of the low-level tasks of real-world object recognition that are still so difficult for computers.[43]

---

41. *See, e.g.*, Tom R. Tyler & John M. Darley, *Building a Law-Abiding Society: Taking Public Views About Morality and the Legitimacy of Legal Authorities Into Account When Formulating Substantive Law*, 28 HOFSTRA L. REV. 707 (2000).

42. *See, e.g.*, Rob Ramey, *Autonomous Legal Reasoning and Ethics: Some Notes on Methodology* (ms. 2015, considered at the ICRC conference).

43. This puts a new spin on whether there ought to be automated war-fighting. One objection to allowing it is that it would not give enough meaningful human control over life and death decisions. *See* Peter Asaro, "*Jus nascendi*, Robotic Weapons and the Martens Clause," *in* ROBOT LAW 367–86 (Ryan Calo, Michael Froomkin and Ian Kerr eds., 2016) (arguing that public morality and human dignity require that humans should make life and death decisions, not machines, and that this requirement should be codified into law). Usually this worry concerns soldiers being replaced by robots who would then make those decisions. But Michael Horowitz asks the following variant on the question: suppose a machine is serving as a general, but a

Similar mechanisms are already in effect, such as the automation of decisions about admitting refugees into a country—there are checklists of questions and answers which determine a person's admissibility.[44] These things are automated right now in the sense that you could program a machine to operate and score the checklist. In fact, people in administrative roles are in effect being used as nothing more than very slow cogs in the conceptual machine that does its decision-making according to algorithms designed by policy makers and civil servants.[45] Likewise for university admissions policies and operations.[46] These conceptual machines are in effect programmed Turing Machines—robots—that happen to have human parts.

Returning to the military context, for situations requiring instant decisions, systems like this might be resident on computers connected to the computer infrastructures featuring in cyber war. They could make lightning fast decisions about how to respond to cyber-attacks.[47] Something like this is already being attempted in the construction of driverless cars.[48] These devices will have to prioritize moral and legal values in making decisions about what to hit and what to avoid in situations where it has become inevitable that there is going to be a

---

human soldier carries out the general's instructions to kill. *See* Michael Horowitz, *The Morality of Robotic Warfare: Assessing the Debate Over Autonomous Weapons*, MICHAELHOROWITZ.COM (Feb. 2015), http://www.michaelchorowitz.com/Documents/HorowitzLAWSEthicsDraftFeb2016.pdf. Perhaps the soldier does this almost "robotically," merely following his training. Would there have been an appropriate level of human control in this process? I do not think that matters so much as whether there has been an appropriate level of control by *morality*. And it might well be that in this situation morality is even more firmly under control of events than if there had been a human general. This is for reasons of the sort mentioned above, e.g., that human decision makers are subject to all manner of human failing.

44. *See, e.g.*, *I-94 Automation*, U.S. CUSTOMS AND BORDER PROTECTION, https://www.cbp.gov/sites/default/files/documents/I-94%20Fact%20Sheet%20-%20FINAL%20(web%20ready).pdf (last visited Apr. 11, 2016) (explaining the new 1-94 form automation for those arriving in the United States).

45. *See id.* (explaining that Customs and Border Protection officers will no longer need to attach an 1-94 to a visitor's passport upon entering the United States).

46. *See* Emmanuel Felton, *Colleges Shifts to Using "Big Data,"* THE HECHINGER REP. (Aug. 21, 2015), http://hechingerreport.org/colleges-shift-to-using-big-data-including-from-social-media-in-admissions-decisions/ (showing the switch in college admissions to the use of algorithms).

47. *See, e.g.*, Jules Zacher, *Automated Weapons Systems and the Launch of the US Nuclear Arsenal: Can the Arsenal Be Made Legitimate?,* U. OF PA. (Feb. 2016) https://www.law.upenn.edu/live/files/5443-zacher——arms-control-treaties-are-a-shampdf. Zacher specifically considers whether putting nuclear weapons under the control of AWS might make them more in conformity with international human rights law and the law of war, for, compared to humans they have superior abilities to collate large amounts of data in real time and to dispassionately implement pre-given, conditional launch directives. *Id.* at 10. Zacher's position is that, while AWS can do better with data and implementing directives, nuclear weapons are inherently illegal. *Id.* That is because even with this technology, by their very nature as having huge and indiscriminate destructive power, they cannot be made to respect distinction, necessity or proportionality. *Id.* So they should not be put under control of AWS. *Id.* Instead, they should be banned. *Id.*

48. *See Driverless Cars News*, ABCNEWS, http://abcnews.go.com/topics/news/technology/driverless-cars.htm (providing news about Google's development of driverless cars).

collision, the only decision remaining being what is going to take the hit.[49] Implementing this in a machine will require either the machine or its programmers to resolve the Trolley Problems[50] long figuring in philosophical and legal thought experiments about what morality requires.

It will be objected that creating an expert system out of our extant laws would not be to robotize the highest function of moral expertise, vis., the figuring out of moral obligations, but only some arbitrary or local conception of them—morality as conceived in the United States, for example, rather than, say, Iraq. Robots can only follow rules, not create them; they can only implement some person's or nation's conception of morality, not investigate to find the one correct moral truth (if there is one).

But this mistakenly assumes that while rules of conduct can be programmed into robots, we cannot program rules for discovering the right rules. In fact, in law there are rules for determining whether a policy has been legitimately enacted.[51] For example, the rules found in the Constitution constrain the rules that may permissibly be enacted in ordinary legislation.[52] And in morality there are a very small number of principles vying for determining morally correct behaviour. For example, there are the principles saying that we should do whatever respects natural rights, or maximizes happiness, or fulfills mutually advantageous contracts.[53] And these principles mostly agree about right conduct.[54] Indeed, there is every hope of a "Theory of Moral Everything" that will unify all plausible moral intuitions and provide argument successfully dismissing unassimilable ones.[55] There is already increasingly less debate about vast parts of morality, as seen in the increasing balance between free market and welfarist features in modern states.

---

49. *See id.* (describing how Google's driverless cars work).

50. *See* Lauren Cassani Davis, *Would You Pull the Trolley Switch? Does it Matter?*, THE ATLANTIC (Oct. 9, 2015), http://www.theatlantic.com/technology/archive/2015/10/trolley-problem-history-psychology-morality-driverless-cars/409732/ ("The trolley dilemmas vividly distilled the distinction between two different concepts of morality: that we should choose the action with the best overall consequences (in philosophy-speak, utilitarianism is the most well-known example of this), like only one person dying instead of five, and the idea that we should always adhere to strict duties, like 'never kill a human being.'").

51. *See, e.g.*, U.S. CONST. art. I.

52. *Id.*

53. *See* Steven Forde, *John Locke and the Natural Law and Natural Rights Tradition*, NAT. L., NAT. RTS. & AM. CONST., http://www.nlnrac.org/earlymodern/locke (last visited Apr. 22, 2016) (stating that human beings are subject to a moral law, and that center to this is each man's duty for self-preservation); *The History of Utilitarianism*, STAN. ENCYCLOPEDIA OF PHIL. (Mar. 27, 2009), http://plato.stanford.edu/entries/utilitarianism-history/ ("[U]tilitarianism is generally held to be the view that the morally right action is the view that produces the most good."); *Contemporary Approaches to the Social Contract Theory*, STAN. ENCYCLOPEDIA OF PHIL. (Mar. 3 1996), http://plato.stanford.edu/entries/contractarianism-contemporary/ (describing social contract theory as "the agreement (or consent) of all individuals subject to collectively enforced social arrangements shows that those arrangements have some normative property . . . .").

54. *See The Nature of Morality and Moral Theories*, U. OF SAN DIEGO, http://home.sandiego.edu/~baber/gender/MoralTheories.html (describing how all moral theories lead to decisions between right and wrong conduct).

55. *Id.*

Here, we do not just happen to have less disagreement; we are discovering truths that are compelling consensus.[56]

IV. ROBOTS AND THE KANTIAN KINGDOM OF ENDS

What then could machines not do in law and morality? Perhaps nothing, depending on how sophisticated we can design them to be. But then the real question becomes what characteristics machines would need in order to fully participate in law and morality. To answer this question we must investigate further into what law and morality are.

Law and morality are devices for the regulation of behaviour.[57] All behaviours are in the service of goals or ends. It may seem that people sometimes behave in certain ways not to bring about an end, but simply because they think it is the right way to behave. However we can see this too as end-directed behaviour simply by ascribing to these people the end of behaving in the right way.

Law and morality tell you what you may do in pursuit of your ends, what you may not do, and what you must do in helping or not obstructing others in the pursuit of their ends. When law and morality prescribe your behaviour they treat you as a legal/moral agent; and when they prescribe behaviour of others towards you they treat you as a legal/moral patient.

What is the relation between having ends and being a legal/moral agent and patient? Immanuel Kant is instructive here.[58] He thought we ought to treat people as ends in themselves, that is, as beings with goals or ends whose attainment matters to them for their own sakes.[59] Because others have ends, others ought to matter to us.[60] That is, we should take it as some reason to do something that

---

56. I said earlier that a creature is only fully able to operate the law if it recognizes that sometimes laws should be broken, e.g., when morality requires this. So, a fully legal being would also have to be a fully moral being. But note that, similarly, to be a fully moral being you must know that sometimes it is permissible to violate moral obligations. Sometimes there are things more important than doing the right thing, and someone who fails to recognize this fails to understand what morality is. As Susan Wolf points out, the world would be a worse place if people were not sometimes politically incorrect—there would be less wit and humor, for example, and less art—and you make a monstrosity of morality if you always defer your needs to it—you have to be allowed to live your life, to pursue the things you happen to care about, sometimes at the expense of moral duty. That is, you have to be a little bit eccentric relative to the demands of morality. Susan Wolf, *Moral Saints* 79 J. OF PHIL. 419–439 (Aug. 1982). And I am tempted to build that into the requirements for the successful automation of law: a full legal automaton would have to be in some ways sometimes idiosyncratic, thence to be fully able to operate morality, and thence, then, to be fully able to operate the law.

57. Thanks to Sheldon Wein for discussion on this point.

58. *See* IMMANUEL KANT, GROUNDWORK OF THE METAPHYSICS OF MORALS 42 (Mary Gregor & Jens Timmermann, trans., revised ed., Cambridge Univ. Press 2012) (1785) (indicating that humanity is an end in itself).

59. *See id.* ("[A]s far as necessary or owed duty to others is concerned, someone who has it in mind to make a lying promise to others will see at once that he wants to make use of another human being *merely as a means*, who does not at the same time contain in himself the end.").

60. *See id.* at 42–43 (indicating that to advance the ends of others, that end in itself must

another's end would be advanced by our doing it. What matters to others should in some degree matter to us because it matters to them.[61] This means we should not just use people as means to our own ends or goals. To treat people as ends and not merely as means entails either not obstructing them or even positively helping them attain their ends in certain circumstances—ones to be specified by rules.[62] Because we all have ends, because our having them requires us to help or not obstruct each other in attaining them in some circumstances, and because this is something to be regulated by rules, we are in effect all co-legislators of the rules we should follow in the so-called kingdom of ends, rules about when we must help each other and when we need only help ourselves.[63]

Moreover, Kant thought that we should do only what we could will without contradiction be done by all people in similar circumstances.[64] For behaviours of ours are morally and legally evaluable only if they are done freely.[65] And what makes us freely behaving agents is us writing laws for our own conduct.[66] The no-contradiction criterion is the test of something's ability to be a law, that is, a principle able to govern everyone always (analogously to the ways laws of nature govern all things always)—if a principle embeds a contradiction, it is impossible always to follow it; necessarily there will be a possible situation where to obey part of the principle you'll have to violate a part that contradicts the first part.[67]

We can move from the latter point to the others: imagine what you could will, without contradiction, be done by all people in similar situations. This is to imagine a law of conduct. Now imagine what states of affairs would result of all people following only permissible laws of conduct. Then these are the states of affairs people are permitted to desire, the ends they are permitted to choose. And since we all must use this same test in writing the laws, we are all co-legislators—co-writers of the laws; we write them collaboratively.

Kant also explains what the relation is between being a moral agent and being

---

also be my end).

61. *Id.*

62. *See id.* at 42 ("[H]umanity could indeed subsist if no one contributed anything to the happiness of others while not intentionally detracting anything from it; but this is still only a negative and not positive agreement with *humanity, as an end in itself*, if everyone does not also try, as far as he can, to advance the ends of others.").

63. *See id.* at 45 ("The concept of every rational being that must consider itself as universally legislating through all the maxims of its will, so as to judge itself and its actions from this point of view, leads to a very fruitful concept attached to it, namely that *of a kingdom of ends*.") (footnote omitted).

64. *See id.* at 36 ("Some actions are such that their maxim cannot even be thought without contradiction as a universal law of nature; let alone that one could will that it should become such.").

65. *See id.* at 46 ("[One should] do no action on a maxim other than in such a way, that it would be consistent with it that it be a universal law, and thus only in such a way that the will could through its maxim consider itself as at the same time universally legislating.").

66. *See* KANT, *supra* note 58, at 45 ("A rational being, however, belongs to the kingdom of ends as a member if it is universally legislating in it, but also itself subject to these laws.").

67. *See id.* at 37 (indicating that some maxims cannot become universal laws of nature because they are contradictory).

able to set ends: being a moral agent consists in being able to write coherent laws, which is the same as requiring and permitting oneself—and everyone else—to aim only at those ends that could result of everyone's following such laws.[68]

And what is it to have ends? It is at the least to be disposed to bring about the states of affairs which are these ends. Meanwhile, having morally permissible ends is the same as being disposed to behave in any of the ways permitted by the laws, and only in those ways; for all such behaviours will tend to have certain outcomes, which can then be taken for the agent's morally permitted goals or ends.[69]

But what is to determine which subset of possible ends a given person will set for herself? It doesn't matter, morally speaking, so long as the ends that get set—chosen—are compatible with the ends of others in the foregoing sense.[70]

We teach some ends to our children and leave their selection of other ends to accidents of their genetics and life-experience.[71] And we could do something similar for robots. Some ends we might design into them; others, we might allow them to choose, perhaps by giving them algorithms that will permute their subsequent experiences into choices of ends. And even though we created the algorithms for end-selection, the ends selected might be unpredictable by us because dependent on what the world throws at the robots, in the same way that a child's experiences are unpredictable, and so likewise some of her ends.[72]

The foregoing Kantian theory of morality is plausibly thought to be a component of all moral systems.[73] They vary only in how much accommodation they require of one person to another Right-leaning theories have it that we owe each other only refraining from interfering with each other in our individual pursuit of our respective ends.[74] At the other end of the spectrum, left-leaning theories

---

68. *See* Christine M. Korsgaard, *Introduction* to KANT, *supra* note 58, at xxxiv ("The committed moral agent has a deep need to place faith in some vision of how the kingdom of ends may actually be realized.").

69. *See* KANT, *supra* note 58, at 34 ("[T]he universality of the law according to which effects happen constitutes . . . a universal law of nature.").

70. *See id.* at 46 (indicating that universal laws of nature are universally legislating). For more on Kantian constraints on permissible ends, *see* Duncan MacIntosh, *Categorically Rational Preferences and the Structure of Morality*, *in* 7 VANCOUVER STUDIES IN COGNITIVE SCIENCE: MODELING RATIONALITY, MORALITY & EVOLUTION 282–301 (Peter A. Danielson, ed., 1998) and Duncan MacIntosh, *The Mutual Limitation of Needs as Bases of Moral Entitlements: A Solution to Braybrooke's Problem*, *in* ENGAGED PHILOSOPHY: ESSAYS IN HONOUR OF DAVID BRAYBROOKE 77–99 (Susan Sherwin & Peter Schotch, eds., 2007).

71. *See* Fred Guterl, *Can Children Teach Themselves?*, SCI. AM. (Feb. 27, 2013), http://blogs.scientificamerican.com/observations/can-children-teach-themselves/ (indicating that children can be taught or can teach themselves).

72. Thanks to Darren Abramson for discussion on the unpredictability of the outcomes of algorithms that take unpredictable world events as inputs.

73. *See* KANT, *supra* note 58, at 34 (discussing this theory as a universal law of nature).

74. *See* G.K. Chesterton, *Negative and Positive Morality*, http://www.chesterton.org/negative-and-positive-morality/ (discussing the difference between negative and positive morality).

have it that we owe each other positive help towards attaining our ends.[75] As a corollary, right-leaning theories forbid people having ends whose pursuit would necessarily involve interfering with other people's ends, while left-leaning theories forbid having ends whose pursuit would make impossible helping others in the pursuit of their ends.[76] Societies work out different positions on these spectrums, and part of being legally competent is the capacity and willingness to operate within one's society's expectations about these matters, and to participate in socially approved methods of resolving such conflicts about this as may arise.[77] For example, in a deliberative democracy this consists in advocating for various positions on the spectrums, accepting the verdict of democratic voting on the matter, and adhering to the rule of laws democratically so enacted.

Now, to some general claims. First, under what conditions should an autonomous device whose behaviour could affect the welfare of moral patients be allowed to be "released into the wild," that is, be allowed to choose its own behaviours? Just when it is able to do the right things for the right reasons (at least in the contexts in which it is expected to function). Second, under what conditions should such a device, rather than its manufacturers or deployers, be held accountable for what it does? When it is genuinely able to select ends, in the Kantian sense of ends, and by the Kantian means. Third, what would a device have to be like in order to satisfy the first two conditions? It would have to be a full moral agent, that is, it would have to satisfy the following four conditions:

i) It would have to be able to make choices because they are arguably the right choices to make.

ii) It would have to be able to offer a justification of its choices and to respond to counter-vailing justifications. This is because moral and legal agents are accountable, in the sense that they must be able to justify themselves in their decisions if challenged, e.g., by pointing out that their choices are consistent with permissible laws and advance permissible ends—we are to be co-legislators, after all.

iii) It would have to be able to morally learn from other moral agents and from its own experience, e.g., be able to learn from others when it has made a mistake about the permissibility of its ends or of its proposed conduct in their advancement. (This is part of co-legislating.) And to be like this, it would have to satisfy the following further condition.

iv) It must be itself a moral patient, that is, have a disposition to have ends.

Must it also be able to know at first-hand what it feels like to be frustrated or helped in the attainment of those ends and in that sense to suffer or be gladdened, thence to be able to decide to do or not do something to someone else because of

75. *Id.*

76. *Id.*

77. MacIntosh, *Categorically Rational Preferences and the Structure of Morality*, *supra* note 70, at 289.

what it would feel like to have this thing done to it? That would seem to be ideal, for it is an indisputable part of morality that one is supposed to be moved by the emotional experience of others. But it might be thought that this is not strictly necessary. Perhaps it is enough that the device have a certain disposition, the disposition to be ever less likely to set a given end for itself the more likely it is that pursuit of that end would be inimical with other agents setting and pursuing their ends. This disposition might be a functionally equivalent surrogate for being moved by the prospective suffering of others not to do things likely to produce that suffering, namely, frustrate them in the pursuit of their ends. And programming such a disposition into a machine seems like something attainable without having to solve the problem of how to make a machine have feelings.[78]

Suppose then that we have a machine that can set ends in this sense: it can chose a state of affairs to be disposed to produce, it can say what is good about this state of affairs when challenged by others, it will only dispose itself to produce this state of affairs if its doing so is compatible with other agents having reasonable prospects of attaining the states of affairs they have set as ends for themselves, and so it will comply with laws the general compliance with which is required to make these several states of affairs attainable. But suppose that this machine cannot suffer, cannot feel frustration at the failure of the attaining of its end. Then it may be that none of us needs to limit our own behaviour so that this machine can attain its end. For no one—not us, and not even the machine—cares whether it comes to pass.

If this is true, then any ends that robots may have ought to command respect from humans, and robots ought to get to vote on which laws ought to prevail and so on which ends ought to have a chance of being brought about, only if the robots can suffer from failing to attain such ends. And this will require solving the problem of how to engineer affect into a machine.

Without the capacity to feel, robots cannot be moral or legal patients, only property, and as such, their mere existence could not change people's moral or legal duties.[79] Contrast this with what happens when a human baby—prone to

78. There may however be special contexts where a machine cannot fulfill the moral duties required of it without being capable of feelings, contexts where it is not enough to do the right thing; you must also feel the right way. Think of palliative care giving, or of nursing more broadly: to do it properly you have to do it with loving compassion. We might still let a machine do the physical tasks involved even if it could not feel (perhaps times are desperate and we are short-staffed), but this would be a morally second best state of affairs. A judge handing out a sentence might likewise be obliged to have certain feelings about it, e.g., as part of expressing the disapprobation of the community. In yet other contexts, however, an agent's lacking feelings might morally improve the situation, e.g., if a killing is necessary and it would be better if it could be done coldly and without the agent doing it feeling guilt. *See* Duncan MacIntosh, *Fire and Forget: A Defense of the Use of Autonomous Weapons Systems in War and Peace* (Oct. 2015) (unpublished paper) (on file with author).

79. *See* John Markoff, *Relax, the Terminator Is Far Away*, N.Y. TIMES (May 25, 2015), http://www.nytimes.com/2015/05/26/science/darpa-robotics-challenge-terminator.html?_r=0 (discussing the difficulty of creating cognitive artificial intelligence and the resulting trend of machine-human pairings).

having feelings upon the attaining or frustrating of its ends—comes to exist: suddenly we have duties to it from the mere fact of its existence.[80] A robot's capacity to make judgments would not matter, except insofar as it might have knowledge useful in figuring out what would advance *our* goals, or what they should be. If a robot were to judge that a certain end would be good to bring about, but the robot did not have a stake in bringing it about, a stake in the sense that the robot would care if the end were not to come about, then the robot's judgement is merely information: we might be curious what reasons the robot had for judging the end a good one; for perhaps these reasons would persuade us of the value of the end. By contrast, if a robot were able to invest concern in the attaining of a thing it judged to be of value, then the robot would exist as more than just a provider of information.

Strangely, an unfeeling robot might be fit to rule us, but not to cast votes with us in collective democratic self-governance. The machine might do better than us at figuring out arrangements in which we each best thrive—imagine telling the machine what our ends are and asking it what rules are such that if we all follow them, we will all be most likely to attain our ends.[81] But that would be different from letting the machine have a vote. For voters are supposed to act as Rousseauian citizens, voting not just the public interest—the interest of everyone else—but their own interest too; the purpose of democratic voting is to sum together everyone's felt preferences in the selection of policies and courses of action.[82] And if robots do not have interests, then they have no interest to take into account in voting either. To fully participate in law is, among other things, to vote, to have rights, to make and waive rights claims, to be someone others must make some accommodation to on account of having ends, to be something that must be accommodative of itself for the same reason, and to have a welfare that can be legally curated, a welfare that would be improved by the attaining of its ends.[83]

---

80. *See* Duncan MacIntosh, *Who Owns Me, Me or my Mother? How to Escape Okin's Problem for Nozick's and Narveson's Theory of Entitlement, in* LIBERTY, GAMES AND CONTRACTS: JAN NARVESON AND THE DEFENCE OF LIBERTARIANISM 157–62 (Malcolm Murray ed., 2007) (discussing the tension between deciding that an offspring is its own person or that it is the property of its mother and how entitlements enjoyed by offspring factor into the calculation).

81. We have already begun to defer to machines in this way on financial matters. *See* Tom C.W. Lin, *Financial Weapons of War*, 100 MINN. L. REV. 1377, 1386 (2016) (discussing the modern financial infrastructure which also serves as a new battlefield in contemporary warfare). At the ICRC conference mentioned above, Thomas Lin presented a paper that has since been published which included a discussion of a financial program that seeks and instantly trades in investments which maximize both investment yield and tax efficiency. *Id.* The program does this better than humans could do and in a way unpredictable by humans, since they cannot compute at the speed of a computer running a program. *Id.* Lin mooted the implications of entrusting such things as equity trading to computers. For a summary of the ICRC conference and a list of attendees, *see* Laura Burgess, *Autonomous Legal Reasoning? Legal and Ethical Issues in the Technologies of Conflict*, INT'L COMM. OF THE RED CROSS: INTERCROSS BLOG (Dec. 7, 2015), http://intercrossblog.icrc.org/blog/048x5za4aqeztdiu3r8f96s8m7lzom?rq=autonomous.

82. *See* TIMOTHY O' HAGAN, ROUSSEAU 79 (Ted Honderick ed., 2005) (discussing Ruosseau's views on the relationship between individual will and corporate will in democracy).

83. *See, e.g.*, *Citizenship Rights and Responsibilities*, U.S. CITIZENSHIP & IMMIGR. SERV. (last visited Apr. 11, 2016), https://www.uscis.gov/citizenship/learners/citizenship-rights-and-

The possibility of us being able to engineer robots so that they have ends both in the functional and in the affective sense, robots that would then have an interest, raises large issues I cannot get into here. For example, if we get to the point where we can make robots like this, would it be politically permissible to do so? This is complicated, for they could be manufactured in large number, they could have the right to vote, and they could overwhelm our own voting rights, for example. They would also have rights to consume resources, and this would have to be balanced against their capacity to make economic contributions benefitting us all. So there would be population control issues. Moreover, we would not be permitted to bring into existence beings who had affect and yet had no means of arranging positive affect. For that would mean we had created feeling beings in full knowledge that their lives would be a torture, something antithetical to according them Kantian respect as beings that have ends.[84] And on and on.

These considerations might even cancel out the original motivation I offered for producing robots to operate the system of law, namely, that we could engineer them to be more objective than us in the administration of the rule of law. However welcome that might be, it might not be worth the price of having yet more stake-holders whose interests would have a claim to be served. On the other hand, if the resulting robots really did comply with Kantian edicts in their selection of ends and in their endorsing of laws for regulating conduct advancing those ends, arguably it would be a certainty that this would be a good thing for everyone—for all of them and all of us. For as David Gauthier would observe, it would then be guaranteed that our permitted ends and theirs would be such that their mutual advancement would yield a surplus of goods from co-operation, and a mutually advantageous distribution of those goods.[85] (Well, guaranteed unless we are facing situations of absolute scarcity of resources, situations where it is increasingly likely that our interactions will be zero-sum games—games where there are winners only if there are also losers—rather than ones generating surpluses from cooperation.)

In brief, then, without ends and an emotional stake in their attainment, you cannot be a moral or a legal patient, because then you have no interest mandating your protection under law and morality. And it is doubtful you could be a full moral agent, either. You might be able to autonomously behave in ways beneficial or harmful to others, but you could not do so being entitled to participate as a stake-holder in the designing of the rules governing this, because you would yourself have no ends whose attainment was emotionally consequential for you, requiring and deserving, therefore, to be brought into respectful relations with others. A fully moral and legal agent must *care*.

---

responsibilities (listing the rights and responsibilities of American citizenship as defined by the United States government).

84. For more on the duties of creators to the persons they create, *see* Duncan MacIntosh, *Who Owns Me, Me or my Mother? How to Escape Okin's Problem for Nozick's and Narveson's Theory of Entitlement, in* LIBERTY, GAMES AND CONTRACTS: JAN NARVESON AND THE DEFENCE OF LIBERTARIANISM 157–62 (Malcolm Murray ed., 2007).

85. *See generally* DAVID GAUTHIER, MORALS BY AGREEMENT 14 (1986).