Intelligent Robotics and Autonomous Agents
Edited by Ronald C. Arkin

Dorigo, Marco, and Marco Colombetti, *Robot Shaping: An Experiment in Behavior Engineering*

Arkin, Ronald C., *Behavior-Based Robotics*

Stone, Peter, *Layered Learning in Multiagent Systems: A Winning Approach to Robotic Soccer*

Wooldridge, Michael, *Reasoning about Rational Agents*

Murphy, Robin R., *An Introduction to AI Robotics*

Mason, Matthew T., *Mechanics of Robotic Manipulation*

Kraus, Sarit, *Strategic Negotiation in Multiagent Environments*

Nolfi, Stefano, and Dario Floreano, *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*

Siegwart, Roland, and Illah R. Nourbakhsh, *Introduction to Autonomous Mobile Robots*

Breazeal, Cynthia L., *Designing Sociable Robots*

Bekey, George A., *Autonomous Robots: From Biological Inspiration to Implementation and Control*

Choset, Howie, Kevin M. Lynch, Seth Hutchinson, George Kantor, Wolfram Burgard, Lydia E. Kavraki, and Sebastian Thrun, *Principles of Robot Motion: Theory, Algorithms, and Implementations*

Thrun, Sebastian, Wolfram Burgard, and Dieter Fox, *Probabilistic Robotics*

Mataric, Maja J., *The Robotics Primer*

Wellman, Michael P., Amy Greenwald, and Peter Stone, *Autonomous Bidding Agents: Strategies and Lessons from the Trading Agent Competition*

Floreano, Dario, and Claudio Mattiussi, *Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies*

Sterling, Leon S., and Kuldar Taveter, *The Art of Agent-Oriented Modeling*

Stoy, Kasper, David Brandt, and David J. Christensen, *An Introduction to Self-Reconfigurable Robots*

Lin, Patrick, Keith Abney, and George A. Bekey, editors, *Robot Ethics: The Ethical and Social Implications of Robotics*

Robot Ethics
The Ethical and Social Implications of Robotics

Edited by Patrick Lin, Keith Abney, and George A. Bekey

The MIT Press
Cambridge, Massachusetts
London, England

2012

Singer, P. W. 2009. *Wired for War. The Robotics Revolution and Conflict in the 21st Century*. New York: Penguin Group.

U.S. Department of Defense. 2009. *FY2009–2034 Unmanned Systems Integrated Roadmap*. Washington, DC: Department of Defense.

Verchio, Donna Marie. 2001. Just say no! The SIrUS Project: Well-intentioned, but unnecessary and superfluous. *Air Force Law Review* 21 (Spring): 224–225.

Wallach, Wendell, and Colin Allen. 2009. *Moral Machines: Teaching Robots Right From Wrong*. New York: Oxford University Press.

11 A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics

Peter M. Asaro

The continued advancement of robotic technologies has already begun to present novel questions of social and moral responsibility. While the overall aim of this collection is to consider the ethical and social issues raised by robotics, this chapter will focus on the legal issues raised by robotics. It starts from the assumption that we might better understand the social and moral issues surrounding robotics through an exploration of how the law might approach these issues. While it is acknowledged that there are instances where what is legal is not necessarily morally esteemed, and what is morally required may not be legal, in general, there is a significant overlap between what is legal and what is moral. Indeed, many of the crucial concepts are shared, and as such this chapter will explore how the law views responsibility, culpability, causality, intentionality, autonomy, and agency. As a philosopher, rather than a lawyer or legal scholar, my concern will be with these theoretical concepts, and how their justificatory frameworks can be used to interpret and apply law to the new kinds of cases, which teleoperated, semi-autonomous, and fully autonomous robotics have already, or, may soon, present. Insofar as some of the issues will also involve matters of industry and community standards, public opinions, and beliefs, as well as social values and public morals, the chapter will consider questions of value. While my concern will be primarily with the law as it is typically understood and applied in the United States, my aim is that these reflections will also prove useful to scholars and lawyers of other legal traditions.

Indeed, the legal issues raised by robotic technologies touch on a number of significant fundamental issues across far-ranging areas of law. In each of these areas, there can be found existing legal precedents and frameworks which either directly apply to robotics cases, or which might be extended and interpreted in various ways so as to be made applicable. My aim is to consider each in turn, as well as to identify the principles that might underlie a coherent legal understanding of the development and use of robotic systems. Furthermore, I will consider the means by which we might judge the potential of robots to have a legal standing of their own. It will thus be helpful to organize the discussion in terms of both the salient types of

robots—teleoperated, semi-autonomous, and autonomous—as well as the principal areas of law, criminal and civil.[1]

The most obvious issues that arise for the application of the law to robotics stem from the challenge that these complex computational systems pose to our traditional notions of intentionality, as well as how and whom to punish for wrongful acts (Bechtel 1985; Moon and Nass 1998). Most of the scholarship on law and robotics to date has focused on treating robots as manufactured products (Asaro 2006, 2007; Schaerer, Kelley, and Nicolescu 2009), subject to civil liability, or on whether robots can themselves become criminally liable (Dennett 1997; Asaro 2007), or the challenges robotic teleoperation poses to legal jurisdiction (Asaro 2011). I will begin by considering the more straightforward cases of semi-autonomous robots, which can be treated much like other commercial products. For these cases, the law has a highly developed set of precedents and principles from the area of law known as *product liabilities*, which can be applied.

I will then consider the implications of increasingly autonomous robots, which begin to approach more sophisticated and human-like performances. At some point in the future, there may be good reasons to consider holding such robots to standards of criminal or civil liability for their actions, as well as compelling reasons to hold their owners and users to higher, or lower, standards of responsibility for the wrongdoings of their robots. These considerations will draw upon a variety of legal areas with similar structures of distributing intention, action, autonomy, and agency. There exist certain similarities between such robots and their owners and controllers, and the ways in which individuals have traditionally been held to account for the wrongdoings of other subordinate intelligent, sentient, conscious, autonomous, and semi-autonomous agents. Examples include laws pertaining to the assignment of responsibility between animals and their owners, employees and their bosses, soldiers and their commanders, and slaves and their masters, as well as *agency law*, in which agents are entrusted with even greater levels of responsibility than is the case with typical subordinates. There are also issues involving whether robots themselves are entitled to legal standing, redress, or even rights, including the ability to sign contracts, be subject to criminal liabilities, or the means by which they might be justly subjected to punishment for crimes. This will bring us to consider the punishments against other kinds of nonhuman legal agents, namely corporations, and what can be learned about robot punishments from corporate punishments.

11.1 Robots and Product Liability

Many of the most common potential harms posed by robotic systems will be covered by civil laws governing product liability. That is, we can treat robots as we do other technological artifacts—such as toys, weapons, cars, or airliners—and expect them to

raise similar legal and moral issues in their production and use. In fact, the companies that currently manufacture robots are already subject to product liability, and retain lawyers who are paid to advise them on their legal responsibilities in producing, advertising, and selling these robots to the general public. Most of the public's current concerns about the possible harms that robots might cause would ultimately fall under this legal interpretation, such as a robotic lawnmower that runs over someone's foot, or a self-driving car that causes a traffic accident.

It will be helpful at this point to review the basic elements of product liability law.[2] Consider, for example, a toy robot that shoots small foam projectiles. If that toy were to cause several children to choke to death, the manufacturer might be held liable under civil law, and be compelled to pay damages to the families that lost children because of the toy. If it can be proven in court that the company was *negligent*, with regard to the defects, risks, and potential hazards arising from the use of their product, then the company could also be criminally, as well as civilly, liable for the damages caused to victims by their product. Legal liability due to negligence in product liability cases depends on either *failures to warn*, or *failures to take proper care* in assessing the potential risks a product poses. A failure to warn occurs when the manufacturer fails to notify consumers of a foreseeable risk, such as using an otherwise safe device in a manner that presents a potential for harm. For example, many power tools display warnings to operators to use eye protection or safety guards, which can greatly reduce the risks of using the device. The legal standard motivates manufacturers to put such warning labels on their products, and, in the preceding example, the manufacturer might avoid liability by putting a label on the package, stating that the robot toy contains parts that are a choking hazard to young children.

A failure to take proper care is more difficult to characterize. The idea is that the manufacturer failed to foresee a risk, which, if they had taken proper care, they would have likely foreseen. This counterfactual notion is typically measured against a somewhat vague community standard of reason, or an industry standard of practice, about just what proper care is expected among similar companies for similar products. In some sense, the more obvious the risk is, according to such a standard, the more likely that the negligence involved rises to the level of criminality.

The potential failure to take proper care, and the reciprocal responsibility to take proper care, is perhaps the central issue in practical robot ethics from a design perspective. 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 anticipating potential future interactions, and the relative autonomy of a robotic product once it is produced. It is likely to be very difficult or impossible to foresee many of the risks posed by sophisticated robots that will be capable of interacting with people and the world in highly complex ways—and may even develop and learn new ways of acting that extend beyond their initial design. Robot ethics shares this

problem with bioengineering 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 controlled after it has been produced and released into the world (Mitcham 1990).

A 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 were taking proper care as measured against their peers. This appeal to a relative measure again points to the vagueness of the concept of proper care, and the inherent difficulty of determining what specific and practical legal and moral duties follow from the obligation to take proper care. This vague concept also fails to tell us what sorts of practices *should* be followed in the design of robots. An obvious role for robot ethics should be to seek to establish standards for the robot industry, which will ensure that the relevant forms of proper care are taken, and I believe this should be one of the primary goals for future robot ethics research.

If the company in question willfully sought to remain ignorant of the risks its robotic products might pose, such as by refusing to test a product or ignoring warnings from designers, then its negligence could also be deemed criminal. This would be a case of *mens rea*, in which the culpable state of mind is one of ignorance, either willfully or unreasonably. That is, if the risks posed are so obvious that they would be recognized by anyone taking the time to consider them, or knowledge of the risks had to be actively avoided, then that ignorance is criminal. Beyond that, if it can be shown that the manufacturer was actually aware of the risk, then this amounts to *recklessness*. Reckless endangerment requires a mental state of foreseeing risks or possible dangers, whether to specific individuals or an uncertain public, though not explicitly intending that any potential victims actually be harmed.[3] In some cases, recklessness can also be proved by showing that a "reasonable person" should have foreseen the risks involved, even if it cannot be proven that the defendant actually had foreseen the risks. An even more severely culpable state of mind would be if the company sold the dangerous toys *knowingly*, in awareness of the fact that they would cause damages, even though they did not intend the damages. And the most severe form of culpable liability is that of having the mental intention to cause the harm, or otherwise *purposely* causing harms. While these are all cases of criminal liability, as we will see later in our discussion of corporate punishment, such cases are almost always settled by awarding punitive monetary damages to victims and their legal advocates, rather than penalties owed to the state, such as imprisoning the guilty parties.[4]

Another interesting aspect of liability is that it can be differentially apportioned. That is to say, for example, one party might be 10 percent responsible, while another

is 90 percent responsible for some harmful event. This kind of analysis of the causal chains resulting in harms is not uncommon in cases involving traffic accidents, airliner crashes, and product liability. In many jurisdictions, there are laws that separate differential causal responsibility from the consequent legal liability. Among these 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, or *strict liability*, which can hold a party fully responsible for compensation. These liability structures are meant to ensure that justice is done, in that the wronged individual is made whole (monetarily) by holding those most able to compensate them fully liable for paying all of the damages, even when they are not fully responsible for causing the harm. Nonetheless, these cases still recognize that various factors and parties contribute differentially to causing some event.

Differential apportionment could prove to be a useful tool when considering issues in robot ethics. For instance, a badly designed object recognition program 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, and so on. This implies that engineers need to think carefully about how the subsystem they are working on could interact with other subsystems—whether as designed or in partial breakdown situations—in potentially harmful ways. That, in turn, would suggest that systems engineering approaches that can manage these complex interactions would become increasingly important for consumer robotics. It also means that manufacturers will need to ensure the quality of the components they use, including software, test the ways in which components interact with each other, as well as prescribe appropriate maintenance regimes to ensure the proper functioning of those components. This is typical of complex and potentially dangerous systems, such as in airliners and industrial robots, and may prove necessary for many consumer robots, as well.

There is, however, a limit to what robot manufacturers, engineers, and designers can do to limit the potential uses of, and harms caused by, their products. This is because other parties, namely the consumers and users of robots, will choose to do various sorts of things with them and will have to assume the responsibility for those choices. For instance, when one uses a product in a manner wholly unintended by its designers and manufacturers, such as using a tent as a parachute, we no longer hold the manufacturer liable for the harms that result. Schaerer, Kelley, and Nicolescu (2009) argue that users should be held liable only in those cases in which it can be shown that they acted with harmful intentions. I disagree with this argument because of the intrinsic flexibility of design inherent in the programmability of robots. Typically, we do not hold manufacturers responsible when the hardware has been tampered with or extensively modified, or when the hardware is running software developed by users or a third party, even when there is no malice involved. We also do not always hold the company that develops a piece of software responsible when it turns out to be

vulnerable to a malicious third party, such as a hacker or virus. Again, the operative legal considerations are causal responsibility and culpable intent. However, in manufacturing a product that is programmable, and thus wildly customizable, a great deal of responsibility lies in the hands of those who do the programming, as well as those who use the robot by giving it various commands.

The challenge presented by programmable general-purpose robots is that it is unreasonable to expect their manufacturers to anticipate all the things their robots might be programmed to do or asked to do, and thus unreasonable to hold them liable for those things. At least, the less foreseeable the uses, the less responsible the manufacturer might be. But there is no clear and definitive line here. At one extreme are cases where the manufacturer ought to be held liable, and at the other extreme cases where the programmer or user ought to be held liable. At one extreme, we would find the narrowly specified applications of the robots for which its manufacturers intended the product to be used. At the other extreme, we might find a highly original custom application or program, which perhaps only that particular programmer or user might have dreamt up.

Like built-in programming, the context in which the robot has been placed and the instructions given to it by its owners may also be the determining, or contributing, causes of some harm, where the robot is the proximate cause. Orders and operator commands are like programming, in some sense, and as natural language processing grows more sophisticated, the two may become increasingly indistinguishable. And even a well-programmed robot can be ordered to do things that might unintentionally cause harms in certain situations. In short, there will always be risks inherent in the use of robots, and at some point the users will be judged to have knowingly assumed these risks in the very act of choosing to use a robot. Properly assessing responsibility in liability cases will be difficult and contested, and will depend on decisions in future cases that establish various legal precedents for interpreting such cases.

It also seems likely that robotic technologies will advance much like computer technologies, in that hackers and amateur enthusiasts will push the envelope of capabilities of new devices as much as commercial manufacturers do, especially in terms of the software and programming of robots. Even iRobot's mild-mannered Roomba vacuum-cleaning robot has a fully programmable version called Create (iRobot.com 2010), and hackers have created their own software development kits (SDKs) to customize the Roomba robot as they see fit, though at their own liability (Kurt 2006). As long as these robotic products have enough safe and legitimate uses, it would be difficult to prohibit or regulate them, just as it would be difficult to hold the manufacturers responsible for any creative, even if dangerous, uses of their products. Cars and guns are also very dangerous consumer products, but it is the users who tend to be held liable for most of the harms they cause, not the manufacturers, because the use of those potentially dangerous products place additional burdens of responsibility on

the user. For manufacturers to be held responsible in those cases, it is usually necessary to show that there is some defect in the product, or that manufacturers misled consumers.

The crucial issue raised by Schaerer, Kelley, and Nicolescu (2009) is whether to hold the manufacturer strictly liable for all the damages (because they are better able to pay compensation and to ensure responsible design), or whether their limited ability to foresee a possible application of their technology should limit their liability in some way. One implication of applying strict liability, as Schaerer, Kelley, and Nicolescu argue, is that doing so may result in consumer robots being designed by manufacturers to limit their liability by making them difficult to be reprogrammed by users, or safeguarding them from obeying commands with hazardous implications. This could include making the open-ended programming of their robots more difficult, or incorporating safety measures intended to prevent harm to humans and property, such as ethical governors (Arkin 2009).[5] Conversely, by shielding individual users from liability, this could also encourage the reckless use of robotic systems by end-users. Cars are causally involved in many unintended harms, yet it is the drivers who are typically held responsible rather than manufacturers. This issue points to a fundamental tension between identifying the causal responsibility of original manufacturers, end-users, and third-parties, and the need for legal policies that can shape the responsible design and use of consumer robots, even if they run counter to our intuitions about causal responsibility.

An additional challenge that we may soon face is determining the extent to which a given robot has the ability to act "of its own accord," either unexpectedly or according to decisions it reaches independently of any person. As robot control programs become more capable of various forms of artificial reasoning and decision-making, these reasoning systems will become more and more like the orders and commands of human operators in terms of their being causally responsible for the robots' actions, and, as such, will tend to obscure the distinction just made between manufacturers and users. While some sophisticated users may actually design their own artificial intelligence systems for their robots, most will rely on the reasoning systems that come with these robots. Thus, liability for faults in that reasoning system might still revert to the manufacturer, except in cases where it can be shown that the user trained or reprogrammed the system to behave in ways in which it was not originally designed to behave. In its general form, the question of where commands and orders arise from is integral to the legal notions of autonomy and agency. There is a growing literature addressing the question of whether robots can be capable of moral autonomy, or even legal responsibility (Wallach and Allen 2009). But missing from these discussions is the recognition that the law does not always hold morally autonomous humans fully responsible for their own actions. The notable cases include those of diminished mental capacity, involuntary actions, or when agents are following orders of a superior.

The next section will consider the possibility that even if a robot could become, in some sense, fully autonomous, then we might not be inclined to hold it legally liable for all of the harms it might cause.

11.2 Vicarious Liability, Agents, and Diminished Responsibility

There are multiple areas of the law that deal with cases in which one independent, autonomous, rational being is acting on behalf of, or in subordination to, another. Often discussed in the robotics literature are laws governing the ownership of domesticated animals; however, there are also analogous cases involving the laws governing the liability of employees and soldiers following orders, as well as historical laws governing the liability of masters for their slaves, and the harms they cause when agents are carrying out the orders of their superiors. The laws governing animals are the simpler cases, as animals are not granted legal standing, though they may be entitled to protections from abuse in many jurisdictions. More complicated are cases where a person can act either on behalf of a superior or on their own behalf, and judging a specific act as being one or the other can have differing legal implications. The area of law dealing with these three-party relationships is called *agency law*,[6] and we will consider this after first considering the legal liabilities surrounding domesticated and wild animals.

It has been recognized that robots might be treated very much like domesticated animals, in that they clearly have some capacity for autonomous action, yet we are not inclined to ascribe to them moral responsibility, or mental culpability, or the rights that we grant to a human person (Caverley 2006; Schaerer, Kelly, and Nicolescu 2009). Domesticated animals are treated as property, and as such any harms to them are treated as property damages to the owner. Because they are domesticated, they are generally seen as not being particularly dangerous if properly kept. Despite this, it is recognized that animals sometimes act on their own volition and cause harms, and so their owners can be held liable for the damages caused by their animals, even though the owners have no culpable intentions. If, however, the owners' behavior was criminally negligent, reckless, or purposeful, then the owners can be held criminally liable for the actions of their animals. For instance, it can be criminal when someone fails to keep his or her animal properly restrained, trains an animal to be vicious, orders an animal to attack, or otherwise intends for the animal to bring about a harm.

We should note that in such cases the intention of the animal is rarely relevant—it does not matter much for legal purposes whether the animal intended the harms it caused or not. Rather, it is the owner's intention that is most relevant. Moreover, in those cases where the animal's intention runs counter to the owner's intention, this can have two different consequences. In cases of domestic animals, where the animal

suddenly behaves erratically, unexpectedly, or disobeys its owner, then this tends to diminish the *mens rea* of the owner, though does not release them from liability, and often motivates the destruction of the animal. However, in cases of exotic or wild animals, such as big cats, nonhuman primates, and poisonous snakes, there is a certain presupposition of their having independent reasoning (i.e., being wild) and being more physically dangerous than domesticated animals. And with the recognition of the intrinsic danger they pose to other people, there is an additional burden of responsibility on the owner. Owning such animals has various restrictions in different states (Bornfreeusa.org 2010; Kelley et al. 2010), and the very act of owning them is recognized as putting other members of the community at risk, should the animal escape or someone accidentally happen upon them. Failing to properly keep such an animal can automatically constitute criminally reckless endangerment, based on the known dangerousness of the animal.

Such a standard might also be applied in robotics. A standard off-the-shelf robot might be considered as being like a domesticated animal, in that its manufacturer has been entrusted to design a robot that is safe to use in most common situations. However, a highly modified, custom programmed, or experimental robot might be seen as being more like a wild animal, which might act in dangerous or unexpected ways. Thus, someone who heavily modifies his or her robot, or builds a highly experimental robot, is also undertaking greater responsibility for potentially endangering the public with that robot. A good example would be someone who armed a robot with a dangerous weapon. Such an act could itself be seen as a form of reckless endangerment, subject to criminal prosecution, even if the robot did not actually harm anyone or destroy any property with the weapon. Similar principles apply in drunk-driving laws. By driving a car while drunk, an individual is putting others at risk, even if they do not actually have an accident. It is because of this increased risk that the activity is deemed criminal (as well as being codified in law as criminal). Building a robot that intentionally, knowingly, or recklessly endangers the public could be similarly viewed as a criminal activity, and laws to this effect should be established. More limited cases of negligent endangerment might be determined to be civilly or criminally liable.

With certain technologies that are known to be dangerous if misused—such as cars, planes, guns, and explosives—there are laws that regulate their ownership and use. This ensures both that the possession and use can be restricted to individuals who are trained and tested on the proper use of a technology, as well as to establish an explicit and traceable connection between a piece of technology and a responsible individual. Thus, the use of an airplane or automobile requires completing a regime of training and testing to obtain an operator's license. The ownership of a gun or explosives requires a license, which also aids in tracking individuals who might obtain such materials for illicit purposes, and in tracking the materials themselves. The ownership

of dangerous exotic animals, and in many jurisdictions even certain particularly aggressive domesticated dog breeds, such as pit bulls, often requires a special license (Wikipedia.org 2010). It would not be unreasonable to expect that certain classes of robots, especially those that are deemed dangerous, either physically or because of their unpredictable behavior or experimental nature of their reasoning systems, might require special licenses to own and operate. Licenses might also be required to prevent children from being able to command dangerous robots, just as they are not allowed to drive cars, until they have reached a certain age and received training in the responsible uses of the technology.

The treatment of robots as animals is appealing because it does not require us to give any special rights or considerations to the robots—our only concern is with the owners. Another interesting area of legal history is the laws governing slavery. The history of these laws goes back to ancient Rome, and they have varied greatly in different times, places, and cultures, up to and including the slave laws of the United States. The U.S. slave laws ultimately treated slaves as property, but included numerous specialized clauses intended to manage the unique difficulties and dangers of enslaving human beings, as well as encoded specific racial aspects of slavery into the laws themselves. For the most part, slaves were treated as expensive animals, so that if a slave damaged the property of someone other than his or her owner, their owner was liable for the damage. Similarly, as property, the slave was protected from harm from individuals other than his or her owner, but such harms were viewed as property damage rather than crimes, such as assault or even murder. Indeed, the laws largely enshrined the ability of owners to harm their own slaves and not be subject to the criminal laws that might otherwise apply. Yet, it was also recognized that slaves exercised a will of their own, and so owners were not held liable for damages caused by their slaves if they had escaped. And, unlike animals, in the act of escaping, slaves were held liable for their own choice of whether to escape or not, though those who aided them were also held liable for assisting them in their escape. A full consideration of the implications of slave law for our understanding of robotics is beyond the scope of this chapter, but will be the subject of further research.

Agency law deals with cases in which one individual acts on behalf of another individual. In these cases, the *agent* acts on behalf of a *principal*. There are various circumstances where these relationships are established, but they generally involve some form of employment, often involving a contract.[7] Whether or not there is a written contract, the liability of the principal for the actions of its agents is derived from the doctrine of *respondeat superior*—that superiors are responsible for the action of their subordinates. Thus, if an employee causes a harm in the conduct of their job, and thus explicitly or implicitly at the discretion of their employer, the employer is liable. This is called *vicarious liability*—when one person or legal entity is liable for the

actions of another. For instance, when a delivery truck damages a parked car, the delivery company, rather than the individual driver, can be held liable. There are exceptions to this, however, which recognize the independent autonomy of employees. One of the employer's defenses against such liability claims is to argue that the employee was acting on their own behalf, and not that of the employer—which generally means showing they were not doing their job in the typical manner. Courts make a distinction between *detours*, in which an employee must digress from the usual manner of carrying out their job in order to achieve the purposes of their employer, and *frolics*, in which the employee is acting solely for their own purposes. Thus, when a driver finds an intersection blocked and must take a different route to make a delivery, the employer would still be liable for the damage to the parked car. However, if the driver had decided to take a different route in order to visit a friend before making the delivery, then the court may decide that this constitutes a frolic and the driver is responsible for the damage to the parked car because the driver was not carrying out their duties as an employee, or fulfilling the will or purpose of the employer, at the time of the accident.

These ideas might be usefully applied to many kinds of service robots. It would seem that, for most uses of a robot to assist a person in daily life, such as driving them around, shopping for them, cleaning and maintaining their home, running errands, and so on, the robot would be little different than a human servant or employee. As such, vicarious liability would apply, and the owner would therefore be liable for any harm caused by that robot in the conduct of their owner's business. This would also include cases of detour, in which the robot was unable to carry out its duties in the normal or directed manner, and sought alternative routes, plans, or strategies for achieving its given goals.

As robots grow more sophisticated and autonomous, we might eventually be tempted to argue that they actually are capable of developing their own purposes. For such a robot, an owner might seek a defense from liability for the actions of a robot which was on a frolic of its own—a robot which, though employed in the service of its owner, caused some harm while pursuing its own purpose. Depending on the ways in which such robots might be programmed, and our ability to review its reasoning and planning processes, we might actually be able to determine from its internal records which purposes it was actually in pursuit of when it caused a particular harm. Of course, it might be that it was pursuing a dual purpose, or that the purposes were obscure, in which case the courts would have to make this determination in much the same manner as they do for human agents. However, this raises several issues regarding whether robots might themselves have legal standing, especially if they are capable of frolicking, or whether they might be subject to penalties and punishments, and it is to these issues that we now turn.

11.3 Rights, Personhood, and Diminished Responsibility

Modern legal systems were established on the 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, over time the law has managed to deal with several kinds of nonpersons, or quasi-persons, and we can look to these for some insights on how we might treat robots that are nonpersons, or quasi-persons. Personhood is a hotly debated concept, and many perspectives in that debate are based in strongly held beliefs from religious faith and philosophical dispositions. Most notably, the status of unborn human fetuses, and the status of severely brain damaged and comatose individuals have led to much debate in the United States over their appropriate legal consideration and rights. Yet, despite strongly differing perspectives on such issues, the legal systems in pluralistic societies have found ways to deal practically with these and several other borderline 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 do not have full legal standing. Of course, the killing of a child is murder in the same way that the killing of an adult is, and so a child is still a legal person in this sense—and, in fact, is entitled to more protections than an adult. 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 full-fledged legal persons can claim temporary impairment of judgment, and thereby diminished responsibility for their actions, given certain circumstances, for example, temporary insanity, or being involuntarily drugged. The point is that some aspects of legal agency can apply to entities that fall short of full-fledged personhood and full responsibility, and it seems reasonable to think that some robots will eventually be granted this kind of quasi-agency in the eyes of the law before they achieve full legal personhood.

11.4 Crime, Punishment, and Personhood in Corporations and Robots

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 this distinction: 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, by "the people," and the debt owed by the wrongdoer is owed to the society. While the punishments may take 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 (retribution); 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 fundamental problems with applying criminal law to robots: (1) criminal actions require a moral agent to perform them, and (2) it is not clear that it is possible to punish a robot. While moral agency is not essential to civil law, moral agency is essential to criminal law, and is deeply connected to our concepts of punishment (retribution, reform, and deterrence). Moral agency might be defined in various ways, but, in criminal law, it ultimately must serve the role of being an autonomous subject who has a culpable mind, and who can be punished. Without moral agency, there can be harm (and hence civil liability), 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 or act of nature, but 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 are capable of recognizing the similarity of their potential choices and actions to those of other moral agents who have been punished for the wrong choices and actions—without this reflexivity of choice by a moral agent, and recognition of similarity between and among 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. Solum (1991–1992) 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 personhood is not impossible in principle for an AI to achieve, 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 a court's determination whether an AI could stand trial as a legal agent in its own right, and not merely as a proxy or agent of some other legal entity. He argues 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—in other words, could the court fine or imprison an AI that mismanages a trust? In cases of quasi-personhood and diminished responsibility, however, children and the mentally impaired are usually shielded from punishment as a result of their limited legal status, specifically because they lack proper moral agency.

There is another relevant example in law of legal responsibility resting in a nonhuman entity, namely corporations. The corporation is a nonhuman entity that has been effectively granted many of the legal rights and responsibilities of a person. Corporations

can (through the actions of their agents) 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 seventeenth-century Lord Chancellor of England put it, corporations have "no soul to damn and no body to kick" (Coffee 1981), so how can they be expected to have a moral 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 antitrust violations, split apart, or dissolved altogether. Though they cannot be imprisoned in criminal cases, responsible agents within the corporation can be prosecuted for their individual actions. As a result of this, and other aspects of corporations being complex sociotechnical systems in which there are many stakeholders with different relations to the monetary wealth of a corporation, it can be difficult to assign a punishment that achieves retribution, reform, and deterrence, while meeting other requirements of justice, such as fairness and proportionality.[8]

Clearly, robots are different in many important respects from corporations. However, there are also many important similarities, and it is no coincidence that Coffee's (1981) seminal paper on corporate punishment draws heavily on Simon's (1947) work on organizational behavior and decision making, and in particular how corporate punishment could influence organizational decision making through deterrence. 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 purpose for the existence of corporations—to make money. The essential purposes of robots may not be so straightforward, and, if they exist at all, they 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 from corporations 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 psychological desires and fears central to being human that may not readily apply to robots—concerning pain, freedom of movement, mortality, and so on. Thus, torture, humiliation, imprisonment, and death are not likely to be effective in achieving retribution, reform, or deterrence in robots. There could be a policy to destroy any robots that do harm, but, as is the case with animals that harm people, it would essentially be a preventative measure to avoid future harms by an individual, 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. In short, there is little sense in trying to apply our traditional notions of punishment to robots directly. This appears to me to be a greater hurdle to ascribing moral agency to robots than other hurdles, such as whether it is possible to effectively program moral decision making.

11.5 Conclusion

I hope that this brief overview of how certain legal concepts might be applied to current and future robots has convinced the reader 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 of problems will remain unresolved. In terms of thinking about robots as manufactured products, many of the most practical and pressing issues facing robotics engineers can be seen as being essentially like those facing other engineers. In 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 and research. It is my belief that robot ethics will have its greatest influence by seeking to define and establish expectations and standards of practice for the robotics industry.

There remain a host of metaethical questions facing robot ethics that lie beyond the scope of 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 construction of truly autonomous technologies demand special consideration in their own right. While there was no room to discuss it in this chapter, the legal perspective can also contribute to framing issues in the use of robots in warfare, though it offers little in the way of determining what social values we should aspire to enshrine in the laws governing the use of lethal robots. In particular, international law, humanitarian law, uniform codes of military conduct, the Geneva Conventions, the Nuremberg Principles, and international laws banning antipersonnel mines and limiting biological, chemical, and nuclear weapons, are all starting points for theorizing the ethics of using robot technologies in warfare, but may fall short in suggesting new standards for the ethical conduct of the kind of warfare that robots might make possible.

Notes

1. In the system of Anglo-American law, a distinction is drawn between criminal and civil law, and within civil law there is a further distinction between the laws of torts and contracts. Tort law deals with property rights and infringements outside of, or in addition to, contractual obligations and crimes, and is primarily concerned with damage to one's person or property and other harms. Thus, one has the right to sue responsible parties for damages that one has suffered, even

if one is not engaged in an explicit legal contract with the other party, and in addition to or regardless of whether the other party also committed a criminal act when causing the damages in question. Tort law seeks justice by compelling wrongdoers to compensate, or "make whole," those who were harmed for their loss (Prosser et al. 1984). Criminal law deals with what we tend to think of as moral wrongdoing or offenses against society, such as theft, assault, murder, etc., and seeks justice by punishing the wrongdoer.

There are several crucial differences between the concepts of criminal damages and civil damages and their accordant penalties. Most generally, for something to be a crime, there must be a law that explicitly stipulates the act in question as being criminal, whereas civil damages can result from a broad range of acts, or even inaction, and need not be explicitly specified in written law. Criminal acts are usually distinguished by their having criminal intent—a culpable state of mind in the individual committing the crime, known in Latin as *mens rea*. While certain forms of negligence can rise to the level of criminality, and can be characterized as nonmental states of ignorance, judgments of criminality typically consider mental states explicitly. Civil law, in comparison, is often indifferent to the mental states of the agents involved. And finally, there are differences in the exactment of punishments for transgressions. Under civil law, the damages are repaired by a transfer of money or property from the liable transgressor to the victim, while in criminal law the debt of the guilty transgressor is owed to the general public at large or the state, for the transgressor has violated the common good. A criminal penalty owed to society need not be evaluated in monetary terms, but might instead be measured in the revocation of liberty within society, expulsion from society, and in cultures of corporeal punishment, the revocation of bodily integrity or life, or the infliction of pain, humiliation, and suffering. In some instances, both frameworks apply, and criminal penalties may be owed over and above the restorative monetary damages owed to the victim of a crime.

2. For more on product liability law, see chapter 17 of Prosser et al. 1984.

3. It is in this way very much like the *doctrine of double effect* in Just War Theory (Walzer 1977), in that it separates knowledge of the possible harms of one's actions from the intention to actually bring those harms about. According to the doctrine of double effect, the killing of innocent civilians is permissible if the intended effect is on a militarily valid target, whereas the killing of civilians is not permissible if the intended effect is actually to harm the civilians.

4. For more on criminal negligence and liability, see chapter 5 of Prosser et al. 1984.

5. This notion is also popular in science fiction, starting with Isaac Asimov's "Three Laws of Robotics" (later four), and the "restraining bolts" in *Star Wars* droids, all of which aim to prevent robots from doing harm, despite maintaining their willingness to obey orders.

6. For more on the legal theory of agency, see Gregory 2001.

7. The principal can also be a corporation, in which case it is unable to act without its agents, which raises certain issues for corporate punishment, as we will see.

8. As Coffee (1981) argues, typical monetary fines against a company hurt shareholders and low-level employees more directly than they hurt the managers and decision makers in a company, which diminishes their ability to deter or reform those who made the bad decisions and thus the fairness of imposing such fines.

Acknowledgments

I would like to thank Lawrence Solum for his thoughts on an early draft of this chapter, Malcolm MacIver for his thoughtful questions and comments, and special thanks go to John Evers for his numerous helpful discussions and legal insights on these topics.

References

Arkin, Ronald. 2009. *Governing Lethal Behavior in Autonomous Robots*. New York: CRC Press.

Asaro, Peter. 2006. What should we want from a robot ethic? *International Review of Information Ethics* 6 (12): 9–16.

Asaro, Peter. 2007. Robots and responsibility from a legal perspective. <http://peterasaro.org/writing/ASARO%20Legal%20Perspective.pdf> (accessed July 14, 2011).

Asaro, Peter. 2011. Remote-control crimes: Roboethics and legal jurisdictions of tele-agency. *IEEE Robotics and Automation Magazine* 18 (1): 68–71.

Bechtel, William. 1985. Attributing responsibility to computer systems. *Metaphilosophy* 16 (4): 296–306.

Bornfreeusa.org. 2010. Exotic animals summary. <http://www.bornfreeusa.org/b4a2_exotic_animals_summary.php> (accessed September 11, 2010).

Caverley, Daniel. 2006. Android science and animal rights: Does an analogy exist? *Connection Science* 18 (4): 403–417.

Coffee, John. 1981. "No soul to damn: No body to kick": An unscandalized inquiry into the problem of corporate punishment. *Michigan Law Review* 79 (3): 386–459.

Dennett, Daniel. 1997. When HAL kills, who's to blame?: Computer ethics. In *HAL's Legacy: 2001's Computer as Dream and Reality*, ed. D. G. Stork, 351–365. Cambridge, MA: MIT Press.

Gregory, William. 2001. *The Law of Agency and Partnership*, 3rd ed. New York: West Group.

iRobot.com. 2010. <http://store.irobot.com/shop/index.jsp?categoryId=3311368> (accessed September 11, 2010).

Kelley, Richard, Enrique Schaerer, Micaela Gomez, and Monica Nicolescu. 2010. *Advanced Robotics* 24 (13): 1861–1871.

Kurt, Tod E. 2006. *Hacking Roomba: ExtremeTech*. New York: Wiley. See also <http://hackingroomba.com/code/> (accessed November 26, 2010).

Mitcham, Carl. 1990. Ethics in bioengineering. *Journal of Business Ethics* 9 (3): 227–231.

Moon, Y., and Clifford Nass. 1998. Are computers scapegoats? Attributions of responsibility in human-computer interaction. *International Journal of Human-Computer Studies* 49 (1): 79–94.

Prosser, William Lloyd, and W. Page Keeton, Dan B. Owens, Robert E. Keeton, and David G. Owen, eds. 1984. *Prosser and Keeton on Torts*. 5th ed. New York: West Group.

Schaerer, Enrique, Richard Kelley, and Monica Nicolescu. 2009. Robots as animals: A framework for liability and responsibility in human-robot interactions. *Proceedings of RO-MAN 2009: The 18th IEEE International Symposium on Robot and Human Interactive Communication*, Toyama, Japan, Sept. 27–Oct. 2, 72–77. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5326244> (accessed July 14, 2011).

Simon, Herbert. 1947. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*. New York: Free Press.

Solum, Lawrence. 1991–1992. Legal personhood for artificial intelligences. *North Carolina Law Review* (April): 1231–1287.

Wallach, Wendell, and Colin Allen. 2009. *Moral Machines: Teaching Robots Right from Wrong*. Oxford, UK: Oxford University Press.

Walzer, Michael. 1977. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*. New York: Basic Books.

Wikipedia.org. 2010. Breed-specific legislation. <http://en.wikipedia.org/wiki/Breed-specific_legislation> (accessed November 26, 2010).

12 Robots and Privacy

M. Ryan Calo

Robots are commonplace today in factories and on battlefields. The consumer market for robots is rapidly catching up. A worldwide survey of robots by the United Nations in 2006 revealed 3.8 million in operation, 2.9 million of which were for personal or service use. By 2007, there were 4.1 million robots working just in people's homes (Singer 2009, 7–8; Sharkey 2008, 3). Microsoft founder Bill Gates has gone so far as to argue in an opinion piece that we are at the point now with personal robots that we were in the 1970s with personal computers, of which there are many billions today (Gates 2007). As these sophisticated machines become more prevalent—as robots leave the factory floor and battlefield and enter the public and private sphere in meaningful numbers—society will shift in unanticipated ways. This chapter explores how the mainstreaming of robots might specifically affect privacy.[1]

It is not hard to imagine why robots raise privacy concerns. Practically by definition, robots are equipped with the ability to sense, process, and record the world around them (Denning et al. 2009; Singer 2009, 67).[2] Robots can go places humans cannot go, see things humans cannot see. Robots are, first and foremost, a human instrument. And, after industrial manufacturing, the principle use to which we've put that instrument has been surveillance.

Yet increasing the power to observe is just one of ways in which robots may implicate privacy within the next decade. This chapter breaks the effects of robots on privacy into three categories—direct surveillance, increased access, and social meaning—with the goal of introducing the reader to a wide variety of issues. Where possible, the chapter points toward ways in which we might mitigate or redress the potential impact of robots on privacy, but acknowledges that, in some cases, redress will be difficult under the current state of privacy law.

As stated, the clearest way in which robots implicate privacy is that they greatly facilitate *direct surveillance*. Robots of all shapes and sizes, equipped with an array of sophisticated sensors and processors, greatly magnify the human capacity to observe. The military and law enforcement have already begun to scale up reliance on robotic technology to better monitor foreign and domestic populations. But robots also