

Killers, fridges, and slaves: a legal journey in robotics

Ugo Pagallo

Received: 17 November 2009 / Accepted: 5 December 2010 / Published online: 22 January 2011
© Springer-Verlag London Limited 2011

Abstract This paper adopts a legal perspective to counter some exaggerations of today's debate on the social understanding of robotics. According to a long and well-established tradition, there is in fact a relative strong consensus among lawyers about some key notions as, say, agency and liability in the current use of robots. However, dealing with a field in rapid evolution, we need to rethink some basic tenets of the contemporary legal framework. In particular, time has come for lawyers to acknowledge that some acts of robots should be considered as a new source of legal responsibility for *others'* behaviour.

Keywords Agency · Legal systems · Liability · Moral accountability · Responsibility · Robots

1 Introduction

Robotics is the area of AI concerned with the use of robots, that is, according to George A. Bekey, machines that “sense, think, and act.” (Bekey 2005)

This field is interdisciplinary par excellence, involving not only artificial intelligence and computer science, but also cybernetics, physics, mathematics, mechanics, electronics, neuroscience, biology, and humanities. Moreover, it comprises a number of different applications that the European Robotics Research Network has classified in eight types, including humanoids, adaptive service robots,

network and outdoor robotics, edutainment, and so forth (Veruggio 2007).

Notwithstanding this complexity, the social understanding of robotics may be summed up with three metaphors or images, two of which suggested by the title of this special issue of AI & Society.

The first metaphor of robots as killers is illustrated by Asimov's warnings in his novels and the first law of robotics, according to which no human being may be injured by robots (what frequently happens in Asimov's stories). One of the most relevant and developed areas of today's robotics, after all, concerns military applications like intelligent weapons, robot soldiers, and even super-human soldiers with sensor systems, augmented reality-devices, or exoskeletons. Therefore, the question does not hinge simply on the possibility of having robots that gun people down as, say, when Robbie CX30 assassinated Bart Matthews in Richard Epstein's story on *The case of the killer robot* (1997). The issue is whether we should admit the moral responsibility of robots in homicides.

The second image is that of ‘fridges.’ Scholars often use this metaphor to prevent some overstatements in the current debate, such as robot killers and other artificial agents outsmarting humans, so that we, as a species, would soon face extinction since intelligent robots will replace us as the next step in evolution (Moravec 1999). Against this sort of techno-determinism and other science fiction speculations, the fridge-metaphor proposes a more sober picture of robots and their peculiar autonomy: “They can deal successfully with their tasks, even if they have the intelligence of a refrigerator.” (Floridi 2007)

Finally, we find the image of robots as ‘slaves.’ Similarly to the way in which slaves were disciplined by Ancient Roman law, we may deem robots as autonomous agents that are however merely ‘things,’ eventually lacking

U. Pagallo (✉)
Law School, University of Turin, Turin, Italy
e-mail: ugo.pagallo@unito.it
URL: http://www.dsg.unito.it/Segreteria/personale/pagallo_c.htm

rights and duties. This third parallelism sheds light on a new form of agency, in that, “like a slave, [the robot] is capable of making decisions which will affect the rights (and, in later law, the liabilities) of its master. By facilitating commercial transactions, autonomous agents have the ability to increase market efficiency. Like a slave, an autonomous agent is capable of doing harm.” (Katz 2008)

In order to assess these metaphors, I examine the social understanding of robots by adopting a legal perspective. A long and well-established tradition on this subject allows me to counter and thus to avoid some misunderstandings of the current debate. Apart from Leibniz’s seminal remarks on law and machines, the “Law of the Automaton” was already a quite popular topic among German scholars in the late 1800s: It suffices to recall Günther’s *Das Automatenrecht* from 1891, Schiller’s *Rechtsverhältnisse des Automaten* from 1898, or Neumond’s *Der Automat* from 1899. More than a century later, a relatively strong consensus still exists about key notions like legal liability and agency in the use of robots. Although it might be meaningless to sue robots for murder, it makes a lot of sense to think about them as ‘intelligent’ fridges. So, what about the possibility to hold robots as similar to slaves in the near future?

All in all, we are dealing with a field which is rapidly progressing and that, furthermore, questions some basic tenets of today’s legal framework. Maybe it is too early to envisage personal responsibility of robots in criminal affairs, as both in the example of liability for sexual offences put forward by Fernando Barrio (2008), and the hypotheses of “the robot kleptomaniac” and “picciotto roboto” illustrated by Reynolds and Ishikawa (2007).

However, time has come to seriously acknowledge that robots’ behaviour should be legally considered as a new source of personal responsibility for *others’* acts (e.g., tort law and vicarious liability in the common law tradition and its counterpart in civil law, that is, ‘objective responsibility’ or liability without fault).

Hence, the paper aims at examining this specific field of the social understanding of robotics, by providing a concise legal journey in four sections.

First of all, I analyse how robots have stimulated scholarly debate on some basic fields of jurisprudence and legal theory. More particularly, I consider the case of hermeneutics and the above-mentioned laws of robotics in Isaac Asimov’s work.

Secondly, I illustrate the current state of the art in legal science: Special attention is paid to the notions of agency and civil (as opposed to criminal) liability.

Thirdly, I explain the reasons why I think we should add a new form of legal responsibility for *others’* behaviour: Robots, just like slaves, are reckoned ‘things’ with significant ‘autonomy’ and, perhaps, some specific duties.

The result is a more accurate definition of ‘agency’ which appears to be one of the most relevant aspects related to the social understanding of robotics. Along with animals and fellow human beings, we will be forced to cope with another source of im/moral behaviour.

2 Legal analogies

There are two ways in which we may interpret the notion of legal analogy in this context.

On the one hand, the role of some metaphors is to grasp a number of legal issues *as if* robots were killers, slaves, and the like, in order to comply with the dogma of civil law as a self-referential system in which analogy may fill its normative gaps. This was the method already used by German scholars in the late 1800s, when tackling the novelty of automatons in legal systems: What if a robot causes harm to a human being?

On the other hand, metaphors can be used to clarify some typical queries concerning the field that scholars usually refer to as jurisprudence (in the common law tradition) or as ‘general theory of law’ (in the civil law tradition). For example, assuming Asimov’s laws of robotics, it is possible to illustrate some classical legal topics that, according to Herbert Hart’s (1961) tripartite approach, concern fundamental legal concepts, legal reasoning, and ethical problems involving the law. Let me start with the quotation of Asimov’s laws in *Runaround* (1942):

1. “A robot may not harm a human being, or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given to it by human beings, except where such orders would conflict with First Law.
3. A robot must protect its own existence, as long as such protection does not conflict with the First or Second Law.”

Then, in *Robots and Empire* (1985), Asimov added a fourth law, the ‘Zeroth’ law:

“0. A robot may not injure humanity, or, through inaction, allow humanity to come to harm.”

By going back to Hart, we can understand these laws in connection with the nature of ‘legal concepts’ and the traditional idea of natural law. As stressed by Paolo Comanducci (2006), the link between laws of robotics and natural law suggests that we reconsider the meaning of legal commands, since natural law was meant to guide our actions in the same way in which the laws of robotics would direct the behaviour of robots. In both cases, law can be viewed as an objective imperative whose infringement would imply a violation of the nature of the agent.

Furthermore, as in the case of natural law, the laws of robotics propose matters of normative hierarchy and legal antinomies as well: A good illustration is offered by Roger Clarke's work on *Asimov's Laws of Robotics* (Clarke 1993), where he indicates various additional implicit laws with which to fill the gaps of Asimov's normative system. In particular, the First Law of Robotics should be integrated by a meta-law, which determines that "a robot may not act unless its actions are subject to the Laws of Robotics." Then, a second section is added to the second law in that "a robot must obey orders given it by superordinate robots." Besides, a new first section should be inserted in the third law, and so forth.

Ethical problems involving the law should be added to the list of possible parallelisms between the laws of robotics and legal theory. We can further raise questions on free will as the basis of agents' responsibility, matters of robot paternalism and conscientious objectors, down to the minimum content of natural law, that is, "the minimum set of principles which, because rationally necessitated—given certain fundamental 'truisms' about human nature and the human predicament—for the securing of purposes shared by all survivable human societies, can be called natural law." (Hart 1961)

However, one of the most relevant and debated issues involves the interpretation of these laws, i.e., both of robotics and of legal systems. For the sake of conciseness, I shall make the point with three popular arguments of traditional legal hermeneutics.

First, the specific property of the laws of robotics, namely their abstract and general nature, entails the difficult task of applying these laws to a given context: Do the circumstances of the context affect how we interpret those general rules?

Secondly, the vagueness of ordinary language, as in the case of crucial terms like 'harm' or 'order,' jeopardizes the possibility of ensuring mechanical observance of the rules: Would it be feasible to develop computable models so as to comprise not only legal norms and concepts but also legal agents?

Finally, proper understanding of the law is characterized by several sets of criteria for interpreting the laws of the system. As exemplified by Asimov's work, robots used to adopt a sort of literal reading in his first novels: Only the extremely more sophisticated robots of the later stories began to employ complex hermeneutical techniques such as strict or extensive interpretations of the laws, evolutionary and teleological readings of the texts, and so forth. Adapting the well-known example of the rule which bans vehicles from a park (Hart 1961), let us suppose a supermarket prohibiting the entrance of pets: What should we think of this norm? Does this rule forbid me to bring along my snake?

Obviously, given the over-articulate approach of legal theoreticians, the final (and fatal) risk of all these abstract analyses is to end up in paralysis. Somehow, what scholars incessantly debate in terms of mechanical versus holistic configurations of the system, literal versus contextual reading of the texts, analytical versus systematic comprehension of the norms may lead to the outcome of some Asimov's plots, i.e., the deadlock of his robots' positronic brains.

Therefore, in the next section, I explain how lawyers can deal with robot-related issues in a more practical way.

3 The state of the art in law

Current legal debate on robots can be divided into two main areas: Criminal law and private or civil law. Although lawyers assume that criminal law, strictly speaking, lies beyond legal robotics, it is instructive to grasp the various reasons why homicides or robberies would necessarily involve only the responsibility of human beings. After examining such reasons in the next Sect. 3.1, I dwell on the more orthodox topic of robots and their role in contractual and extra-contractual obligations between humans (Sect. 3.2).

3.1 Robots and the criminal law

The common legal standpoint excludes robots from any kind of criminal responsibility because they lack psychological components such as intentions or consciousness, i.e., the set of preconditions for attributing liability to someone in the case of violation of criminal laws like murdering or stealing. Robots would not indeed be epistemically aware of their own behaviour like 'wishing' or 'wanting' to act in a certain way. Although many legal systems provide for the criminal responsibility and agency of some 'artificial' persons like companies, organizations, governments, or corporations, their liability is always reducible to an aggregation of human beings as the only relevant source of their action.

However, it is highly debatable to claim that robots lack all types of agenthood: After all, we already have proper 'artificial agents' that are interactive, autonomous, and adaptable. According to the criteria pointed out by Allen et al. (2000) in their work on the status of artificial moral agents, robots respond to stimuli by changing the values of their own properties or inner states and, what is more, they are able both to modify these states without external stimuli and to improve the rules through which those very states change. These capabilities do not only imply the unpredictability of robots' behaviour as well as their ability to act on human behalf. This scenario suggests that we are

dealing with a new source of moral agency, since robots, like animals, children and, obviously, adult human beings, can cause morally qualifiable actions as good and evil (McFarland 2008). Consequently, robots may represent a meaningful target of human censorship. We can, for instance, do the following things: “(a) monitoring and modification (i.e., ‘maintenance’); (b) removal to a disconnected component of Cyberspace; (c) annihilation from Cyberspace (deletion without backup).” (Floridi and Sanders 2005, 24)

Yet, by extending the class of morally accountable agents so as to include the artificial agency of robots, we need not admit either their moral responsibility or their criminal liability. As in the cases of children’s actions or the behaviour of animals, the reason hinges on the need of differentiating the source of relevant moral actions from the evaluation of agents as being morally responsible for a certain behaviour. This is why Floridi and Sanders, who acknowledge the moral accountability of artificial agents, readily concede “that it would be ridiculous to praise or blame an artificial agent for its behaviour or charge it with a moral accusation.” (*op. cit.*, 17) Once we can tell moral accountability from moral responsibility, we can address Daniel Dennett’s question on “who’s to blame, when HAL kills?” (1997, 351), by saying “that HAL is accountable—though not responsible—if it meets the conditions defining agenthood.” (Floridi and Sanders 2005, 26)

Consequently, according to the current state of the art in criminal law, it would be pointless to debate before a judge whether or not a robot should be considered a ‘killer,’ a ‘robber,’ and so forth. Even if we assume that some sort of moral accountability is a necessary requirement for legal responsibility, the former does not represent the sufficient condition of the latter, because respondents ought to be subject to the ordinary process of moral appreciation in order to determine whether or not they are guilty in the name of the law. Otherwise, by blurring accountability and responsibility, we would be forced back to the days when criminal trials were commonly performed against animals and even lifeless things (Ewald 1995). Indeed, the reasons underpinning the legitimacy of inflicting punishment in modern criminal law such as the theory of retribution, of special and general prevention, would be devoid of meaning. Can we reckon robots paying their debt to society? Can we correct their moral character so that robots fully understand why they ought not to repeat an evil action? Should we punish them so as to dissuade human beings from committing similar wrongs?

Hence, as long as lawyers believe that homicides and other criminal matters necessarily presuppose the responsibility of human beings, the state of the art in legal science thinks of robots not as ‘killers’ or ‘robbers.’ Rather, its mental image is that “robots are completely unremarkable

technological artifacts, no different than toasters or cars.” (Asaro 2007, 2) Let us now examine how far such an idea reaches.

3.2 Robots and the private law

Today’s legal robotics are relevant in the realm of law that belongs to the private law field, namely contractual and extra-contractual obligations (in this context, I use the expression ‘private law’ and not ‘civil law’ as opposed to ‘criminal law’ in order to prevent misunderstandings when comparing the ‘common law’ to the ‘civil law’ tradition).

I should also add that, while examining some aspects regarding the production and use of robots, I set strict contractual obligations aside, because conditions, terms, and clauses depend both on the voluntary agreement between private persons that a court will enforce, and on the commercial or non-commercial nature of that agreement: It is enough to stress the difference between surgical robotics and robot toys, between industrial and military robotics, and so forth.

Beyond the panoply of strict contractual obligations, we find what an English or US lawyer would define as torts, that is, “obligations between private persons that are imposed by Government” to compensate damage done by wrongdoing (Abernathy 2006, 646) More particularly, we can distinguish three types of liability, that is, liability for harm caused by intentional torts, negligence-related tortious liability and liability without fault (strict liability).

First, liability for an intentional tort is established when a person has voluntarily performed the wrongful action.

Secondly, there is liability based on lack of due care when the ‘reasonable’ person fails to guard against ‘foreseeable’ harm.

Finally, strict liability is established without fault as in the paradigmatic case of liability for defective products, which may be due to a lack of information about certain features of the artefact. This is why you have seen those extremely detailed and sometimes strange labels on products, by which manufacturers warn about risks or dangers involving the improper use of the artefact.

Regardless of the kind of tort we are coping with, “when multiple parties or multiple events have set in motion a chain of events that leads to the plaintiff’s harm (...) American and British courts use the doctrine of ‘proximate cause’ (sometimes called ‘legal cause’) to guide juries in deciding where to cut the chain of responsibility” (Abernathy 2006, 655). In some cases, such a liability can be apportioned because of contributory negligence, namely when the plaintiff’s negligent conduct has contributed to her own injuries, or when there are two or more tortfeasors liable for the plaintiff’s injuries.

Compared to this traditional legal framework, there are two main reasons why robots raise new problems, implying that they should not be considered simple ‘enhanced toasters.’

On the one hand, these artefacts are (and will increasingly be) capable to learn the features of their surrounding environment and of the living beings who inhabit it, while gaining knowledge or skills from their own behaviour. This capability means that robots will not only be unpredictable for their users but for their human designers as well. “So, without necessarily imagining some Sci-Fi scenarios where robots are provided with consciousness, free will and emotions, in a few years we are going to cohabit with robots endowed with self knowledge and autonomy—in the engineering meaning of these words.” (Veruggio 2007, 27)

On the other hand, it is likely that autonomous robots will create new forms of legal agency, that is, the relationship by which a party grants authority for another to act on her behalf so as to deal with a third party. “Accordingly, the legal responsibility for the actions of a robot falls on the individual who grants the robot permission to act on their behalf. (...) 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.” (Asaro 2007, 3)

The result is that both the unpredictability of robots’ behaviour and their capability to act on human behalf call for a rethinking of the traditional legal framework. If robots can hardly become ‘killers,’ it is also unlikely we can go on conceiving them as simple ‘fridges.’ Let us examine the third image of robots as possible ‘legal slaves.’

4 A step further in legal robotics

Taking the autonomy of artificial agents seriously, some scholars have proposed a suggestive parallelism between robots and slaves insofar as the jurists of Ancient Rome would have anticipated many of today’s issues involving artificial agents and robots, “by defining an advanced legal framework to cover the rights and obligations arising from slave ownership.” (Katz 2008, 2)

This does not mean, of course, that we should treat current artificial agents or tomorrow’s intelligent robots as if they were our modern slaves. Even though we reject the idea that robots represent a new source of moral agency (see above 3.1), legal systems provide for a number of sanctions in the case of intentional misuse of power, vandalism, etc., so that robots, as informational objects, could properly be reckoned moral patients that deserve respect and protection as such (Pagallo 2010). In the terms of Terrell Bynum’s account of information ethics, evil

appears “as everything that damages or impoverishes” the informational nature of the universe (Bynum 2006, 17). Thus, in the hypotheses of humans who unjustly damage or destroy their own artificial companions, we may envisage forms of prosecution in order to preserve consistency between robots and their owners.

Putting aside the ethical aspects concerning new possible reasons for charging humans, the parallelism between robots and slaves casts light on two further fields of the foreseeable future of legal robotics.

First, we have to examine how the law can cope with both the enforcement of rights and obligations created by robots and the question of liability for damages caused by them.

Secondly, we need to widen our perspective so as to take into account the very possibility that robots will soon represent a new source of personal responsibility for others’ behaviour.

Let us take a brief look at both of these aspects involving the parallelism between robots and slaves. In the next section, I consider the realm of contractual obligations (Sect. 4.1); then, the field of extra-contractual responsibility (Sect. 4.2).

4.1 Robots and contractual obligations

The first reason for comparing the status of robots with that of slaves in Ancient Rome is that slaves were considered ‘things’ that, nevertheless, played a crucial role both in trade and in commerce: The élite, as in the paradigmatic case of the emperor’s slaves, were estate managers, bankers, and merchants. They had the legal capacity to enter into binding contracts, to represent their masters, to hold important jobs as public servants or for their masters’ family business, to amass, manage, and make use of property. Although most of the slaves certainly had no rights to claim against their own masters, some slaves enjoyed a significant “autonomy” (Štaerman and Trofimova 1975, 53). Consequently, by considering how today’s artificial agents negotiate, make contracts, establish rights and obligations between humans, is there something we could learn from Ancient Roman law?

From this point of view, one of the most interesting mechanisms provided by Roman law is the ‘peculium.’ In the phrasing of the Digest of Justinian, it is “the sum of money or property granted by the head of the household to a slave or son-in-power. Although considered for some purposes as a separate unit, and so allowing a business run by slaves to be used almost as a limited company, it remained technically the property of the head of the household.” (Watson 1988, xxxv-xxxvi)

As a sort of proto-limited company, the peculium aimed at striking a balance between the claim of the masters not

to be ruined by their slaves' businesses and commercial activities, and the interest of the slaves' counterparties to safely transact with them. While, most of the time, masters' liability was limited to the very value of their slaves' peculium, the legal security of the latter guaranteed slaves' counterparties that obligations would have been met. Thereby, depending on both the kind of activities and status of the slaves as *dispensatores*, *ordinarii*, etc. (for the long list see Štaerman and Trofimova 1975, 82), there were different types of lawsuits or *actiones*: *exercitoria*, *institoria*, *tributaria*, etc.

Some scholars (Katz 2008) have thus suggested to apply this old mechanism to contemporary transactions mediated by artificial agents and tomorrow's intelligent robots. Given the increasing extent of their autonomy, a new sort of peculium could in fact represent the right way to approach and balance the different human interests involved. Whereas, by employing robots or artificial agents to do business, transactions, or contracts, people could claim a limited liability, robots' peculium would guarantee their human counterparties, or other robots, that obligations would really be met.

Besides, when compared to other artificial agents and the typical issue of anonymity on the Internet, most of the interactions with robots will have the advantage of avoiding such a hard issue of anonymity, in that transactions, contracts, and businesses will often be in the 'real world.' This does not mean, of course, that we will need no business models as they have been proposed in the case of artificial agents: It is enough to recall the insurance model illustrated by Curtis Karnow (1996), or the authentication model of Andrew Katz (2008). Robots will indeed raise issues of trustfulness, reliability, traceability, identifiability, and the like, along with psychological problems related to the very interactions with robots as matters of attachment and feelings of subordination, deviations in human emotions, etc. (Veruggio 2007, 29)

Nonetheless, from a legal viewpoint, we should not miss the crucial point: The very idea of the peculium as well as the parallelism between robots and slaves is so attractive, because they show a sound way to forestall any legislation that might prevent the use of robots due to their risks and the consequent excessive burden on the owners (rather than, say, on the producers and designers) of robots. By striking a balance between people's claim not to be dilapidated by their robots' activities and the interest of the robots' counterparties to be protected when transacting with them, an updated form of peculium seems particularly interesting in order to address a new generation of contractual obligations and a novel source of agency as well. Whether or not you are ready to admit that robots would be 'legal persons' (Solum 1992; Teubner 2007), the characteristic pragmatism of the jurists of Ancient Rome indicates

how to achieve both forms of limited liability for the owners of robots and of business warranty for robots' counterparties.

4.2 Robots and extra-contractual obligations

Legal robotics does not only concern the enforcement of rights and obligations created by robots' business, because robots also raise problems of extra-contractual responsibility for damages caused by them. This scenario transcends the mechanism of peculium and involves what Roman jurists defined in terms of Aquilian protection (Zimmermann 1988, 1017); namely, the form of responsibility which stems from the general idea that people are held liable for unlawful or accidental damages caused to others because of their personal fault: *Alterum non laedere* (Pagallo 2009).

However, as I stressed in the previous Sect. 3.1, this type of extra-contractual responsibility comprises specific forms of 'strict liability torts' which correspond to the idea of 'objective responsibility' or liability without fault in the civil law tradition. Regardless of any illicit or culpable behaviour, in other words, people are held liable both for the damages caused by their own dangerous activities, as in the case of some torts of 'product liability,' and for the harm caused by their own children, animals, and even employees. By considering that robots are interactive, autonomous, and adaptable, we will thus need a new sort of legal responsibility for others' behaviour: If the mechanism of peculium can guarantee a form of legal accountability for what robots do in the field of contractual obligations, it is likely that we will have a new kind of liability without fault for the consequences of the behaviour of robots in the realm of extra-contractual obligations.

In order to illustrate how such a responsibility may be constructed, it is important to understand how the burden of proof is allocated in these cases.

Sometimes, law imposes liability without regard to the intent of the subject or her use of ordinary care. Employers, for example, are often held liable for any illicit action the employees engage in under their working contract activities. In the case of robots, such a policy could obviously be mitigated so as to prevent the risk that people think twice before using or employing robots at all. We could perhaps make insurance compulsory as we have done in most legal systems with cars. We might also extend the mechanism of peculium by determining that human extra-contractual liability should be limited to the value of their own robots' portfolio (plus, eventually, the compulsory insurance set above).

Anyhow, legal systems also provide for limits to such liability without fault. This is what typically happens to parents who evade responsibility for their children's

behaviour when they prove they could not prevent their actions. This is also what occurs to the owners of animals when they prove that a fortuitous event happened. While regarding the set of dangerous activities, some legal systems exclude liability when it is proved you have taken all the ‘appropriate measures’ in order to prevent any sort of damage, we may guess what sort of limited responsibility fits our robots. Once we agree they are neither killers nor simple fridges, should the behaviour of our modern slaves be legally assimilated to our children’s actions? Alternatively, should we assume that the behaviour of robots is ontologically or intrinsically dangerous? Or, as David McFarland (2008) claims, should we compare robots’ behaviour with the actions of our own pets?

5 Conclusions

We are unlikely to run across a single metaphor to grasp the next generation of robot-related legal issues in the realm of extra-contractual obligations. It is plausible that that responsibility will vary according to the different typology of robots we are confronted with. In some cases, it will be a matter of preventing the actions of robots (*robots as kids*); in others, we ought to prove that a fortuitous event has occurred (*robots as pets*). Then, when discussing forms of liability without fault, we will have to decide whether human liability should be limited to the value of robots’ peculium and the amount of their possible compulsory insurance (*robots as employees*).

In any event, lawyers will soon need to sort out a new kind of extra-contractual responsibility for others’ actions because robots represent a decisive step in a novel generation of cases involving the accountability of artificial agents. As their human counterparts, robots *fully act* as state-transition systems: They do not only interact with each other and their environment insofar as they can change their properties and inner states, without that change being a direct response to interaction. Even though they have no moral responsibility (for *robots cannot be killers*), they are sometimes accountable (for *robots can really be bad fridges*). Robots will increasingly be the source of evil or good actions, and we will be forced to mull over the possibility to conceive them as legally liable for what they do with the peculium of their owners (*robots as slaves*).

Therefore, along with a specific set of contractual obligations, we will face two sorts of novel responsibilities.

On the one hand, we will have a new *moral* responsibility because robots are agents that improve the informational nature of the universe. Since ‘good’ or ‘evil’ can conveniently be represented as anything that enhances or damages the informational complexity of the whole, we

should preserve, cultivate, and enrich the properties of our own robots because they are informational objects par excellence. It is not unlikely, on this basis, that new forms of prosecution could be envisaged against humans to avoid owners from unjustly damaging or suppressing their robots (Pagallo 2010).

On the other hand, *legal* responsibility will need to be further extended: Besides contractual obligations and cases of liability without fault involving animals and human beings, we are going to be liable for what some artificial agents autonomously do. The necessity to introduce a new kind of extra-contractual protection depends on the fact that robots do act: Whether or not they are ‘really’ conscious, ‘mechanically’ conscious (Aleksander and Dunmall 2003), or ‘phenomenologically’ conscious (Franklin 2003), robots are capable of gaining knowledge from their own behaviour, thus becoming unpredictable. Whereas we will assess their actions in terms of possible harm and damage, it is highly probable that the key legal point in fields like edutainment-related robotics, health care, or life quality applications will concern how we educate, treat, or manage our robots.

As it happens with children and pets, this kind of extra-contractual obligation does not exclude robots’ autonomy since they will keep a certain sphere of freedom which is compatible with the legal responsibility of the owners. What is new about this kind of liability without fault is that, very soon, it will include legal responsibility for what an artificial state-transition system *chooses* to do!

This innovative notion of agency represents one of the most relevant topics for the social understanding of robots today. We should be prepared to accept a new category of behaviour which is not purely human nor barely animal, yet produces multiple relevant legal effects. Robots are here to stay, so the aim of the law should be to wisely discipline our mutual relationships.

References

- Abernathy CF (2006) Law in the United States. Thomson West, St. Paul
- Aleksander I, Dunmall B (2003) Axioms and tests for the presence of minimal consciousness in agents. In: Holland O (ed) Machine consciousness. Imprint Academic, Exeter, pp 8–18
- Allen C, Varner G, Zinser J (2000) Prolegomena to any future artificial moral agent. J Exp Theor Artif Intell 12:251–261
- Asaro P (2007) Robots and responsibility from a legal perspective. In: Proceedings of the IEEE conference on robotics and automation: workshop on roboethics. Rome, 14 April 2007
- Asimov I (1942) Runaround. Doubleday, New York
- Asimov I (1985) Robots and empire. Doubleday, New York
- Barrio F (2008) Autonomous robots and the law. Society for computers and law. Retrieved August 15th, 2009, from <http://www.scl.org/site.aspx?i=ho0>

- Bekey GA (2005) *Autonomous robots: from biological inspiration to implementation and control*. The Mit Press, Cambridge
- Bynum TW (2006) Flourishing ethics. *Ethics Inf Technol* 8:157–173
- Clarke R (1993) Asimov's laws of robotics: implications for information technology. *IEEE computer* 26(12):53–61 and (1994) 27(1):57–66
- Comanducci P (2006) *Le tre leggi della robotica e l'insegnamento della filosofia del diritto. Materiali per una storia della cultura giuridica XXXVI* 1:191–197
- Dennett D (1997) When HAL kills, who's to blame? In: Stork D (ed) *HAL's legacy: 2001's computer as dream and reality*. MIT Press, Cambridge, pp 351–365
- Epstein RG (1997) *The case of the killer robot*. Wiley, New York
- Ewald W (1995) Comparative jurisprudence (I): what was it like to try a rat. *Am J Comp Law* 143:1889–2149
- Floridi L (2007) Artificial companions and their philosophical challenges. *E-mentor* 5(22):84–86
- Floridi L, Sanders JW (2005) On the morality of artificial agents. *Mind Mach* 14(3):349–379
- Franklin S (2003) IDA: a conscious artifact? In: Holland O (ed) *Machine consciousness*. Imprint Academic, Exeter, pp 47–66
- Hart H (1961) *The concept of law*. Clarendon Press, Oxford
- Karnow CEA (1996) Liability for distributed artificial intelligence. *Berkeley Technol Law J* 11:147–183
- Katz A (2008) Intelligent agents and internet commerce in ancient Rome. Society for computers and law. Retrieved August 15th, 2009, from <http://www.scl.org/site.aspx?i=ho0>
- McFarland D (2008) *Guilty robots, happy dogs: the question of alien minds*. Oxford University Press, New York
- Moravec H (1999) *Robot: mere machine to transcendent mind*. Oxford University Press, New York
- Pagallo U (2009) Sul principio di responsabilità giuridica in rete. *Il diritto dell'informazione e dell'informatica XXV*(4–5):705–734
- Pagallo U (2010) The human master with a modern slave? Some remarks on robotics, ethics, and the law. In: Arias-Oliva M, Bynum TW, Rogerson S, Torres-Corona T (eds) *EthiComp 2010: the “backwards, forwards and sideways” changes of ICT*. Universitat Rovira I Virgili, Tarragona, pp 397–404
- Reynolds C, Ishikawa M (2007) Robotic thugs. In: Bynum TW, Rogerson S, Murata K (eds) *EthiComp 2007: glocalisation*. Global e-SCM Research Center & Meiji University, Tokyo, pp 487–492
- Solum LB (1992) Legal personhood for artificial intelligence. *North Carolina law rev* 70:1231–1287
- Štaerman EM, Trofimova MK (1975) *La schiavitù nell'Italia imperiale. I-III secolo*. Editori Riuniti, Roma
- Teubner G (2007) Rights of non-humans? Electronic agents and animals as new actors in politics and law. Max Weber Lecture at the European University Institute of Fiesole, Italy on 17th January 2007
- Veruggio G (2007) Euron roboethics roadmap. In: *Proceedings Euron roboethics atelier*. February 27th–March 3rd, 2006, Genoa, Italy
- Watson A (1988) *The digest of Justinian, vol I*. University of Pennsylvania Press, Philadelphia
- Zimmermann R (1988) *The law of obligations. Roman foundations of the civilian tradition*. Clarendon, Oxford