Robots and Responsibility from a Legal Perspective

Peter M. Asaro, *Member, IEEE*

Abstract—This paper considers how legal theory, or jurisprudence, might be applied to robots. This is done with the intention of determining what concepts and approaches to robot ethics might be gained from taking a legal perspective. In many cases, legal theory is suggestive of possible approaches to problems that will require further work to evaluate. It concludes that legal theory does allow us to define certain classes of ethical problems that correspond to traditional and well-defined legal problems, while other difficult practical and meta-ethical problems can not be solved by legal theory alone.

I. INTRODUCTION

IT has been argued that a good place to begin thinking about ethics in robots, is to look to the existing legal frameworks and how these might apply to robots today and in the near future [1]. In this presentation I want to consider a few of the most significant ways in which we might use jurisprudence in order to get a better understanding of robot ethics. Primarily, this will consist in examing the nature of legal responsibilities involved in the production and use of robots and will be in this sense very general. I believe this is a better place to start thinking about robot ethics, rather than the immediate issues involved in the specific design choices of, or possible moral prohibitions on, the engineer, because those specific issues require a well defined sense of how responsibility is focused, transferred and distributed in and around robots. I also believe that it is only by understanding what makes robots unique from other technologies that we can begin to think about what makes robot ethics distinct from engineering ethics more generally. To do this, we will focus on the concept of legal responsibility, and begin by considering robots as being like any other technological device or product, and from there move to thinking about what could make robots different in the eyes of the law, and what special considerations robotics engineers might need to make as a result.

It is important to be clear that legal responsibility is not exactly the same thing as moral responsibility. Still, I believe it represents an excellent starting point for thinking about robot ethics for several reasons. As others have already noted [2], there is no single generally accepted moral theory, and only a few generally accepted morals. And while there are differing legal interpretations of cases, and differing legal opinions among judges, the legal system ultimately tends to do a pretty good job of settling questions of responsibility. Thus, by beginning to think about these issues from the perspective of legal responsibility, we are more likely to arrive at practical answers. This is because both 1) it is likely that legal requirements will be how robotics engineers will find themselves initially compelled to build robots ethically, and so the legal framework will structure those pressures and their technological solutions, and 2) the legal framework provides a system for understanding agency and responsibility, so we will not need to wait for a final resolution of which moral theory is "right" or what moral agency "really is" in order to begin to address the ethical issues currently facing robotics.

We might think of legal responsibility as a subset of moral responsibility. There is certainly a large overlap between what is legal and what is moral, even under differing moral theories. Indeed the disagreements between them represent a relatively small set of cases where what is morally acceptable, or required, is in violation of the law (*e.g.* civil disobedience, or speeding an injured person to the hospital), and the relatively large set of actions which are legally acceptable but morally despicable (*e.g.* being rude or obnoxious, making racists statements, or violating someone's trust outside any legally-binding contracts). While these cases do arise in real life, it is also safe to assume that the vast majority of practical decisions faced by humans, and potentially by robots, will be of the sort that legal and moral theories will largely agree on what the appropriate actions are. As such, building a robot that is capable of safeguarding the legal responsibility of those who build it and use it, would at least be a good start in building one which has moral responsibility.

How then can the law help us in our thinking about robots? There are several relevant aspects of the law, and we will consider each in turn, but first a brief overview. In the most straightforward sense, the law has a highly developed set of cases and principles that apply to *product liability*, and we can apply these to the treatment of robots as commercial products. As robots begin to approach more sophisticated human-like performances, it seems likely that they might be treated as *quasi-agents* or quasi-persons by the law, enjoying only partial rights and duties. A closely related concept will be that of *diminished responsibility*, in which agents are considered as being not fully responsible for their own actions. This will bring us to the more abstract concept of *agency* itself in the law, and how responsibility is transferred to different agents. Finally we will consider *corporate punishment*, which is relevant both as it applies to cases of wrongdoing in product liability, but also because it addresses the problem of legal punishments aimed at non-human agents, namely corporations.

P. M. Asaro is with the HUMLab, Umeå University of Umeå, Sweden, +46 + (0)90 - 786 9286 (e-mail: peterasaro@sbcglobal.net).

II. RESPONSIBILITY AND LIABILITY: ROBOTS AS PRODUCTS

In the system of Anglo-American law, a distinction is drawn between criminal and civil law. Civil law is traditionally called tort law, and deals primarily with property rights and infringements, such as damage to property or other harms, and seeks justice by compelling wrongdoers to compensate those who were harmed for their loss. Criminal law deals with what we often think of as moral wrongdoing, stealing, murder, *etc.*, and seeks justice by punishing the wrongdoer. The difference is that between someone building a toy robot which shoots little plastic missiles that causes several small children to choke to death, and someone who builds a robot with a built-in bomb that kills a number of people on a public street. In each case there is a robot causing death, but in the first case the parents of the children would file a lawsuit against the manufacturer seeking monetary compensation, and in the second case the government would find, arrest, prosecute and punish the individuals responsible. I want to set criminal law aside for the moment, however, as civil law is actually more relevant to robots as they now exist insofar as they might be capable of material wrongdoing.

Even if we make no assumptions about the intentions, consciousness, or moral agency of robots, we can still apply the basics of civil law to robots as they now exist. That is, we can assume that robots are completely unremarkable technological artifacts, no different than toasters or cars, and there are still legal and moral issues connected with their production and use. This point is easily demonstrated by noting that the companies that manufacture robots, such as the *Furby* and *AIBO,* most certainly employ and retain lawyers who are paid to advise them on their legal responsibilities in producing, advertising and selling these robots to the general public. Moreover, I believe that many of the concerns about the possible harms that robots might cause would ultimately fall under this mundane interpretation.

The relevant legal concept in cases like our toy robot that chokes small children is *negligence*. Negligence implies that the manufacturer failed to do something that was morally or legally required, and thus they can be held responsible for certain harms produced by their product. Legally culpable forms of negligence depend upon either *failures to warn*, or *failures to take proper care*. A *failure to warn* occurs when the manufacturer was knowingly aware of a risk or danger but failed to notify consumers of this risk. This is the reason why there are now so many warning labels on various products, and in the example above the manufacture might avoid liability by putting a label on the package stating that the robot contains parts that are a choking hazard to young children. A *failure to take proper care or avoid foreseeable risks* is more difficult to prove in court because it is more abstract, and involves cases where the manufacturer cannot be shown to have known about a risk or danger from the product. In these cases, it is argued that the given danger or risk was in some sense obvious or easily foreseeable, even if the manufacturer failed to

recognize it. In order to prove this, lawyers often bring in experts to testify that those risks were obvious, and so forth.

Another interesting aspect of liability is that it can differentially apportioned. That is to say, for example, one party might be 10% responsible, while another is 90% responsible for some harmful event. This kind of analysis of the causal chains resulting in harms is not uncommon, especially in traffic accidents and product liability cases. In many jurisdictions there are laws imposing joint and several liability, which holds all parties equally responsible for compensation, even if they are not equally responsible for the harm. Nonetheless, these cases still recognize that various factors and parties contribute differentially to some event.

This is an important consideration to keep in mind when thinking about robot ethics. For instance, a badly designed object recognition algorithm might be responsible for some damage caused by a robot, but a bad camera could also contribute, as could a weak battery, or a malfunctioning actuator, *etc.* Moreover, the context in which the robot has been placed, and perhaps the instructions given by its owners may also be the principle, or contributing, causes of some harm in which a robot is the proximate cause. In short, there is a limit to what robot engineers and designers can do to limit the potential uses and harms caused by their products because other parties, namely the consumers and users of robots, will choose to do all sorts of things with them and will have to assume the responsibility for those choices. Similarly, there will always be risks inherent in the use of robots, and at some point the users will be judged by the court to have knowingly assumed these risks in the very act of choosing to use a robot.

The potential *failure to take proper care*, and the reciprocal responsibility to take proper care, is perhaps the central issue in practical robot ethics. What constitutes proper care, and what risks might be foreseeable, or in principle unforeseeable, is a deep and vexing problem. This is due to the inherent complexity of potential future interactions, and the relative autonomy of the product once it is produced. Sophisticated robots that will be capable of interacting with people and the world in highly complex ways, and that may develop and learn new ways of interacting which extend beyond their intended design, present a difficult future to attempt to foresee risks in. Robot ethics shares this double-edged problem with the bio-engineering ethics–both the difficulty in predicting the future interactions of a product when the full scope of possible interactions can at best only be estimated, and in producing a product that is an intrinsically dynamic and evolving system whose behavior may not be easily guided after it has been produced.

The classic defense against charges of *failures to warn* and *failures to take proper care* is the *industry standard defense*. The basic argument of the *industry standard defense* is that the manufacturer acted in accordance with the stated or unstated standards of the industry they are participating in. Thus, they were merely doing what other similar manufacturers were doing, and they were thus taking

proper care as measured against their peers. This need for a relative measure again points to the vagueness of the concept, and the inherent difficulty of determining what specific and practical legal and moral duties follow from the obligation to take proper care. This kind of defense also fails to tell us what sorts of practices *should* be followed in the design of robots. That is, robot ethics should be concerned with the establishment of standards for the robot industry which will ensure that the relevant forms of proper care are taken. There is quite a bit more to be said about this topic, obviously, but for now we should stay with the law.

III. AGENTS, QUASI-AGENTS AND DIMINISHED RESPONSIBILITY

The law offers several ways of thinking about the distribution of responsibility in complex cases. As we saw in the previous section, responsibility for a single event can be divided amongst several parties, and each party can even be given a quantitative share of the total responsibility. We will now consider how even a single party's responsibility can be divided and distributed. Modern legal systems were established on a presupposition that all legal entities are persons. While a robot might someday be considered a person, we are not likely to face this situation any time soon. However, the law has also been designed to deal with several kinds of non-persons, or quasi-persons, and we can look to these for some insights on how we might treat robots that are non-persons, or quasi-persons.

Personhood is a hotly debated concept, and many perspectives in that debate are strongly held beliefs based in religious faith and philosophical dispositions. Most notably, the case of unborn human fetuses, and the case of severely brain damaged and comatose individuals have led to much debate in the United States over their appropriate legal status and rights. Yet, despite strongly differing perspectives on such issues, the legal systems in pluralistic societies have found ways to deal practically with several border-line cases of personhood.

Minor children are a prime example of quasi-persons. Minors do not enjoy the full rights of personhood that adults do. In particular they cannot sign contracts or become involved in various sorts of legal arrangements because they do not have the right to do so as minors. They can become involved in such arrangements only through the actions of their parents or legal guardians. In this sense they are not legal persons. In another sense, the killing of a child is murder in the same way that the killing of an adult is, and so a child is a legal person in this sense. Children can thus be considered a type of quasi-person, or legal quasi-agent. The case of permanently mentally-impaired people can be quite similar to children. Even fully-fledged persons can claim temporary impairments of judgment, and thereby *diminished responsibility* for their actions given certain circumstances, *e.g.* temporary insanity. The point is that some aspects of legal agency can apply to entities which fall short of fully-fledged personhood, and full responsibility, and it seems reasonable to think that some robots will eventually become

a kind of quasi-agent in the view of the law before they achieve personhood.

The concept of personhood is deeply tied to the notion of agency. The law also deals explicitly with agency and, interestingly enough, it does so in order to address cases in which the power of agency is transferred between parties. The law of agency is a highly specialized field that deals mainly with the talent agents of athletes and entertainers, and to some extent insurance, travel, and real estate agents. These agents are empowered by their employers, whom they thereby represent for the purpose of negotiating contracts and making various agreements on their behalf. The individuals are bound by the contracts that their agents sign just as if they had signed them themselves, except in cases where one can prove misconduct on the part of the agent. To act as someone's agent is to enact their legal powers from afar, and is in this sense a form of distribution of legal agency.

The possible application to robotics seems inviting–robots could be seen in many cases as agents acting on the behalf of others. Accordingly, the legal responsibility for the actions of a robot falls on the individual who grants the robot permission to act on their behalf. If it is not already clearly enough implied by the law, it might be advisable to make a law which makes such legal responsibilities explicit. Such a law might, however, place a too heavy burden on the owners of robots, preventing the adoption of robots due to risk, or unfairly protecting manufacturers who might share in the responsibility of misbehaving robots due to poor designs.

IV. CRIME, PUNISHMENT AND PERSONHOOD IN CORPORATIONS AND ROBOTS

Crime and punishment are central concepts in both law and morality, yet they might seem out of place in a discussion of robot ethics. While we can imagine a humanoid robot of such sophistication that it is effectively, or indistinguishably, a person, these robots will be easier to find in science fiction than in science fact for a long time to come. There are, however, technologically possible robots that may approach actions that we might consider, at least at first glance, to be criminal. If so, how might the law instruct us to treat such cases?

As stated earlier, criminal law is concerned with punishing wrongdoers, whereas civil law is primarily concerned with compelling wrongdoers to compensate those harmed. There is an important principle underlying the distinction: that crimes deserve to be punished, regardless of any compensation to those directly harmed by the crime. Put another way, the harmed party in a crime is the whole of society. Thus, the case is prosecuted by the state or "the people", and the debt owed by the wrongdoer is owed to the society. While the punishments may different forms, the point of punishment is traditionally conceived of as being corrective in one or more senses: that the wrongdoer pays their debt to society (justice); that the wrongdoer is to be reformed so as not to repeat the offense (reform); or that

other people in society will be deterred from committing a similar wrong (deterrence).

There are two principle problems with applying criminal law to robots: 1) Criminal actions require a moral agent to perform them, and 2) How is it possible to punish a robot? Moral agency is deeply connected to our concepts of punishment. Moral agency might be defined in various ways, but it ultimately must serve as the subject who is punished. Without moral agency, their can be harm but not guilt. Thus, there is no debt incurred to society unless there is a moral agent to incur it–it is merely an accident and not a crime. Similarly, only a moral agent can be reformed, which implies the development or correction of a moral character–otherwise it is merely the fixing of a problem. And finally, deterrence only makes sense when moral agents recognize the similarity of their potential choices and actions to those of another moral agent who has been punished for the wrong choices and actions–without this reflexivity of choice by a moral agent, and recognition of similarity between moral agents, punishment cannot possibly result in deterrence.

We saw in the previous section that it is more likely that we will treat robots as quasi-persons long before they achieve full personhood. Lawrence Solum [5] has given careful consideration to the question of whether an artificial intelligence (AI) might be able to achieve legal personhood, using a thought experiment in which an AI acts as the manager of a trust. He concludes that while person hood is not impossible in principle, it is also not clear how we would know that any particular AI has achieved it. The same argument could be applied to robots. Solum imagines a legal Turing test in which it comes down to the determination of a court whether an AI could stand trial as a legal agent in its own right, and not merely a proxy or agent of some other legal entity. However, it seems that a court would ultimately base its decision on whether the robot in question has moral agency, and whether it is possible to punish it–can you fine or imprison an AI that mismanages a trust? In cases of quasi-personhood and diminished responsibility, children and the mentally impaired are usually shielded from punishment as a result of their legal status.

There is, however, in the law a relevant case of legal responsibility resting in a non-human, namely corporations. The limited liability corporation is a non-human entity that has been effectively granted the legal rights of a person. Corporations can own property, sign contracts, and be held liable for negligence. In certain cases, corporations can even be punished for criminal activities such as fraud, criminal negligence and causing environmental damage. A crucial aspect of treating corporations as persons depends on the ability to punish them, though this is not nearly so straightforward as it is for human persons. As a 17th century Lord Chancellor of England put it, corporations have no soul to damn and no body to kick, so how can they be expected to have a conscience? Of course, corporations exist to make money, for themselves or stockholders, and as such can be given monetary punishments, and in certain cases such as anti-trust violations, split apart or dissolved

altogether. They cannot, however, be imprisoned. And as a result of this, and other aspects of their being complex socio-technical systems in which there are many stake-holders differently related to the monetary wealth of a corporation, it can be difficult to assign a punishment that achieves justice, reform and deterrence while meeting the requirements of fairness, such as proportionality.

Clearly, robots are different in many important respects from corporations. However, there are also many important similarities, and it is no coincidence that John Coffee's [3] seminal paper on corporate punishment draws heavily on Herbert Simon's [4] work on organizational behavior and decision making. Nonetheless, a great deal of work needs to be done in order to judge just how fruitful this analogy is. While monetary penalties work as punishments for corporations, this is because they target the essential reason for the existence of corporations–to make money. The essential purposes of robots may not be so straightforward, will vary from robot to robot, and may not take a form that can be easily or fairly penalized by a court.

The most obvious difference is that robots do have bodies to kick, though it is not clear that kicking them would achieve the traditional goals of punishment. The various forms of corporal punishment presuppose additional desires and fears of being human that may not readily apply to robots–pain, freedom of movement, mortality, *etc.*. Thus, torture, imprisonment and destruction are not likely to be effective in achieving justice, reform or deterrence in robots. There may be a policy to destroy any robots that do harm but, as is the case with animals that harm people, it would be a preventative measure to avoid future harms rather than a true punishment. Whether it might be possible to build in a technological means to enable genuine punishment in robots is an open question.

V. CONCLUSIONS

I hope that this brief overview of how certain legal concepts might be applied to current and future robots has convinced you that jurisprudence is a good place to begin framing some of the issues in robot ethics. I do not claim that this is the only viable approach, or that it will be capable of resolving every issue in robot ethics. Rather, I maintain that we can delineate different classes of ethical problems, some of which will have straightforward solutions from a legal perspective, while other classes will remain unresolved. In terms of thinking about robots as manufactured products, many of the most practical and pressing issues face robotics engineers can be seen as being essentially like those facing other engineers. I these cases, it is necessary to take proper care in imagining, assessing and mitigating the potential risks of a technology. Just what this means for robotics will, of course, differ from other technologies, and should be the focus of further discussion.

There are also a host of meta-ethical questions facing robot ethics that are largely ignored by the legal perspective. While moral agency is significant to the legal perspective, jurisprudence alone cannot determine or define just what moral agency is. Similarly, the ethical questions facing the

building of truly autonomous technologies demands its own treatment. While there was no room to discuss it in this paper, the legal perspective can also go a long way toward framing issues in the use of robots in warfare. In particular, international law, humanitarian law, uniform codes of military conduct, the Geneva Conventions, the Nuremberg Principles, and international laws banning anti-personnel mines, are all eminently relevant to theorizing the ethics of using robot technologies in warfare, and will be a subject for further study.

REFERENCES

[1]   Peter Asaro, "What Should We Want from a Robot Ethic?" *International Review of Information Ethics*, forthcoming.

[2]   Colin Allen, Gary Vamer and Jason Zinser, "Prolegomena to any future artificial moral agent," *Journal of Experimental and Theoretical Artificial Intelligence,* **12** (2000), pp. 251-261.

[3]   John C. Coffee, Jr., "'No Soul to Damn: No Body to Kick': An Unscandalized Inquiry into the Problem of Corporate Punishment," *Michigan Law Review*, Vol. 79, No. 3 (Jan. 1981), pp. 386-459.

[4]   Herbert Simon, *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*. New York, NY: Free Press.

[5]   Lawrence Solum, "Legal Personhood for Artificial Intelligences," *North Carolina Law Review*, (April 1992), pp. 1231-1287.