Chapter 1
Introduction

HELENA:	*You mean you make them start to work as soon as they're made?*
DOMIN:	*Sorry. It's more like working in the way a new piece of furniture works…*
HELENA:	*How do you mean?*
DOMIN:	*Much the same as going to school for a person. They learn to speak, write, and do arithmetic. They have a phenomenal memory. If one reads them a twenty-volume encyclopaedia, they could repeat it back to you word for word, but they never think up anything original. They'd make fine university professors.*

Karel Ĉapek*, Rossum's Universal Robots,*
Introductory Scene

Abstract The aim of this book is to introduce laypersons to the complex set of principles, concepts, and ways of legal reasoning that govern the design, construction, supply and use of robotics technology today. In light of the classical distinction between legal plain and legal hard cases, attention is drawn to the cases where the disagreement among lawyers regards either the meaning of the terms framing the legal question, or the ways such terms are related to each other in legal reasoning, or the role of the principles that are at stake in the case. Paradoxically, the fact that a strong consensus still exists in the field of the laws of robots becomes clearer when the behaviour of robots falls within the loopholes of the system, provoking a new generation of hard cases.

The two different magnitudes of complexity as explored in this book, robotics technology and the law, challenge not only each other, but also today's society. Following the term Isaac Asimov coined in his 1942 novel, *Runaround*, "robotics" is the field dealing with the design and construction of a quantity of machines as varied as network centric-applications, adaptive robot servants, robot soldiers, unmanned ground and underwater vehicles, robot toys and even robot nannies. Robotics today is one of the most exciting fields of scientific research and technology, spanning several disciplines, such as artificial intelligence ("AI") and computer science, cybernetics, physics and mathematics, electronics and mechanics, neuroscience, biology and the humanities. Despite the multiplicity of robotic applications, some argue that we are dealing with machines built basically upon the mainstream "sense-think-act" paradigm of AI research (Bekey 2005). Sebastian Thrun, director of the AI Laboratory at Stanford, California, similarly reckons that robots are machines with the ability to "perceive something complex and make appropriate decisions" (in Singer 2009: 77). Others stress that robots are those machines able to learn and adapt to changes in environments. The UN World 2005 Robotics Report proposes a general definition of robot as a reprogrammable machine operating in a semi- or fully autonomous way, so as to perform manufacturing operations (*e.g.*, industrial robots), or provide "services useful to the well-being of humans" (*e.g.*, service robots).

These definitions do not dispel all doubts. References to the autonomy or intelligence of robots often are a source of misunderstanding. Consider the UK Ministry of Defence's Joint Doctrine Note on "unmanned aircraft systems" dated 30 March 2011. The notion of autonomy there is connected to a system "capable of understanding higher level intent and direction." Moreover, according to the Note, "estimates of when artificial intelligence will be achieved (as opposed to complex and clever automated systems) vary, but the consensus seems to lie between more than 5 years and less than 15 years, with some outliers far later than this." Opponents find this statement "ludicrous": in *Automating Warfare* (2011), Noel Sharkey affirms that, apart from the metaphorical use of the words, robots are not going to be "capable of understanding higher level intent," nor will they think like human beings in the foreseeable future. Likewise, Kenneth Himma argues in *Artificial Agency* (2007) that robots and other artificial agents ("AAs") do not meet the necessary and sufficient conditions required for properly claiming they engage in autonomous behaviour, as AAs lack the requisites of consciousness, free will and intent.

Sci-Fi scenarios aside, certain types of robots are already challenging tenets of social interaction, basic rules among nations, and even cornerstones of the law. "Even if they have the intelligence of a refrigerator"

(Floridi 2007), robots can improve the set of instructions through which their inner states change, and transform such properties without external stimuli: therefore, they can deal successfully with their tasks by exerting control over their own actions without any direct intervention by humans. As the 2007 EURON Roboethics Roadmap states, "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 2006). This specific autonomy of the robot, taking decisions of its own, seems particularly critical in such fields as military robotic technology: the United States military forces fund more than one-half of the American research and development ("R&D") in AI today. Consequently, looking at certain military robotic applications is instructive in shedding further light on the notion of robots that can rule (*nomos*) over themselves (*auto*) and, thus, are autonomous in a general sense.

For example, in the field of unmanned aerial vehicles ("UAVs"), a distinction should be drawn between "autonomous" and "semi-autonomous" machines. Some drones, such as the US Air Force's RQ-1 and MQ-1 Predators, have to be considered semi-autonomous. Others are fully "independent of real time UAV-pilot control input," according to the UK Defence Standards definition of autonomous flight. Think of the Global Hawk and the US Navy's anti-ship missile defence system, the Phalanx CIWS, operating completely alone. Some 40 countries currently are developing even more sophisticated forms of autonomous lethal weapons and other types of robot soldiers, a development summed up by scholars as "killer robots" (Sparrow 2007; Krishnan 2009), "robotic lethal behaviour" (Arkin 2007), or "autonomous military robotics" (Lin et al. 2008). Although these machines are not conscious of themselves and do not enjoy any "higher level intent and direction," they can act and decide beyond the direct control of humans. Norbert Wiener justly warned about the "autonomy of robots" in *The Human Use of Human Beings* (1950): the use of robots in battle might lower the requirements of declaring or entering into war, invoke a disproportionate use of force, violate the principle of discrimination and immunity, and might even provoke accidental wars. By considering the impact of today's robot soldiers on traditional categories of *ius ad bellum* (*i.e.*, when and how resort to war can be justified) and *ius in bello* (*i.e.*, what can justly be done in war), it can be remarked that the menace of robotic behaviour is as old as the very idea of "robot."

The word "robot" was used for the first time in Karel Ĉapek's 1920 play, *Rossum's Universal Robots*. The plot revolves around a factory producing artificial persons, "robots," whose rebellion ultimately leads to the extinction of the human race. In the second act, individuals at the headquarters of

R.U.R., the world manufacturer of thousands of robots located on a remote island, wonder why these machines are revolting against humanity. Dr. Gall, Head of the Physiology and Research Department at R.U.R., reckons that the "crucial mistake" they made was to turn some of these machines into "robot soldiers."

> This is just the same old evil as Europe has always committed. They just couldn't leave their damned politics alone and so they taught the robots to go to war, they took the robots and turned them into soldiers and that was a crime against humanity (Ĉapek 1920, Act 2).

Reality, at times, outpaces fantasy: since 2005, combat air patrols by US drones have increased by 1,200 % and, under President Barack Obama, the frequency of such strikes in Pakistan has risen tenfold "from one every 40 days during George Bush's presidency to one every four" (*The Economist*, 8 October 2011, p. 32). Significantly, Christof Heyns, Special Rapporteur on extrajudicial executions, urged in his 2010 Report to the UN General Assembly that Secretary-General Ban Ki-moon convene a group of experts in order to address "the fundamental question of whether lethal force should ever be permitted to be fully automated."

Robotics behaviour has appeared to be a source of risk and potential threat in other realms as well: the financial troubles in late 2008 may have been facilitated by the use of "robo-traders" such as AI brokers, electronic agents and smart digital interfaces. Since the early 2000s, experiments with Zero Intelligent ("ZI") agents, developed by the University of Pennsylvania and Lehman Brothers, have shown troubling similarities to the greediness of human speculators. In *Rights of Non Humans?* (2007), Günther Teubner sums up these concerns, claiming that robotics technology and other smart artificial agents raise problems of alienation and reification in social life that already troubled Karl Marx (*Entfremdung*) and Martin Heidegger (*Verdinglichung*). The overall idea is that autonomous AAs "create aggressive new action centres as basic productive institutions" so that we should bring the "economic, social and technical transactions run by electronic agents… back under human control" (Teubner 2007: 21).

Admittedly, the use of robo-traders in financial markets and autonomous lethal weapons on battlefields is alarming. However, let us avoid sweeping generalizations. Rather than machines that necessarily "alienate" (Marx) or "reify" (Heidegger) human life, we should pay attention to the number of robotics applications that, according to the UN World 2005 Report, provide "services useful to the well-being of humans." To start with, think of intelligent vehicles driving themselves on highways, a popular subject of Sci-Fi movies such as Michael Keaton's Batmobile in *Batman* (1989), or, for that

matter, the smarter AI cars in *Demolition Man* (1992), *Timecop* (1993), *Minority Report* (2002) and *I, Robot* (2004). Over the past decade, research (*e.g.,* Stanford University and Carnegie Mellon), business (General Motors and Volkswagen), and both (Google), have made this dream come true. To cut to the chase, the Nevada Governor in June 2011 signed a bill into law that for the first time ever authorizes the use of driverless cars on public roads. Of course, this is not to say that today's AI chauffeurs are as sophisticated as the Sci-Fi cars in Hollywood movies. Moreover, the Nevada Assembly (36–6) and Senate (20–1) acknowledged that "regulations authorizing the operation of autonomous vehicles on highways within the State of Nevada" may take a long time. Still, *pace* Teubner, robotic automation might not be a bad thing, once we recall that the autonomy of human drivers causes around 1.3 million accidents and 41,000 deaths on EU roads every year.

Likewise, contemplate certain useful applications in the industrial and service sectors. For example, a new generation of unmanned water-surface and underwater vehicles for remote exploration began in the 1990s to undertake emergency and hazard management work, by preventing damage, alerting controllers, fixing oil leaks, and so forth. Some of these underwater robots became popular in 2010, when they were employed for stopping the BP oil spill in the Caribbean Sea. In addition, a number of artificial companions and helpers at home, such as robot toys and robot nannies, are programmed in the field of service robots for domestic or personal use, to provide love and take care of children and the elderly. In the show business and music industry, consider the success story of the Japanese pop star robot singer HRP-4C. Developed by the Institute of Advanced Industrial Science and Technology's media interaction group, this amazing "divabot" is capable of singing, dancing, "breathing," and even performing her (!) shows. While HRP-4C uses the Vocaloid software developed by Yamaha, as well as a VocaListener to synthetize the notes of the songs, a VocaWatcher program allows HRP-4C to analyse individuals' facial tics as this divabot moves her hips and belts out a tune. Although it may be conceded that a robotic Maria Callas would be more stimulating than the current robotic Lady Gaga, it is difficult to see why this machine should *a priori* be likened to her more troubling cousins, robo-traders and robot soldiers. Some of these AI nannies and show biz pop girls raise a number of psychological issues concerning feelings of subordination, attachment, trustworthiness, etc. Yet, going back to some current picture of robotics, *e.g.*, Teubner's *Rights of Non Humans?*, it is problematic to dismiss such robots as an expression of "aggressive new action centres."

Robotic applications bring about a new set of constraints and opportunities that transform, reshape and even enrich individual and social environments.

This twofold aspect of robotics, as a source of good and evil, has been stressed by a number of scholars who, interestingly, insist on the impact of both military robotics technology and the service robots "useful to the well-being of humans," illustrated by the UN 2005 Report. Introducing a special issue of *AI & Society* on "the social impact of AI" (2011), Greg Michaelson and Ruth Aylett emphasize that "the recent advancements in the now mature discipline of Artificial Intelligence… have rekindled problematic social and ethical questions about our relationships with machines," adding to the tension between "killer robots" and "friendly fridges." Similarly, in the introduction to the special issue of *Philosophy & Technology* on "robotics: war and peace" (2011), John Sullins reckons that the ethical questions about our relationships to robots can be fruitfully addressed in connection with the following spectrum: at one end, "robots of war" such as MQ-9 Reapers or C-3PO Terminators may be presented as emblems of the "aggressive new action centres" of Teubner's version of robotics; at the other end of the spectrum are "robots of peace," such as the Japanese pop singer HRP-4C or, say, the da Vinci surgery system in the medical sector. What is common to robotics, from this point of view, ultimately revolves around the normative challenges of this technology, that is, "why we should, or should not, deploy these systems in our homes and battlefields" (Sullins 2011).

The kinds of robotic applications we are willing to implement is a crucial question today for ethics, economics, philosophy of technology, psychology and other fields. Here, the focus is not on how the manifold applications of robotics technology obey the "laws" of disciplines such as mathematics, physics, neuroscience, biology, and so forth. Rather, attention is drawn to the reasons why such machines should, or should not, be deployed in accordance with the aim of the moral, political and economic fields, in governing the process of technological innovation. Figure 1.1 below shows how the different magnitudes of complexity concerning the "laws of robots" can be illustrated:

Let us now augment the intricacy of this model by focusing on the second magnitude of complexity as explored in this book. In addition to multiple robotic applications and the laws of such disciplines as AI and computer science, cybernetics, and so on, that which is under scrutiny concerns the legal challenges facing this field: "the laws of robots." The first problem is identifying what is common to robotics through the lens of the "laws of the law."

Traditionally, when determining "what the law is," scholars distinguish the law from other academic fields, such as politics, ethics or economics. However, certain scholars affirm that the law ultimately depends on such
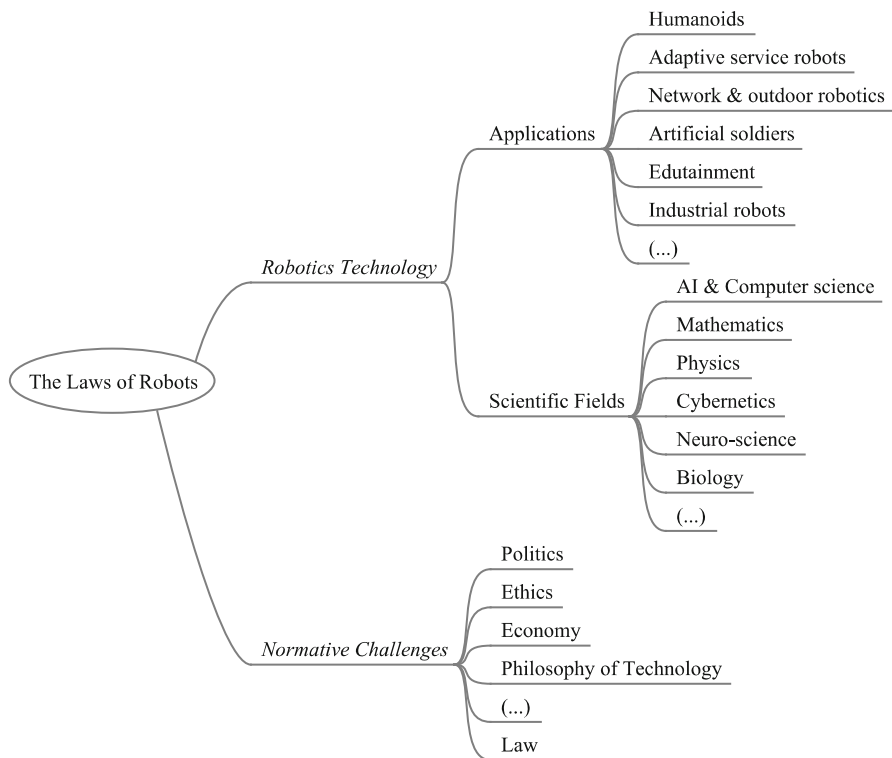
Fig. 1.1 The magnitudes of complexity of robotics technology

fields: a realist would trace the law back to politics, an advocate of the natural law tradition to ethics, an expert of the economic analysis of law (as well as an orthodox Marxist) to economics, a techno-determinist scholar to technology, and so forth. It suffices to mention the thesis of a "reductionist," such as the Italian philosopher Benedetto Croce. In *Riduzione della filosofia del diritto alla filosofia dell'economia* (1907), Croce sums up the efforts of legal philosophers to distinguish their own field of law from morals, by the image of the "Cape Horn" of legal science. The overall idea is that lawyers, trying to circumnavigate this issue, end up in a "conceptual storm" and "wreckage." In light of today's debate in legal theory, and how the variations of positivism (both inclusive and exclusive), realism, institutionalism, and different traditions of natural law, perceive the connection between law and morals, some words on the normative fabric of the legal phenomenon seem necessary, in order to clarify the legal approach of this book to the laws of robots. The nature of law and its connection with the moral sphere can be

properly understood by examining circumstances under which individuals (and robots) are confronted with responsibility.[1]

Reflect on cases where responsibility is imposed on individuals for harm resulting from their own fault. This is typical when an individual voluntarily performs a wrong prohibited by law, *e.g.*, tiny robotic helicopters employed in a jewellery heist. In criminal law, the legal accountability for this kind of behaviour is entwined with the notion of the moral responsibility of the individual and the idea of blameworthiness. Criminal defendants ought to be subject to the ordinary process of moral assessment in order to determine whether they are guilty under the law. In civil (as opposed to criminal) law, the general idea is similar, in that individuals are held liable for unlawful or accidental damages caused to others due to personal fault. This idea is traditionally summed up by the Roman maxim, *alterum non laedere*, that is, "do not injure others." Although further examples can be given, it should be clear that legal and moral reasons can overlap. We return to this below.

However, there are other circumstances in which individuals find themselves confronted with legal responsibility and yet, the actor's moral responsibility is not at stake. The first case of legal (as opposed to moral) responsibility refers to the idea that "everything which is not prohibited is allowed." In criminal law, this principle is connected to the clause of immunity summed up, in continental Europe, with the formula of the principle of legality, *i.e.*, "no crime, nor punishment without a criminal law" (*nullum crimen nulla poena sine lege*). Even though certain behaviours might be deemed as morally wrong, *e.g.*, spying on individuals through domestic robots, individuals can be held criminally liable for that behaviour only on the basis of an explicit criminal norm. In the wording of Article 7 of the 1950 European Convention on Human Rights, "[n]o one shall be held guilty of any criminal offence on account of any act or omission which did not constitute a criminal offence under national or international law at the time when it was committed."[2] *Vice versa*, there are cases where the law establishes no-fault liability, that is, regardless of the person's intent or ordinary care. Although a conduct may be deemed morally sound, a statute or a specific norm can establish liability for that behaviour. An example of this can

---

[1] The connection between the law and such fields as politics, economy, and technology, is further examined in Chap. 5.

[2] As lawyers know, there is a savings provision pursuant to art. 7(2) of the Convention, which states: "This article shall not prejudice the trial and punishment of any person for any act or omission which, at the time when it was committed, was criminal according the general principles of law recognized by civilized nations." The aim of this provision is to cover such exceptional cases as the Nuremberg trial against the Nazis.

be seen with editors, publishers and media owners (newspapers, TV channels, radio, etc.), parties who are liable for damages caused by their employees, notwithstanding their eventual illicit or culpable behaviour. This mechanism is invoked in many other types of cases where the law imposes liability regardless of the person's intention. Besides individuals' responsibility for the behaviour of their pets and, in most legal systems, their children, this type of strict liability applies to most producers and users of robots.

Going back to Croce's Cape Horn in legal theory, we can shed further light on the normative efforts of the law from a broader perspective, that is, by distinguishing plain from hard cases (*e.g.*, Hart 1961; and Dworkin 1986). A way to circumnavigate Croce's problem exists: we can avoid storms and conceptual wreckages by drawing attention to all the cases where a complex set of concepts and notions in legal reasoning are at work and, still, leave no doubts as to how to apply the clauses and conditions of responsibility/liability in the legal field. According to Herbert Hart, these are the cases where the legal issues are pretty plain, that is, "where the general terms seem to need no interpretation and where the recognition of instances seems unproblematic or 'automatic'… where there is general agreement in judgements as to the applicability of the classifying terms" (Hart 1994: 123). Clauses of immunity in criminal law and cases of no-fault liability in tort law may thus represent a class of such plain cases, in that, here, the distinction between an individual's moral and legal responsibility is not an issue at all. Throughout this book, we are going to see further examples of this general agreement on how the principles, norms, and rules of the legal system work: namely, cases of responsibility pursuant to the liability model in accomplice cases of criminal law (Chap. 3), cases of responsibility that depend on the voluntary agreement between private persons in the civil law field (Chap. 4), down to the strict liability hinging on the idea of dangerous activities in tort law (Chap. 5). This network of concepts in legal reasoning allows scholars to examine matters of unpredictability and risk as provoked by robots, as was the case with previous technological innovations.

Still, there are cases where scholars (and parties to a lawsuit) may disagree. Here, the storms and conceptual wreckages of Croce's Cape Horn represent a class of legal issues that scholars dub as hard cases, for instance, where the disagreement may regard the meaning of the terms that frame the legal question, or the ways such terms are related to each other in legal reasoning, or the role of the principles that are at stake in the case. However, which principles, which concepts, and which ways of legal reasoning, at times end up in a sort of legal stalemate has to be determined in connection with the norms and provisions established by statutes, international agreements, or the case law of the common (as opposed to the civil) law tradition. Work on the logic and nature of the law, such as Croce's own research in legal philosophy, in other words is
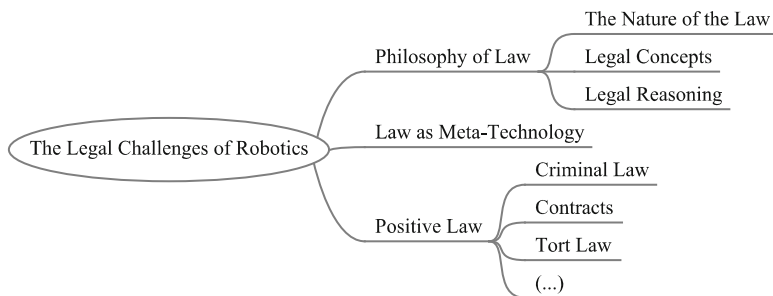
Fig. 1.2 A philosophy of law for lawyers and a work in positive law for philosophers

a necessary, but insufficient ingredient of the analysis: in order to determine whether a legal issue appears hard, or plain, we need the knowledge of experts in positive law as much as the efforts of legal philosophers. For example, regarding the military employment of robotic applications, focus should be on the 1907 Hague Convention, the four Geneva Conventions from 1949, and the two 1977 additional Protocols, which define the current laws of war and the international framework of humanitarian law. In the case of, say, the civilian use of unmanned aerial vehicles, attention should be drawn to the 1948 Chicago Convention on International Civil Aviation and, in Europe, the EU Regulation 216/2008. In the case of the civilian use of unmanned water-surface and underwater vehicles, the legal point of reference is the 1972 IMO COLREGs Convention on maritime law.

This twofold approach to the laws of robots, that is, both the perspective of legal philosophers and the knowledge of experts in positive law, can be summed up with a sort of interface, or level of abstraction,[3] through which this book aims to describe, examine, and argue about the laws of robots. What I propose here is to approach the laws of the law establishing the conditions of legitimacy for the design, production, and use of robots, conceiving the law as meta-technology, *i.e.*, as a means to govern other technological means. This perspective sheds further light on topics of legal philosophy (*e.g.*, the nature of the law, concepts, legal reasoning), as well as provisions of positive law. Figure 1.2 sums up this level of abstraction:

As seen from Book IV of Plato's *The Republic*, this idea is not new: "The regulations which we are prescribing, my good Adeimantus, are not, as might

---

[3]On the methodology of the "level of abstraction," this author draws on Luciano Floridi's work. See *The Method of Levels of Abstraction* (2008) and, more recently, the second volume of Floridi's *Principia Philosophiae Informationis*, namely Information Ethics (2013). By varying the "interface," the "set of observables" changes accordingly: more details on this method in Sect. 2.1.3.

be supposed, a number of great principles, but trifles (Plato 2006)." In this context, the regulative efforts of the law can be illustrated with the thesis of the *Pure Theory of Law* (1934/2002) and *General Theory of the Law and the State* (1945/1949). Here, Hans Kelsen provides a classical account of the law as "a specific social technique of a coercive order" enforced through the menace of physical sanctions: "if A, then B." The legal formula shows "what should be" (*Sollen*, ought to), rather than "what is" (*Sein*, is), namely, punitive sanctions (B) that should follow terms and conditions of legal accountability (A), rather than effects (B) that follow natural causes (A). The distinction between normativity and natural causality means that the aim of the law, to govern the conditions of legitimacy for technological innovation (A), hinges on what should happen in terms of legal responsibility (B). In the phrasing of the *General Theory of the Law and the State* (1949: 26): "What distinguishes the legal order from all other social orders is the fact that it regulates human behaviour by means of a specific technique." Once such technique regulates other techniques and, moreover, the process of technological innovation, we may accordingly conceive the law as a meta-technology.

To be sure, law can be considered as a form of meta-technology without buying Kelsen's ontological commitment. The stance this book adopts does not imply either that the law is merely a means of social control, or that there are no other meta-technological mechanisms. Rather, the level of abstraction defined by law as meta-technology aims, first, to describe how legal systems deal with the process of technological innovation, through such a complex network of concepts, as agency, accountability, liability, burdens of proofs, clauses of immunity, or unjust damages. The analysis dwells on the conditions of legitimacy for the design, construction, and use of robots, as scholars have done since they started examining the impact of automation on the law in the late nineteenth century. Think of Günther's *Das Automatenrecht* (1892), Schels' *Der strafrechtliche Schutz des Automen* (1897), Schiller's *Rechtsverhältinesse des Automen* and Ertel's *Der Automatenmissbrauch und seine Charakterisierung als Delikt*, both from 1898, to Neumond's *Der Automat* in 1899. More than a century later, there is still a relatively strong consensus: in a great number of cases, the rules that govern the design, production and use of such machines (Kelsen's A) are unchallenged, as well as the consequences in terms of legal responsibility (B).

Then, *pace* Kelsen, we should pay attention to the impact of robotics technology on the formalisms of the law, and how we grasp the meaning of certain key terms concerning the aim of the law to govern the process of technological innovation. This impact brings us back to the hard cases of the law, and how we should address them. Some affirm "there is no possibility of treating the question raised by the various cases as if there were one

uniquely correct answer to be found, as distinct from an answer which is a reasonable compromise between many conflicting interests" (Hart 1961: 128). Others, as Ronald Dworkin and followers of the "right answer" thesis, on the contrary interpret the law in a morally coherent way, so that, given the nature of the legal question and the history and background of the issue, *e.g.*, whether to ban robot soldiers through a UN sponsored agreement, lawyers could obtain the solution that best justifies or fits the integrity of the law.

That suggested here is restricting the focus of the analysis and summarizing the complex set of principles, norms and rules establishing the conditions of legitimacy for the design, production and use of robots, through the concepts of legal responsibility (Kelsen's B) and agency (*i.e.*, a key term of Kelsen's A). This stricter perspective emphasizes that which all cases concerning the laws of robots have in common, namely, the conditions whereby legal agents, both human and artificial, are confronted with responsibility. Whether a unique right answer exists (*e.g.*, Dworkin), or not (Hart), we have to preliminarily ascertain the terms through which the law frames technological research and development, so as to take sides in today's debate. Theoretically speaking, three legal notions of agenthood are at stake:

  (i) Legal persons with rights (and duties) of their own;
 (ii) Proper agents establishing rights and obligations in civil law;
(iii) Sources of responsibility for other agents in the system.

Likewise, the different types of cases where agents are confronted with legal responsibility should be stressed:

  (i) The aforementioned clauses of immunity (*e.g.*, the principle of legality);
 (ii) Conditions of strict liability (*e.g.*, no-fault responsibility of editors);
(iii) Cases of responsibility for damages that depend on fault (*e.g.*, intentional torts).

On this basis, three different levels of analysis can be distinguished:

  (i) The different ways robots do act in legal systems (Kelsen's A);
 (ii) The consequences following from the production and use of such machines (Kelsen's B);
(iii) The overall impact of technology on legal systems, so as to determine whether a case is plain, or hard (*e.g.*, Dworkin vs. Hart).

Table 1.1 summarizes this approach with nine possible scenarios:

The legal observables of responsibility for the behaviour of robots in light of Table 1.1 clarify the philosophical challenges of the field, *e.g.*, its hard cases, and the matters of responsibility in positive law, *e.g.*, robotic crimes.

Table 1.1 The behaviour of robots and nine ideal-typical conditions of legal responsibility

Responsible robot	*Immunity*	*Strict liability*	*Unjust damages*
As legal person	I-1	SL-1	UD-1
As proper agent	I-2	SL-2	UD-2
As source of damage	I-3	SL-3	UD-3

Let us become acquainted with such ideal-typical conditions of responsibility for the behaviours of robots:

"I-1," "SL-1," and "UD-1" have in common that robots should be considered as proper persons with rights (and duties) of their own, that is, the thesis of what I call the front of Robotic Liberation. "I-1" means that a person is protected by clauses of immunity, *e.g.*, the principle of legality. "SL-1" stands for cases of no-fault responsibility of the robot as being *sui iuris*. Finally, "UD-1" concerns protection against harm provoked by others: for example, the State, contractual counterparties, third parties in tort law.

"I-2," "SL-2," and "UD-2" share the idea that (some types of) robots can properly be conceived as strict agents in business law: for example, with negotiations and contracts. "I-2" has to do with clauses of immunity in the civil (as opposed to the criminal) law field, such as protection pursuant to safe harbour clauses. "SL-2" *vice versa* emphasizes liability of this robot agent, regardless of intentions or personal fault. Then, "UD-2" stresses that such agents should be protected against unjust damages.

Finally, "I-3," "SL-3," and "UD-3" summarize the traditional viewpoint of scholars that robots would not affect basic cornerstones of the law. As simple tools, and not agents, in the legal system, robots can only represent a source of responsibility for other agents. Therefore, "I-3" means that humans, as well as artificial persons such as corporations, evade responsibility for damage provoked by robots, *e.g.*, clauses of immunity in the laws of war. "SL-3" highlights today's strict liability policies for the design, construction and use of robots. "UD-3" concerns cases of responsibility for human negligence or intentional wrongdoing, which have to be added to the previous hypothesis of no-fault responsibility.

In light of Table 1.1, the complex network of concepts, through which the law aims to govern the process of technological innovation, results in the traditional focus on the question of "Who pays?" This question suggests three scenarios for a hard case in positive law. The disagreement can concern:

  (i) The legal personhood of robots and their constitutional rights;
 (ii) The legal accountability of robots in contracts and how this autonomy impacts other fields of the law;
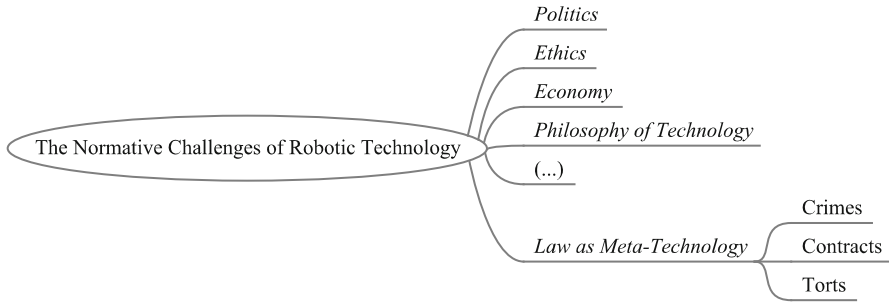(iii) New types of human responsibility for others' behaviour.

Fig. 1.3 Three legal fields for responsible robots

Once such possible candidates for a hard case in the laws of robots are grasped, we have to augment the intricacy of the model: "Who pays?" often means different things in such fields as criminal law, contracts, and torts. The level of autonomy that at times is sufficient to produce relevant effects in the field of contractual obligations (that is, "I-2," "SL-2," and "UD-2"), arguably is insufficient to bring robots before judges and have them declared guilty in criminal courts (*e.g.*, "SL-1"). Likewise, when considering robots as a source of responsibility for other agents in the system ("I-3," "SL-3," and "UD-3"), attention should be drawn to the different ways we say an "agent pays its debt." In criminal law, think of the different reasons underpinning the legitimacy of inflicting punishment, *e.g.*, the theory of retribution, or of special and general prevention. In civil (as opposed to criminal) law, reflect on obligations imposed by the government that can even overrule clauses and conditions of responsibility established by the parties to a contract. In tort law, individuals are held responsible for unjust damages inflicted upon third parties, that is, harm provoked to other agents in the system. This field-sensitivity suggests refining the focus of the model by grasping the specific features of each field of the law. This stricter perspective is illustrated with a new scheme in Fig. 1.3:

By increasing the resolution of this model, new (classes of) legal issues follow as a result. Chap. 3 below explores the popular debate on robotics technology and criminal law, averting Sci-Fi scenarios, *e.g.*, criminally accountable robots. After examining matters of legal responsibility and agenthood (Chap. 2), the aim is to show that robots are affecting basic tenets of the law in two different ways. First, these machines are inducing some problems that are specific to criminal law, mostly to do with clauses of immunity. Besides the immunity of military and political authorities for the use of robots in battle, we have to determine whether the behaviour of robots

falls within the loopholes of the system, necessitating the intervention of lawmakers at both national and international levels, as they did in the early 1990s when establishing a new class of computer crimes. Then, a second class of legal issues concerns how the growing autonomy of robots affects key notions of the system, such as reasonability, predictability, or foreseeability, on which an individual's fault depends. Certain scholars have suggested a failure of causation, since it would be difficult to predict what types of harm may supervene (Karnow 1996). This is a class of hard cases that criminal lawyers share with experts in tort law and contracts: for example, think of clauses and conditions between private persons often crucial in determining the party who is liable for robots involved in criminal enterprises. It should be stressed that in 2010 some criminals used tiny robotic helicopters in a jewellery heist.[4] After matters of reasonable foreseeability in criminal law, such a class of hard cases has to be further examined in the fields of contracts and torts.

The starting point of Chap. 4 is the 2005 "World Robotics"-Report of the UN and the Economic Commission for Europe, mainly focusing on "robots of peace" such as environmental robots, surgical robots and edutainment robots. Here, responsibility and legal accountability for the design, construction and use of robots, are framed as a matter of risk and predictability in contractual obligations. In addition to artificial doctors and cognitive automata such as commercial software-agents, some riskier applications, *e.g.*, ZI agents and unmanned ground vehicles (UGVs), stand for a further set of hard cases. Besides a new type of legal agent for contracts (*i.e.*, "I-2," "SL-2," and "UD-2"), the ability of robots to produce, through their own intentional acts, rights and obligations on behalf of humans, entails the risk that individuals can be financially ruined by their robots' activities. Some reckon that "the best method of accident control may be to cut back on the scale of the activity" through strict liability policies (Posner 1973: 180). Yet, it is feasible to avert legislation that makes individuals think twice before using or producing robots at all: consider new models of insurance and legal accountability for such machines, *e.g.*, the "digital peculium" of robots. Contrary to traditional forms of distributing responsibility and risk, "only robots shall pay" could, at times, be a sound approach to the contract problem (Chopra and White 2011).

Chapter 5 looks at extra-contractual responsibility, *i.e.*, when robots damage third parties rather than their contractual counterparties. What common

---

[4]*Nature*, 22 September 2011, p. 399.

lawyers define as torts deals with obligations between private persons imposed by the government to compensate for damage done by wrongdoing. In the civil law tradition, this idea of extra-contractual responsibility can be traced back to the Ancient Roman-law status of Aquilian protection, as the form of responsibility stemming from the general idea that individuals are liable for unlawful or accidental damages caused to others due to personal fault. The new class of hard cases that the growing autonomy of robots is likely to induce, concerns how we should interpret a novel kind of liability for the behaviour of others. For the first time ever, legal systems will hold humans responsible for what an artificial state-transition system "decides" to do. Moreover, this kind of liability crucially depends on the different kinds of robots with which we are dealing: a robot nanny, a robot toy, a robot chauffeur, a robot employee, and so forth. This is one of the most innovative aspects in the field of the laws of robots, as traditional forms of responsibility for the behaviour of children, pets, or employees, have to be complemented with new strict liability policies (*e.g.*, Posner); or, alternatively, mitigated through insurance models, authentication systems, and the mechanism of allocating the burden of proof.

Chapter 6 brings us back to the law as meta-technology. From the different classes of hard cases as previously mentioned, it does not follow that the aim of the law to govern the process of technological innovation, necessarily falls short in coping with its own purpose. In light of Table 1.1 (*i.e.*, "Is", "SLs," and "UDs"), we can pinpoint cases and classes of specific legal disagreements and yet, most of the time, a relatively strong consensus on both the conditions of legitimacy for the design, construction and use of robots, and the consequences in terms of responsibility, can luckily be found. Paradoxically, this general agreement makes it easier to identify potential hard cases in the field. By distinguishing between concepts of personhood (*i.e.*, "I-1," "SL-1," and "UD-1"), traditional immunity ("I-3"), causation ("UD-3"), artificial agency in contracts ("I-2," "SL-2," and "UD-2"), and new types of responsibility in tort law ("SL-3"), we can determine which cases should be taken seriously or be given priority. For example, certain scholars reckon that the legal personality of robots does not seem necessary or even convenient in the foreseeable future (Sartor 2009). However, you can be a supporter of the front of Robotic Liberation and still admit that the regulation of new robotic crimes ("I-3") should have priority over the three "1s" of Table 1.1: I-1, SL-1, and UD-1.

The conclusion of this book summarizes how scholars address the challenges of this field as first coined by Asimov in the early 1940s: "robotics." More than seventy years later, it is remarkable how his plots foresee many of the crucial issues of today's debate: the legal personhood of robots,

questions of logic on how the "laws of the law" have to be interpreted, up to the design of machines that should comprehend and process such sophisticated information as the current laws of war and rules of engagement. Between law and literature, the message of Asimov's stories seems to be clear: since robots are here to stay, the aim of the law should be to wisely govern our mutual relationships.