DOI: http://dx.doi.org/10.5235/17579961.5.2.214 (2013) 5(2) LIT 214–247

Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules

Andrea Bertolini*

I. FRAMING THE PROBLEM: VAGUE DEFINITIONS AND THE NEED FOR REGULATION

Robots are often said to be the technology of the future;[1] yet such a statement is apt to mislead, its most evident weakness being the very notion of 'robot' upon which it rests. By that term, indeed, very different applications are acknowledged,[2] so that at least in some cases (industrial robots, to name one specific kind) they are already widely used, although they might go unnoticed. In most cases, however, it is hard to anticipate what results can and will be achieved by scientific research in the near future,[3] and the most common depictions may then—at a later date—appear to be quite far from reality.[4]

* Post-Doctoral Fellow in Private Law at Santa Anna School of Advanced Studies (SSSA), member of the Robolaw Unit of SSSA. The research leading to this paper received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no 289092 (RoboLaw). All websites accessed October 2013.

1 See William Henry Gates, 'A Robot in Every Home' *Scientific American* (January 2007), suggesting that the fate of robots will be similar to that of computers over the last few decades; see also Elizabeth Corcoran, *The Robots Are Coming!*, www.forbes.com/2006/08/17/robot-egang-history_06egang_cz_ec_0817robotintro.html. A more detailed analysis is made in M Ryan Calo, 'Open Robotics' (2011) 70(3) *Maryland Law Review* 571 ff. See also Gianmarco Verruggio and Fiorella Operto, 'Roboethics: Social and Ethical Implications of Robotics' in Bruno Siciliano and Oussama Khatib (eds), *Handbook of Robotics* (Springer, 2008) 1519.

2 See Bruno Siciliano and Oussama Khatib, 'Introduction' in Siciliano and Khatib (n 1) 1 ff.

3 In his novel *I, Robot* (Voyager Classics, 2013), Isaac Asimov depicted a world where there was a pervasive diffusion of robots, possessing a 'positronic' brain whose functioning was similar to that of human beings, and whose creator—a scientist born in 1982—claimed they were capable even of feelings and moral conduct beyond human possibilities, thus amounting overall to better beings.

4 For instance, one may take the example of artificial intelligence to see how the notion of strong AI—see Alan Turing, 'Computing Machinery and Intelligence' (1950) 59 *Mind* 433—is today outclassed by the narrower and more limited notion of weak or light AI: see below, section IV; for a detailed discussion see Luciano Floridi, *Philosophy and Computing: An Introduction* (Routledge, Kindle edn 1999) pos 2862 ff, and thus future robots are most likely not going to resemble their current literary depictions.

Finally, some applications will never come into existence, either because of intrinsic technical limitations, or because of active policy decisions of national and supranational authorities (see below, section IV).

This lack of precision derives to a great extent from science fiction,[5] which has always lured the imagination of the many, thanks to the works of bright intellectuals and genial novelists. The stories of sophisticated machines capable of serving human beings with their extraordinary powers and abilities, yet which—in most cases at least—have ended up revolting against their own creators or users,[6] expose the ancestral attraction and fear of men attempting to tame the laws of nature and make themselves equal to gods.

Quite clearly, the notion of robot diffused today among the public at large is closely related to the evocative images of the *Maschinenmensch*[7] in Fritz Lang's 1927 masterpiece, rather than to the complex and articulated taxonomies offered by scholars, which include applications such as the steering aid mounted on tractors.[8] However, the lack of precision in defining the precise object of the analysis could prove most harmful. On a very general level, overlapping the fictional take and the existing—or reasonably foreseeable—technologies may trigger those very feelings of uneasiness and fear, so well described in the literature, which could ultimately impair the possibility of developing useful tools for everyday life. More narrowly, failing to identify the peculiar differences that characterise specific applications may result in insufficient, inefficient or ineffective measures being taken to provide the correct incentives for the development of desirable robotic technologies.

It may be claimed that, in view of existing scientific uncertainty, no action should be taken since 'overly rigid regulation may stifle innovation'.[9] Such a statement is, however,

5 Sam N Lehman-Wilzig, 'Frankenstein Unbound' [1981] *Futures* 444 makes the argument that 'since it is beyond human capability to distinguish *a priori* the truly impossible from the merely fantastic, all possibilities must be taken into account. Thus science fiction's utility in outlining the problem.' In the current article, the opposite argument is made that science fiction should not be guiding legal analysis for the latter to be meaningful, since it does not provide a more accurate estimate of what future technological development will be.

6 See Edward Tenner, *Why Things Bite Back: Technology and the Revenge of Unintended Consequences* (Vintage, 1997) 5 ff.

7 The *Maschinenmensch* (literally Machine-man) was created by the inventor Rotwang in order to bring back to life his love, Hel, who years before had left him in order to marry Fredersen, the master of Metropolis, and then died giving birth to his son Freder. In an act of revenge Rotwang kidnapped Maria and gave the robot its exterior aspect. The robot is meant to bring chaos to the city and lead men to murder through lust.

8 Michael Decker, 'Technology Assessment of Service Robotics: Preliminary Thoughts Guided by Case Studies' in Michael Decker and Mathias Gutman (eds), *Robo- and Informationethics: Some Fundamentals* (Lit Verlag, 2012) 64.

9 This appears to be the claim of Aneta Podsiadla, 'What Robotics Can Learn from the Contemporary Problems of Information Technology Sector: Privacy by Design as a Product Safety Standard—Compliance and Enforcement', paper delivered at the conference 'We Robot: Getting Down to Business', Stanford, 8 April 2013, http://blogs.law.stanford.edu/werobot/agenda. To this end, Podsiadla cites the Robolaw project (www.robolaw.eu) as an example of such a risky if not erroneous move, contrasting it with her idea of first assessing already existing applicable regulation. Yet, the Robolaw project began its analysis by describing already existing legal principles and norms, which without any further adaptation would apply to robotic

simplistic and cannot be accepted. Regulation, in fact, is *per se* a comprehensive term under which very different tools are included. National and supranational authorities may intervene with binding laws, as well as with so-called soft law or standards, which parties may freely decide to adopt. Different approaches lead to different outcomes that need not be rigid. Even if regulation was narrowly defined as national legislative intervention (top-down mandatory regulation), its purpose may well rather be to foster innovation[10] and to provide correct incentives where the market itself would otherwise fail,[11] sometimes because of the inadequacies of the legal system. Finally, such a perspective tends to exclude the role of regulation in the very shaping of technologies, whereas robotics, like any other field of technological development, is only good so long as it serves a purpose and rests on the fundamental principles expressed and accepted in the society in which it is meant to be used. In other words, technology is neither good, nor bad, nor neutral, and law, as well as other social sciences, ought to intervene at an early stage to provide guidance in its design and creation.[12]

The analysis in this paper will focus—within the overall framework briefly sketched—on liability issues, trying to identify the correct paradigm under which robotic applications ought to be considered. For this purpose the very notion of 'robot' will be discussed (section II), showing that all attempts at providing an encompassing definition are a fruitless exercise: robotic applications are extremely diverse and more insight is gained by keeping them separate. This, on the one hand, excludes the very possibility and necessity of elaborating an autonomous set of liability rules for torts involving robots; on the other hand, it inclines one to conclude that robotic applications need to be addressed individually, by identifying a specific distinctive trait, which will trigger—as necessary—a change in the legal analysis.

applications: see *Inventory of Current State of Robolaw* (Robolaw Grant Agreement No 289092, D3.1, 2012). The concluding remark pursuant to which 'regulation of robotics is currently unnecessary' (p 50) appears rather bold if not naïve, in particular as the purport of a general—and therefore necessarily incomplete—survey of existing applicable regulation; moreover it conflicts directly with some of the author's remarks suggesting for instance the adoption of a liability cap in the US legal system (47) and of a negligence standard in Europe, to replace strict liability (46). The real issue is therefore what, when and how to regulate and the extremely diverse nature of robotic applications does not allow for a universal answer to those questions.

10 This is most certainly the case with the Korean Act No 9014, 28 March 2008 on Intelligent Robots Development and Distribution Promotion Act (IRDDPA).

11 Specifically to the theme here discussed, the claim Calo (n 1) 601 ff appears to be making is precisely that existing norms ought to be modified or integrated so as to shield researchers and producers of robotic technologies from liability claims, which could otherwise impede the development of a fruitful market: see below, section XI.

12 Reference to fundamental rights and constitutional principles is most certainly essential; also Podsiadla (n 9) *passim* and in particular 49 suggests that privacy should be taken into account when designing a robot so that it complies with existing standards and regulations. See also Martin Meister, 'Investigating the Robot in the Loop: Technology Assessment in the Interdisciplinary Research Field Service Robotics' in Decker and Gutman (n 8) 38; but other criteria may be called upon to decide how to shape forthcoming applications.

Moving from this consideration, the two most commonly recurring characteristics—namely autonomy and the ability of a robot to learn—identified both by the legal and technological discourse as being the essential turning points for the assessment of liability, will be analysed. The argument will be made that a strong and a weak notion of autonomy need to be kept separate (section III); the former, derived from the philosophical notion of moral agency (section IV), would surely force a change in the existing legal paradigm; the latter, corresponding to the technical aspect of the control and functioning system of the robot (section V), is not *per se* sufficient to justify a change in the rules for the ascription of liability, and the comparison made with animals appears to be misleading (section VI). The ability to learn (section VII), then, narrowed down in its meaning, still does not suffice to justify a change in existing liability schemes, because the robot would still be either performing a program or exerting a freedom, which was attributed to it by its producer.

If all these alternative solutions have to be disregarded, robots may and shall be deemed products, thus objects and not subjects of law. Existing product liability rules are then briefly addressed (section VIII) for the purpose of showing that they are not inherently inadequate to tackle issues of liability arising from the use of robots. The notion of foreseeability and the development risk defence are therefore identified as the criteria that allow sufficient elasticity into the system (section IX); at the same time the possibility of treating robots as having legal personhood is briefly sketched so as to identify the framework within which it might be considered as a plausible alternative to the application of existing norms (section X).

Finally, it will be shown that the analysis conducted does not necessarily imply that the incentives provided by existing regulation are always desirable: indeed product liability rules are often ineffective. At the same time it is not the mere technological aspects that prevail in assessing the appropriateness of existing norms, but rather a policy argument, taking into account the specific nature and desirability of the given application, as well as the specific market failures involving its use and diffusion (section XI).

II. THE QUEST FOR A DEFINITION: A POINTLESS EXERCISE

The *Merriam Webster* dictionary defines 'robot' as

> 1a: a machine that looks like a human being and performs various complex acts (as walking or talking) of a human being; *also*: a similar but fictional machine whose lack of capacity for human emotions is often emphasized … 2: a device that automatically performs complicated often repetitive tasks; 3: a mechanism guided by automatic controls.[13]

13 www.merriam-webster.com/dictionary/robot.

This definition is clearly influenced by the literary depiction of robots, and is thus incomplete.[14] It is incomplete since many applications do not walk or talk and can either be quite simple, such as a vacuum cleaner, or complex, like a surgical or industrial robot. Some are conceived to mimic human emotions or animal behaviours, for the purpose of keeping the elderly or children company;[15] others are being developed to perform operations which entail a certain degree of creativity (softbots) or even provide a first assessment of the medical condition of a patient,[16] thus elaborating complicated data in a very different fashion from one time to another. Finally, as concerns the resemblance to human traits, studies show that, beyond a given point, users find that aspect awkward and unsettling, and so designers tend to preserve clear-cut signs of the mechanical and artificial nature of machines so that they will be more easily accepted in human environments.[17]

A more comprehensive definition is offered by the *Oxford English Dictionary*, which includes *crawlers*:[18]

> 1) a machine capable of carrying out a complex series of actions automatically, especially one programmable by a computer: … (especially in science fiction) a machine resembling human being and able to replicate certain human movements and functions automatically: … a person who behaves in a mechanical or unemotional manner: … 2) another term for crawler (in the computing sense) … 3) (South African) a set of automatic traffic lights: …

The same criticism raised above could be repeated, yet it suffices to point out how inadequately descriptive of real robotic applications such definitions are, and how misleading; if one had to determine what kind of technology qualifies as a robot, the proffered criteria would induce wrong conclusions in most cases. Definitions offered by researchers are always more precise and narrowly tailored to accommodate the specific field of interest of the speaker,[19] yet the outcome is fragmented and contradictory if considered together.[20] Finally, a lowest common denominator approach is unsatisfactory as well. Defining a robot as a machine which autonomously performs a task[21] is at most a synecdoche,

14 For a taxonomy which takes into account the different ethical issues raised, see Verruggio and Operto (n 1) 1151 ff.

15 See PARO, www.parorobots.com.

16 See WATSON, www-03.ibm.com/innovation/us/watson.

17 For a minimal and essential reference see Masahiro Mori, 'The Uncanny Valley' [2012] *IEEE Robotics & Automation Magazine* 98, translation by Karl F MacDorman and Norri Kageki of the original 1970 seminal article.

18 Software used by search engines to analyse systematically all data circulating on a network: see http://en.wikipedia.org/wiki/Web_crawler.

19 For a complete survey see Pericle Salvini, *Taxonomy of Robotic Technologies* (Robolaw Grant Agreement No 289092, D4.1, 2013), 17.

20 See George A Bekey, 'Current Trends in Robotics: Technology and Ethics' in Patrick Lin, Keith Abney and George A Bekey (eds), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press, 2012) 17.

21 The one offered in the text is a translation from the Italian '*una macchina che svolge autonomamente un lavoro*', found in Amedeo Santosuosso, Chiara Boscarato and Franco Caroleo, 'Robot e diritto: una prima ricognizione' [2012] *La Nuova Giurisprudenza Civile Commentata* 494, 498.

since it identifies the peculiar trait of an entire set of applications by reference to one of its possible control mechanisms,[22] and ultimately pointless because it is so general that it fails to provide sufficient guidance when attempting to distinguish a robot from other applications which operate unattended.

The reason why there is not and could never be a satisfactory definition of the term 'robot' is its a-technical nature, both from an engineering and a legal point of view. Being derived from science fiction,[23] the word solely means labour and more precisely enslaved labour. The technologies that have developed and the applications that exist are so diverse that maintaining the use of that term may only serve the purpose of synthesis, allowing one to indicate an extensive set of objects. Therefore, rather than a definition, a classification ought to be created, where various criteria are considered, such as: (i) embodiment or nature; (ii) level of autonomy; (iii) function; (iv) environment; and (v) human-robot interaction,[24] and individual applications should then be analysed accordingly.

If, then, a notion of robot is to be elaborated for merely descriptive—thus neither qualifying nor discriminating—purposes, it may be as follows: *a machine which (i) may either have a tangible physical body, allowing it to interact with the external world, or rather have an intangible nature—such as a software or program, (ii) which in its functioning is alternatively directly controlled or simply supervised by a human being, or may even act autonomously in order to (iii) perform tasks, which present different degrees of complexity (repetitive or not) and may entail the adoption of non-predetermined choices among possible alternatives, yet aimed at attaining a result or provide information for further judgment, as so determined by its user, creator or programmer, (iv) including but not limited to the modification of the external environment, and which in so doing may (v) interact and cooperate with humans in various forms and degrees.*

The consequence for the purpose of the present analysis is that we may not address the legal issues posed by robots unitarily, since the inherent technical differences between robotic applications cannot be overlooked without losing insight.[25] At the same time, given that all attempts to identify the common trait of all robotic applications appear to be fruitless, attention should rather be devoted to isolating the traits that could be of relevance in changing the paradigm within which to frame single robotic applications and the liability issues they raise. Thus a higher degree of precision needs to be reached by social scientists when describing the aspects considered to be forcing a change in

22 See Salvini (n 19) 8.

23 Karel Čapek, *R.U.R. (Rossumovi univerzální roboti)* (1922).

24 See Salvini (n 19) 22 ff.

25 Cf Christophe Leroux *et al*, *Suggestion for a Green Paper on Legal Issues in Robotics: Contribution to Deliverable D.3.2.1 on ELS Issues in Robotics* (2012), www.eurobotics-project.eu/cms/upload/PDF/euRobotics_Deliverable_D.3.2.1_ELS_IssuesInRobotics.pdf, 7. These authors instead attempt to discuss the legal issues posed by robots unitarily, according to a top-down approach, with the aim of developing sets of rules which could be common to most if not all applications.

existing legal concepts, starting with a definition of all possible notions of the frequently recalled 'autonomy'.

III. THE DIFFERENT NOTIONS OF AUTONOMY

When discussing issues of liability, it is frequently claimed that, because (some) robots are 'autonomous', existing legal norms appear to be inadequate.[26] Robots interact in the environment in an unpredictable fashion, which the programmer or producer cannot foresee or to some extent control; this should ultimately suggest the recognition of legal personhood of the machine itself,[27] for the purpose of holding it directly liable.

However, such statements are often ambiguous, since the particular meaning of autonomy is not always sufficiently specified.[28] As already stated, autonomy could be considered, in a technological perspective, as one of the possible control mechanisms for robotic applications, where alternatives are possible, among which tele-operation, tele-presence and supervision might be included.[29] Even in such a case, though, it could still be disputed whether a robot operating without constant human supervision, but which could be controlled in a moment of need (say in case of a malfunction), would or would not qualify as autonomous. Yet the ambiguity of the term is much greater.[30] Hence, in order to determine whether such characteristic—namely autonomy—may justify a shift in the choice of the optimal rules for the ascription of liability, it needs to be considered according to all different perspectives.

There are three meanings of autonomy intended in a more or less explicit fashion by social scientists when discussing robotic applications: (i) self-awareness or self-consciousness, leading to free will and thus identifying a moral agent,[31] (ii) the ability to intelligently interact in the operating environment,[32] and (iii) the ability to

26 Curtis EA Karnow, 'The Application of Traditional Tort Theory to Embodied Machine Intelligence', We Robot (n 9) 1 ff; Andreas Matthias, 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata' (2004) 6 *Ethics and Information Technology* 175.

27 Leroux *et al* (n 25) 60 ff.

28 Discussing artificial intelligence, Karnow (n 26) 3 states something which could easily be transposed to the broader notion of autonomy: 'The notion of intelligence as applied to machines is often just shorthand for "I don't know how they do that so quickly," an amazement borne of ignorance. We might in that way ascribe intelligence to Apple's Siri, which can respond to basic voice commands with vaguely contextually correct responses, missile defense systems which distinguish hostile intruders, and stock market programs which in fractions of a second calculate the best price and execute trades. The apparent magic of these advanced technologies is generally a function of speed outside the human scale, and of the observer's ignorance of the programs being used.'

29 For a discussion of all three see Salvini (n 19) 8.

30 For discussion see Willem FG Haselager, 'Robotics, Philosophy and the Problems of Autonomy' (2005) 3 *Pragmatics & Cognition* 517 ff.

31 See also Herman T Tavani, 'Ethical Aspects of Autonomous Systems' in Decker and Gutman (n 8) 99, discussing whether robots could be deemed 'full-ethical-agents'.

32 Santosuosso, Boscarato and Caroleo (n 21) 449.

learn.[33] From the outset, it is clear that only the second of these features may relate to (or reference) the aspect of control; the first has greater philosophical implications and only secondarily a technical relevance; and the third is quite a distinct concept referring to a different characteristic of the machine, which still influences its behaviour and—depending on how the concept is refined (see below, section VI)—which may bring about 'unforeseeable' outcomes. However, there is no bijection between autonomy and unpredictability of outcome: an autonomous behaviour may still be completely predictable if it corresponds to a program the machine was conceived to complete;[34] and, equally, something unpredictable may occur when a machine is operating under the direct and constant control of a human being.

Therefore, since reference to autonomy when dealing with a machine's ability to learn appears superfluous and potentially misleading, this aspect will be addressed separately, after discussing the two alternative notions of strong (hypothesis (i) in section IV) and weak (hypothesis (ii) in section V) autonomy.[35]

IV. STRONG AUTONOMY OR SETTING THE GOLEM FREE

From a philosophical perspective, responsibility may only be ascribed to a moral agent. A moral agent is defined as a subject whose actions are autonomous in that (i) they lack determination, and thus are free, and (ii) they pursue an endeavour which is properly the subject's own.[36] By contrast, in all cases where given certain conditions an outcome results without further external intervention a mere process is being observed, which—when applied to animals or other beings—qualifies as a behaviour. A behaviour is completely explained by an 'as if relation' which to the contrary does not suffice to describe an action identified 'in terms of means-end-relationship';[37] it is in fact 'only

33 In this sense see Karnow (n 26) 2; Matthias (n 26).

34 This idea appears to be shared by Karnow (n 26) 2, who states that autonomous vehicles are not necessarily unpredictable and thus are not 'interesting' for the purposes of his analysis because they lack the kind of autonomy which he deems would change the legal analysis. No matter how complex its functioning, a driverless car only performs the various tasks it was originally programmed to carry out.

35 Mathias Gutman, Benjamin Rathgeber and Tareq Syed, 'Action and Autonomy: A Hidden Dilemma in Artificial Autonomous Systems' in Decker and Gutman (n 8) 231 ff.

36 Tomis Kapitan, 'The Free Will Problem' in Robert Audi (ed), *Cambridge Dictionary of Philosophy* (Cambridge University Press, 2nd edn 1999) 326. Such a definition corresponds to that of full ethical agents identified by James H Moor, 'Four Kinds of Ethical Robots' (2009) 72 *Philosophy Now* 12, who however distinguishes other intermediate stages where ethical considerations may be relevant even if the robot is not self-conscious. The perspective adopted here is opposite: the possibility of behaving morally is deemed the criterion imposing the acknowledgment of the existence of a subject and not just of an object. Haselager (n 30) 521 ff instead discusses whether it is necessary to show free will in order to be deemed a moral agent, and identifies the possibility of having one's own goals as the purport of the integration of mind (or control system) and body, aiming at achieving homeostasis.

37 See Gutman, Rathgeber and Syed (n 35) 237.

the determination of a specified end that implies the necessity of actions of a specified kind'.[38]

Pursuant to this definition it can be empirically observed that actions pertain to humans alone, and autonomy—which we may call strong—shall thus be equated to the ability to reason and decide intentionally.[39] Most certainly no existing robotic application satisfies these fundamental conditions and it follows that none qualifies as autonomous in a strong sense.

Robots are in fact programmed to perform a task and are designed in a way to achieve the desired result most effectively: therefore, on the one hand, they present the highest degree possible of 'heteronomous determination', and show no understanding of semantics; on the other hand, they excel at syntactics.[40] A machine could be considered a moral agent if and only if it met those minimal requirements,[41] and thus was able to set its own goals as well as, being self-conscious, exert free will.[42] It is disputed, from a technical point of view, whether such a machine could actually be built;[43] recent studies on artificial intelligence show a more narrow scope directed at achieving a specific functional result rather than replicating the mechanisms of the human brain.[44] That said, for the present analysis, it is superfluous to assess the possibilities and likelihood of the various technologies that might at a later date be developed; here, it suffices to discuss the consequences that would follow if it became possible and such technological capability was actually achieved.

38 *Ibid*, 237.

39 See Dieter Sturma, 'Autonomie: Über Personen Künstliche Intelligenz und Robotik' in T Christaller and J Wehner (eds), *Autonome Maschinen* (Westdeutscher Verlag, 2003) 43.

40 See Bert-Jaap Koops, Mireille Hildebrandt and David-Oliver Jaquet-Chiffelle, 'Bridging the Accountability Gap: Rights for New Entities in the Information Society?' (2010) 11(2) *Minnesota Journal of Law, Science & Technology* 497, 528.

41 See Gutman, Rathgeber and Syed (n 35) 237; with a lesser degree of precision Tavani (n 31) seems to admit the possibility of having moral agents who yet do not satisfy these requirements. His position is not persuasive, however, since he is not providing sufficiently precise arguments to ground such a statement (the morality of an electronic agent incapable of exerting free will) otherwise.

42 If one takes the *Cambridge Dictionary of Philosophy,* the entry 'autonomy' directly forwards the reader to 'free will problem': see Kapitan (n 36) 326 ff; see also the discussion summarised by Tavani (n 31) 97 ff.

43 For a discussion see Joachim Hertzberg and Raja Chatila, 'AI Reasoning Methods for Robotics' in Siciliano and Khatib (n 1) 208: 'Reasoning requires that the reasoner … has an explicit representation of parts or aspects of its environment to reason about …' The engineering problem is that of identifying formalism suitable for representing knowledge to be used by the machine, which will be further distinguished in two sub-problems: 'epistemological adequacy: does the formalism allow the targeted aspects of the environment to be expressed compactly and precisely?' and 'computational adequacy: does the formalism allow typical inferences to be drawn effectively or efficiently?'. There is however a trade-off between an epistemologically satisfactory formalism and the possibility of inferring conclusions for the solution of problems (see also p 221).

44 On the distinction between GOFAI (good old fashioned AI) and LAI (light AI) see Floridi (n 4) pos 2862 ff. Essentially the former entails the construction of a machine 'whose behaviour would eventually be at least comparable, if not superior, to the behaviour characterising intelligent human beings in similar circumstances'; the latter instead simply aims at achieving a specific functionality.

By applying an adjusted version of the Turing test,[45] it could be said that a robot ought to be deemed autonomous in a strong sense if—and only if—it could develop rational explanations for its actions, in this way showing intention.[46] From a philosophical standpoint, if such a level of self-awareness was reached, robots would stop being objects and become subjects, capable of acting autonomously and therefore equal to human beings,[47] while—from a legal point of view—they would become '*Träger von Rechten*'.[48] Concerning liability rules, artificial entities could thus be held personally responsible,[49] without the need to identify the human behind them. Such a technological achievement would force a clear-cut shift in the paradigm utilised so far in analysing the issues connected with the harmful consequences arising from the presence of robots in society, but the law would most likely be apt at addressing such new problems effectively.

Within the more limited scope of liability issues, existing rules would in fact suffice. Such artificial entities would be capable of intentionally causing harm, would appreciate their own freedom and existence, and would thus fear criminal punishment;[50] as subjects and not mere objects they could own property and therefore have an estate with which to face the claims of victims of their misconduct.

It would be a matter of policy, or rather a political question, whether such entities should be granted equal rights or rather some form of diminished legal capacity. Jurists could resort to some traditional tools such as the Italian notion of *capacità giuridica*,[51] so as to discriminate among different beings. Racial laws, as well as the regulation of the rights of slaves in the Roman Empire, could provide alternative criteria for the purpose of operating such choices.[52] Yet, if equal treatment was denied, such a decision could rightly be considered discriminatory, triggering the ever-recurring question of whether or not such differences are justified.[53]

Finally, despite appearing provocative at first, it ought to be pondered whether such artificial persons with greater analytical capabilities or power as well as self-awareness

45 See Turing (n 4). The Turing test is not devoid of shortcomings and may oversimplify the robot's task by reducing the human being's freedom to act and thus forcing him to behave more like a robot than vice versa. On all these aspects see Floridi (n 4) pos 2598 ff.

46 See Gutman, Rathgeber and Syed (n 35) 240.

47 Sturma (n 39) 52.

48 Andreas Matthias, *Automaten als Träger von Rechten* (Logos, 2010).

49 See Matthias (n 26) 181–2.

50 The lack of self-consciousness is the most relevant argument that can be made in order to exclude the need to extend criminal liability to existing or future robotic application. If a machine cannot assess the value of its own existence and freedom, the threat to restrain it or even dismantle or disassemble it would represent no effective menace. See also Koops, Hildebrandt and Jaquet-Chiffelle (n 40) *passim* and in particular 560.

51 Very briefly see Cosimo Massimo Bianca, *Diritto civile*, vol 1, *La norma giuridica i soggetti* (Giuffrè, 2002) 213 ff.

52 See Lehman-Wilzig (n 5) 449.

53 For a discussion of whether some form of discrimination may be beneficial, see the considerations of Ronald Dworkin in *Taking Rights Seriously* (Duckworth, 1977) 225 ff.

might instead take control and decide for themselves what degree of freedom and rights to grant us.[54]

The analysis may then be pushed a step further in order to assess whether it is admissible to conceive of—beyond technical considerations—such types of artificial beings. Pursuant to Gutman, Rathgeber and Syed, an argument can be made to deny the theoretical possibility of an artificial moral agent.

As is well known, the Kantian *categorical imperative*[55] forbids that a counterpart be reduced to a pure means to the actor's own end; the relationship between *alter* and *ego* needs to be construed so as to place both subjects on the same level, and thus the second ought to be able to pursue its own desires. The artificial agent instead would still be created for a given purpose and this 'is a status, that cannot be undone by any decision of the [Artificial System]'. At the same time, even if we considered the case where the person decided to create the artificial system not as a tool, but rather as an end in itself, such a choice would be the person's, and therefore he would be responsible for it.[56] Since the fundamental decision to create the machine as a being was the human's, its freedom would intrinsically be denied, and with that its status as a moral being; therefore the conclusion can be drawn that the very notion of an artificial moral agent represents a *contradictio in adjecto*.

Such a philosophical argument could, though, be deemed insufficient by those who thought that, irrespective of any other consideration, if such a machine was created it ought to be recognised as having legal rights and duties. Therefore the deontological question shall be asked whether such autonomous robots, capable of exerting free will and pursuing their own goals, should be created, and the Golem set free.[57] Put in other words, we could say:

> Auf die Frage einer künstlichen Person, warum wir sie in maschineller Form überhaupt zur Existenz gebracht hätten, wären wir kaum besser vorbereitet als der unglückliche Dr. Frankestein. Wenn aber ernsthaft Projekte der künstlichen Erzeugung von Bewusstsein erwogen werden sollen, dann wäre es ratsam zu fragen, ob es überhaupt rechtfertigungsfähige Gründe dafür geben kann, auf technologischem Wege neue Bewusstseinsformen mit existenziellen und ethischen Eigenschaften zu entwickeln.[58]

54 Koops, Hildebrandt and Jaquet-Chiffelle (n 40) 561.

55 'Denn vernünftige Wesen stehen aller unter dem Gesetz, dass jedes derselben sich selbst und alle andere niemals bloß als Mittel, sondern jederzeit zugleich als Zweck an sich selbst behandeln solle.' Translation: 'For all rational beings stand under the law that each of them himself and all others should never be treated merely as a means, but always at the same time as an end in himself.'

56 Gutman, Rathgeber and Syed (n 35) 254.

57 See Barbara Henry, *Dal golem ai cyborgs: trasmigrazioni nell'immaginario* (Belforte Salomone, 2013).

58 Sturma (n 39) 52. Translation: 'To the question of an artificial person, asking why at all we brought it into existence in a machine-form, we would not be better prepared to answer than the unfortunate Dr. Frankenstein. Although, if projects for the artificial creation of consciousness were to be seriously pondered, it would be wise to ask whether justifiable reasons can be at all offered to develop, through artificial ways, new forms of consciousness with existential and ethical qualities.'

The question the Monster puts to Dr Frankenstein, asking for the reason for its existence,[59] cannot be simplistically answered 'because we could';[60] technological possibility does not *per se* ground a sufficient argument. Yet, a utilitarian[61] consideration can be made to oppose the creation of such entities. If robots were to show greater abilities (either in terms of analytical skills or power) than humans—which is the very purpose behind their creation—and were made free to decide for themselves, they would face the choice between good or bad. The consequences of their actions could be particularly beneficial or harmful to humans, according to the choices upon which they were grounded. The robot could in fact decide for itself and its notion of good could be in conflict with that of human beings, as much as choices beneficial to humans may be harmful to animals. So by creating such a being men would expose themselves to the risk of losing control or rather of being dominated by a superior entity, and this could not be prevented or avoided by human action.[62] Conversely, if the robot was programmed in such a way that it could not harm man, or so as to make decisions which are moral from man's perspective, and thus potentially conflicting with its own interest, it would again become a mere tool, determined in its behaviour, and thus not free. Therefore no matter how sophisticated, such a robot would still be an object and not a subject, which, from a legal perspective, would not justify the abovementioned shift in the applicable paradigm.

V. WEAK AUTONOMY

Autonomy could then be understood as the ability to operate without human supervision in a complex environment, assessing and evaluating data:[63] in this sense a driverless vehicle, an autonomous drone and a vacuum cleaner may qualify, despite presenting different degrees of complexity, and thus of autonomy.

Such a skill certainly represents much of the purpose for developing robotic technologies in the first place, since it allows humans to increase productivity and to free up their time.[64] From a philosophical perspective, we shall define this as a weak form of autonomy, pursuant to which the behaviour is not determined by the external intervention of another being; yet the robot is completing a task in order to achieve a goal set by

59 Mary Shelley, *Frankenstein* (Kindle edn, 2013) pos 1617 ff.

60 This is the answer David gives Holloway in the film *Prometheus* (Scott, 2012) to the same question as the Monster asks Dr Frankenstein.

61 Dan W Brock, 'Utilitarianism' in *Cambridge Dictionary of Philosophy* (n 36) 942 ff.

62 The same conclusion is reached by Lehman-Wilzig (n 5) 445: 'In sum one cannot give robots the Promethean fire-gift of intelligence and still hope to keep them sacked. One way or another, then, robot freedom must lead to some harmful behaviour even if well intentioned.'

63 Santosuosso, Boscarato and Caroleo (n 21) 499.

64 For a discussion of how technological development is always aimed at increasing the quality and amount of available time, see Floridi (n 4) pos 4696 ff.

the agent, namely the human being himself.[65] Within this horizon the machine may be provided with the highest possible degree of autonomy—more precisely 'heteronomous autonomy'[66] or the ability to choose between ends—going far beyond the capability to act without human supervision, embracing the ability to acquire data and elaborate it, up to the point of becoming aware of the environment and interacting with it. In such a scenario the human being showing 'autonomous heteronomy'[67] is capable of setting ends, which the machine then accomplishes, freely deciding for itself how to perform the task assigned. From a moral perspective, though, the artifact, no matter how sophisticated it may be, is not properly 'acting' in the philosophical sense described above; it is merely 'producing functional states'.[68]

Such a principle is not foreign to legal theory, since the notion of agency precisely reflects the dual nature of a subject (an agent) that acts towards an end set by another (a principal), producing direct effects in his patrimonial sphere, so long as it is acting within its powers or so long as it appears that way to a third person in good faith.[69]

The choices of the agent are to some extent free, in that he may choose how to perform the intended task,[70] since if they were not—and the subject was merely communicating a choice otherwise completely determined in its content by the principal—we may then be facing the different and more limited figure of a *nuncius*.[71] Such a similarity does not entail stating that existing legal rules would permit us to consider a machine the 'agent' of a human being, since the former may not yet qualify as a legal person.[72] Rather, it highlights how, even when faced with a high degree of autonomous judgment and decision-making by a subject—namely the kind of full-fledged autonomy that is typical of an adult human being—the law may allow effects to be directly produced on another subject, which is held responsible for having identified the desired outcome to be achieved. It should thus be further noted that

65 See Gutman, Rathgeber and Syed (n 35) 236.

66 *Ibid*, 246–7.

67 *Ibid*, 246.

68 *Ibid*, 247.

69 A peculiar case in US law, *Hoddeson v Koos Bros*, 47 NJ Super 224, 135 A 2d 702 (App Div 1957), clearly illustrates how far the principle can be pushed, stating that apparent authority can be established when 'a proprietor of a place of business by his dereliction of duty enables one who is not his agent conspicuously to act as such and ostensibly to transact the proprietor's business with a patron in the establishment … [I]n such circumstances the law will not permit the proprietor defensively to avail himself of the impostor's lack of authority and thus escape liability.'

70 The agent may choose which contract to enter into among possible alternatives, so as to best serve the purposes of his principal. Failing to do so may in fact trigger his liability towards the principal, whenever it may be proved that he served another subject's interests even if not his own (so-called conflict of interest). Such principles are common to most legal systems. See Bryan A Garner, *Black's Law Dictionary* (Thompson West, 8th edn 2007) 67 ff.

71 A mere *nuncius* is for instance under Italian law the person who materially substitutes the spouse in a wedding, in case the party is physically withheld (eg, is at war). See Paolo Gallo, *Trattato del contratto*, vol 2 *Il contenuto—gli effetti* (Utet, 2010) 1490–1.

72 See in the same sense Koops, Hildebrandt and Jaquet-Chiffelle (n 40) 512.

> [t]his determination of the functionality of an artificial system remains adequate even if the ends are realised via steps which are only determinable by their outcome and not by specific single steps. For example, if neural networks are used, which may be described as black-boxes considering the internal states of the net itself, the outcome has to be functionally equivalent to the determined ends.[73]

The overall consequence of this analysis is that, short of strong autonomy, machines cannot be deemed moral agents; in fact even if they might be programmed to act pursuant to moral rules and their decisions appeared to be moral they would still not be free and aware.[74]

The legal implication of such a statement is that the human behind the functioning of the robot ought to be held responsible for its actions.[75] The choice is thus restricted to two possible subjects: the owner/user and the producer. The latter, pursuant to existing norms, is the one to bear the consequences of any harm caused by its product. Yet the argument is sometimes made that because of the autonomous behaviour of the robot leading to some degree of unpredictability, other subjects could be held responsible, namely the user, pursuant to rules such as those that assess the keeper's liability for domesticated animals.

VI. ROBOTS AS ANIMALS

Robots are in fact in some cases compared to domesticated animals,[76] but the reasons for such a claimed similitude are not compelling. Indeed, it has been shown that (weakly autonomous) robots and animals behave—and thus 'act' depending on the natural or environmental conditions[77]—without the intervention of a human exerting direct control; yet this does not suffice to equate the two or to force a change in the existing legal paradigm.

73 Gutman, Rathgeber and Syed (n 35) 247.

74 *Ibid*, 254.

75 Besides, moral implications holding the robot liable *per se*, thus as an entity provided with legal personhood, would only serve the purposes of capping liability. Unless the robot was given the ability to earn its own income, someone else would in fact be contributing its assets and therefore ultimately would be responsible for funding any damages awarded in relation to the harm caused by the robot. See section X.

76 The idea was already anticipated by Lehman-Wilzig (n 5) 448; and see Enrique Schaerer, Richard Kelley and Monica Nicolescu, 'Robots as Animals: A Framework for Liability and Responsibility in Human-Robot Interactions', paper delivered at the 18th IEE International Symposium on Robot and Human Interactive Communication, Toyoma, Japan, 27 September–2 October 2009, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2271466, 72 ff. For these authors the theory of robots as animals is justified as a conservative yet realistic alternative, until truly autonomous robots—capable of passing the Turing test—are achieved. See, too, Chiara Boscarato, 'Who is Responsible for a Robot's Actions? An Initial Examination of Italian Law within a European Perspective' in Bibi van de Berg and Laura Klaming (eds), *Technologies on the Stand: Legal and Ethical Questions in Neurosciences and Robotics* (Wolf, 2011) 393 ff.

77 See Gutman, Rathgeber and Syed (n 35) 236.

The rationale behind vicarious liability rules holding an owner or keeper of an animal liable is animals' intrinsic danger, even when domesticated and trained, due to the unpredictably of their behaviour. Because animals cannot hold property and compensate for the damage they may cause, the owner, who is the one gaining personal advantage from the presence of the animal in society, is required to bear the costs of his choice.

Even if at a first glance some similarities with robots appear, there are still some fundamental differences which ultimately lead to the conclusion that such an equation is flawed. A robot that operates unattended is not necessarily behaving unpredictably,[78] even when it performs very complex tasks and analyses data derived from the active environment it is set into. A self-driving vehicle,[79] for instance, or an assisting robot, is programmed to act within a specific kind of environment (the street or the home) and the actions it is called upon to perform, no matter how complex and vast in number, can be enumerated. A self-driving car is programmed to consider the position of other vehicles, of potential passers-by, the destination of the person sitting inside it, traffic, and all street indications and rules.[80] The way it reacts to the multitude of these external signals does not necessarily render all its decisions unpredictable, just as the complexity of the data and the difficulty encountered by the human being in understanding its functioning does not necessarily imply that it does not follow a programmed logic.[81]

Deviations from the desired outcome may still occur (for instance leading to a crash), since the extreme complexity of the variables may induce errors in their appreciation, and at that point the lawyer may be called upon to assess whether that mistake was foreseeable (both under a strict liability rule such as in product liability cases, or under a general negligence standard—see below, sections VIII ff) and thus the producer ought to be held liable, or rather not, leaving the damage on the harmed party, or attributing it pursuant to other rules.[82]

If a given negative outcome was to be foreseen, it would be possible to prevent it, either by designing the product differently (so long as the alternative harm-preventing

78 For a similar view, see Karnow (n 26) 3.

79 Alberto Broggi *et al*, 'Intelligent Vehicles' in Siciliano and Khatib (n 1) 1176: 'an intelligent vehicle is defined as a vehicle enhanced with perception, reasoning and acting devices, that enable the automation of driving tasks such as safe lane following, obstacle avoidance, overtaking slower traffic, following vehicles ahead, assessing and avoiding dangerous situations, and determining the route.'

80 See *ibid*, 1178 ff and 1181; the most complex aspect from an engineering point of view is the understanding of the road environment.

81 See the extract from Karnow at n 28.

82 With respect to the circulation of vehicles, many legal systems compel the owner to purchase third-party insurance covering damage caused to others (even when the vehicle is not personally driven by the owner, who is the insured party under the insurance contract): see the Italian law of 24 December 1969, n 990. If self-driven cars offered at least the same degree of safety as human-driven cars, the current cost of insurance for vehicles would not increase; if, by removing man from the loop, they actually became safer—according to Broggi *et al* (n 79) 1177, up to 90% of traffic accidents are caused by humans, see also 1191—its costs may be expected to fall.

design was cost-effective)[83] or ultimately by not granting the given degree of freedom and specific technical capacity to the machine.[84] To put it another way, if it was not possible to design a safe enough self-driving vehicle,[85] it ought not to have been made completely autonomous.[86]

Ultimately, so long as a machine is executing the programs it was conceived to perform, even if said programs entailed acquiring external inputs and interacting with the environment, its behaviour may be deemed predictable for the purposes of the application of product liability standards and, pursuant to those very rules, there may be cases where liability is assessed, and others where it is denied. So long as a driverless vehicle is driving[87] and not completing a different task for which it was not originally designed,

83 An objective standard was elaborated under US law in order to ascribe liability for defective design, where the social cost imposed by the product on society is weighed against the cost of a safer alternative. The producer is held liable only when the first exceeds the second: see below, section IX, and n 133.

84 Santosuosso, Boscarato and Caroleo (n 21) 508 ff offer the example of a Ro-dog whose function is to assist the blind and help them move around autonomously. It is clear that such a technology is only desirable so long as it assures the same degree of safety as that offered by a well-trained real life dog. Therefore, unless the producer or designer of such an assistive technology device can guarantee that its product can interact with humans on the streets, recognise different kinds of terrains and the difficulties they may pose for the disabled, he ought not to release it onto the market.

85 The minimal degree of safety a robot should be required to guarantee is that offered by existing non-robotic technologies which could be used to perform the same task. That is to say, we might desire self-driving cars on our streets so long as they can offer at least the same degree of safety as a traditional human-driven vehicle.

86 A quite sophisticated robot (Dust-bot) was designed in Pisa to collect garbage from a real neighbourhood in an experiment which lasted a few weeks. Despite the robot being able to make its way through the traffic and around people walking by and all around it, without being supervised or remotely controlled, a help centre was available, in case the robot required intervention, and safety mechanisms were embedded. For more sophisticated robots distant supervision may be required in order to reinforce safety measures at all times when human interaction is deemed potentially, even if remotely, dangerous. A help centre connected through cloud mechanisms as well as over the internet could provide the additional precaution.

87 Matthias (n 26) 176 gives the example of a modified version of the Mars rover *Pathfinder* with a built-in 'navigation and control system that enables it to avoid obstacles autonomously' (ie without the intervention of an operator). Such a robot would be able to identify a terrain and assess and memorise the difficulties it encounters when crossing it the first time, so that if it later recognised a similar kind of soil it could elaborate a more successful strategy to go over or around it. According to Matthias, such a mechanism would amount to a learning ability of the rover, which would cause the outcome of its behaviour to be unpredictable to the point that it would highlight a responsibility gap: the producer/programmer would have lost control over its 'creature' (see also Andreas Matthias, 'From Coder to Creator: Responsibility Issues in Intelligent Artifact Design' in Rocci Luppicini and Rebecca Adell (eds), *Handbook of Research in Technoethics* (Hersher, 2008) 635 ff) since it is acting based on data which it was not originally provided with and basing its autonomous decisions on it. If it crashes due to a wrong assessment of the characteristics of a given terrain—according to Matthias—it is incorrect to hold the programmer liable, since it had *no control* over what has happened. However, this claim is open to question. First, it could be discussed whether such an ability would amount to actual learning; secondly, in any case it would not cause the behaviour to be unpredictable since the robot is specifically performing the task it was built for: identifying a terrain, assessing its characteristics while proceeding over it, eventually comparing them to other data stored in its memory, evaluating the best possible way to cross it, and finally storing more data in its database. It is safe to assume that if such a product was to be developed the producer and programmer would

then it cannot be deemed so intrinsically different from any other product irrespective of how sophisticated it is. Different considerations should instead apply if the robot was actually modifying itself in order to perform tasks diverging from those for which it was originally conceived (see section VII).

An animal, on the other hand, is a living thing with its own character, vicious or otherwise; it is capable of performing both actions that it was trained to do but also—and actually mostly—things it determines to do for itself, out of choice or instinct. The unpredictability of its actions in fact strictly depends on its nature or erratic behaviour which causes it to deviate from its training. The walking dog that chases a cat in the street, thus tripping up his owner, is following its instinct and deviating from its training; by contrast, an autonomous robot—which has not been given the ability to modify itself—may only erroneously assess variables and come to the wrong solution for its task, or assess them correctly and make a choice contemplated by its program which still leads to undesired consequences.

However, it should be said that the fact that a robot's weak autonomy cannot be equated to an animal's does not necessarily imply that the owner or user of the robot may never be held liable. It is obvious that if the owner or a third party misuses a product and thus causes harm to someone he may be called upon to bear such consequences as may derive from that misuse. Therefore, if the instructions provided with an automatic vacuum cleaner recommend that the user keep the door of the room where the robot is operating shut, and the user fails to do so, letting the machine stray, he might be held liable for damage suffered by a neighbour. A normal negligence standard could be successfully applied in such simple cases as well as in any other where it was clear that the robot was a tool being used to cause harm.[88] Furthermore, there may be conditions where it could prove useful to hold the owner or user liable because he is the one in the best position possible to intervene and avoid harm, or rather compensate the damage when it occurred, irrespective of whether he was at fault in causing it.[89] In the case of a self-driving vehicle, for instance, a possible solution would be to hold the owner liable, forcing him to purchase third-party insurance for the circulation of the vehicle, and

load all available data on terrains (on Earth, Mars as well as any other planet) into the rover's memory, test it with samples available or which could be created or imitated in a lab, and then road-test it extensively. The outcome would thus be much less unpredictable than thought, since it is reasonable to assume that such a machine would be built according to the highest degree of scientific knowledge available. From a legal point of view, the application of product liability rules would raise no issues, whereas holding the robot directly liable would solve no problem. The assets the robots may use to compensate for damage caused would in fact have to be provided by the producer, user, owner or any other third party, which would ultimately be the one bearing the consequences of the undesired outcome: see below, section X.

88 This seems to be one of the cases in which Schaerer, Kelley and Nicolescu (n 76) would apply the liability of the keeper of an animal, but such equation is not necessary in order to come to such a conclusion. The same can be said of the lawnmower example discussed at p 76.

89 These are normally the grounds for affirming vicarious liability, such as in a parent-child or employer-employee relationship: see below, section VIII.

leave it to him, rather than the victim, to decide whether to sue the producer in recourse, in all cases where the defectiveness of the vehicle was then ascertained.[90] It should be noted, however, that in such a case, the decision to hold the owner liable would not be influenced by the autonomous nature of the vehicle, but rather by a policy argument quite common in the field of the optimal choice of liability rules.

We may thus conclude that the weak notion of autonomy, unlike the strong one, does not *per se* force a change in the existing legal paradigm; and the ability of a robot to operate unattended and interact in a complex environment does not suffice on its own to ground an argument for changing the existing set of liability rules.

VII. THE ABILITY TO LEARN AND THE RESPONSIBILITY GAP

The notion of learning[91] is itself quite general and needs to be further specified before we can assess whether it actually forces a change in the existing paradigm for the ascription of liability.[92] Most certainly not every learning capacity induces the same kind of considerations, in particular with respect to the problem of (un)foreseeability[93] of the tortious outcome.

In the first place, we need to distinguish what may be considered the mere appearance of a learning process when instead the machine is merely executing a program, even if not originally installed at the point of purchase or release. This will be the case where a robot gains access to a cloud database in order to retrieve information and instructions, if not programs (apps they might be called) in order to perform an additional task—different from the one it was originally programmed for—or rather a further specification of it. A robot cook could download instructions on how to prepare a special recipe its owner just requested as well as the software updates its producer may release in order to fix its functioning bugs. No differently from a computer, a similar machine would not be actively learning but rather applying new software designed for that specific purpose. Such technology would allow flexibility and increase capacities for a specific purpose, yet it would still entail executing a program. In such cases it is safe to assume that no problem would emerge with existing norms (namely product liability): any malfunctioning of the program could be ascribed alternatively to the programmer or to the producer of

90 Under most states' legislation a product will be deemed defective because of either a manufacturing or designing defect, or because of failure to warn of the potential risks associated with its use: see below, sections VIII ff.

91 Hertzberg and Chatila (n 42) 219 define learning 'as the ability to improve the system's own performance or knowledge based on its experience'.

92 So claims Matthias (n 26) 175 ff; Karnow (n 26) 17–18 rather concedes that the problem posed by robots capable of learning could be solved through the evolution of existing liability rules, namely product liability.

93 The problem is identified with the application of the notion of foreseeability, which is relevant to product liability rules, by Karnow (n 26) 14 ff.

the hardware—in particular to the producer if the former carefully conformed to the standards identified by the latter for the design of software to operate on the machine or even authorised its release as an aftermarket product or service meant to be used on its apparatus.

A different form of learning derives from interactions with the external environment: the more complex the robot, the more sensors and actuators it has,[94] the more it will be able to derive information and inputs from the environment in which it operates. According to some authors, a robot's ability to acquire and elaborate data in order to complete its tasks constitutes actual learning:

> Presently there are machines in development or already in use which are able to decide on a course of action and to act without human intervention. The rules by which they act are not fixed during the production process, but can be changed during the operation of the machine, *by the machine itself*. This is what we call machine learning.[95]

Such a notion of learning is, however, extremely wide and needs to be narrowed down, isolating those forms of self-modification of the machine and of its functioning that may increase the level of unforeseeability of the output.

There are two technical approaches to artificial intelligence which need to be taken into account:[96] neural nets and genetic algorithms. The former is the attempt to emulate the functioning of the neural network[97] in a living system. The process of storing information modifies the system itself, and thus data cannot be accessed and controlled or modified at a later moment. A machine so conceived would learn by functioning, as if it was trained through a process of trial and error.[98] In such a perspective, more than the programming phase, the subsequent exploration provides the actual design of the system and cannot be distinguished from it.[99] The latter instead entails a very high degree of self-modification of the machine. Simplifying the principle underlying evolutionary robotics techniques, it could be said that the machine, which is created to accomplish a given task, is the product of a 'repeated process of selective reproduction, random muta-

94 In this sense embodiment plays a central role since a robot given an external body will interact with the surrounding environment more easily and will be able to affect it negatively as well. On this see also Karnow (n 26) 7.

95 See Matthias (n 26) 177.

96 Matthias (n 26) 178 ff identifies four: symbolic systems, connectionism and neural nets, genetic algorithms, and autonomous agents. The first, however, he deems unproblematic since the information pursuant to which the machine determines its operation is 'stored inside the system in the form of explicit, distinct, quasi-linguistic symbols' which thus 'can be inspected at any time and, should need arise, be corrected'. The last category, defined as 'artificial entities that fulfill a certain, often quite narrow purpose, by moving autonomously through some "space" and acting without human supervision', has been addressed in the previous paragraphs.

97 See Hertzberg and Chatila (n 42) 220.

98 See also David E Moriarty, Alan C Schultz and John J Grefenstette, 'Algorithms for Reinforcement Learning' (1999) 11 *Journal of Artificial Intelligence Research* 199 ff.

99 On reinforced learning see Hertzberg and Chatila (n 42) 220.

tion, and genetic recombination'.[100] Here, instead of programming a robot with detailed instructions on how to complete a specific task,

> [a]n initial population of different artificial chromosomes, each encoding the control system … of a robot is randomly created. Each of these chromosomes is then decoded into a corresponding controller … and downloaded into the processor of the robot. The robot is then let free to act … according to a genetically specified controller while its performance for a given task is automatically evaluated … The fittest individuals are allowed to reproduce by generating copies of their chromosomes … The newly obtained population is tested again on the same robot. This process is repeated for a number of generations until an individual is born which satisfies the fitness function set by the user.[101]

According to Matthias,[102] the circumstance that the information stored in an artificial neural network cannot be accessed and controlled at any given moment in time, and the absolute absence of influence in the output obtained through genetic programming methods, cause a fundamental loss of control on the part of the programmer, which makes the attribution of liability unjustified. Said circumstances would therefore highlight the existence of a so-called 'responsibility gap'.

Such loss of control is, however, more apparent than real, being mostly restrained to the design phase. It should in fact be observed that despite allowing a greater degree of unpredictability of the machine's behaviour, such programming techniques mostly influence the conception of the robot more than its day-to-day operation. If a neural network requires training in order to perfect its skills and accomplish a given task, the development phase of the machine ought to include that very training. Once released onto the market the product is supposed to have learnt or perfected a sufficient skill to interact safely, at least as safely as the existing non-robotic—or even non-learning robotic—applications can. This is the case, for instance, with a walking-dog for the blind. Until its training is complete and the dog can perform the tasks for which it is required, it cannot be sold or employed for the assistance of the disabled, and no different kind of reasoning should apply to a robot performing the same task.

Such a perspective is even more evident when one comes to evolutionary robotics. First, the technique is most often confined to laboratory experiments, frequently software simulations of interactions which in reality never occur.[103] Secondly, the purpose of said technique is to develop otherwise unconceived solutions for the functioning of a machine, whose performance is measured against a fitness function: the outcome pursued is the best possible 'individual' for the task, thus not an ever changing or self-modifying application.

100 See Dario Floreano, Phil Husbands and Stefano Nolfi, 'Evolutionary Robotics' in Siciliano and Khatib (n 1) 1423 ff.

101 *Ibid*, 1424.

102 Matthias (n 26) 181 ff. Similar considerations can be found in Karnow (n 26) 4 ff.

103 Floreano, Husbands and Nolfi (n 100) 1428–9.

In nature, in fact, the genetic sequence of a being does not modify itself over time (unless pathologies occur); similarly the algorithm of the single machine, which is a part of an evolutionary robotics study, is given and does not change during the experiment, but is modified through (re)production of a new specimen. Then, once the desired outcome is obtained, the evolution process is deemed complete. In both cases the ability of a robot to modify itself is indeed limited or can be actively limited after the completion of the design phase and before it is released onto the market and commercialised. Therefore, even if designing such applications does not entail coding complex lines of software into a specific language, but rather requires the use of alternative—and to some extent more sophisticated—methods of production, this does not *per se* influence the final consideration that it is the programmer—or *creator* if we want to call it that[104]—who has control over the general/global outcome. It is in fact the producer's decision as to what kind of technique to use in order to achieve the best result possible, both in terms of sophistication and functionality of the robot as well as safety; only the producer could in fact devise and conceive possible methods aimed at preventing damage deriving from the proper—or even improper—use of its product.

In other words, if foreseeability is a matter of experience, and thus of the repetition of interactions between the environment and the machine, great insight can be gained during the testing and development phase by the producer of a robot as of any other kind of technological application.[105]

Finally, even if we assumed that an ability to modify itself was granted to the robot after the moment it was introduced into the market, we still need to consider that this would be the active decision of the producer or programmer to provide its machine with a given capacity. It is clear, though, that such a possibility should only be allowed when it is sufficiently safe to do so, in light of the devices or measures that could be built in (according to existing knowledge) so as to prevent undesired consequences.

To better explain such a concept we may resort to the example of an adaptive elevator using 'artificial neural networks and reinforcement learning algorithms'[106] in order to better assess traffic patterns and minimise waiting periods. Such a robot is not learning to complete any additional tasks other than the one it was conceived for (as the Mars rover described at footnote 87 above), yet it is improving its effectiveness and efficiency over time. In this example, the robot 'leaves an important executive waiting for half an hour in the 34th floor, so that he cannot attend a business meeting'[107] and therefore damage results as a consequence thereof. Yet, holding the producer liable in such cases appears to be a satisfactory and straightforward solution, at least based on two different considerations. First, it is foreseeable that the patterns of use of an elevator in a

104 Matthias, 'From Coder to Creator' (n 87) 175 ff.
105 Karnow (n 26) 18.
106 The example is given by Matthias (n 26) 176.
107 *Ibid*, 176.

large building with different kinds of offices and business hours as well as with different kinds of users pursuing various occupations may vary over time. Therefore, the program that allows the elevator to learn should enable it to identify potential outliers (say, for instance, a conference attracting a vast number of individuals to a particular floor for a limited number of days) and not base its decision entirely on them; that is to say, the elevator ought not to be made an 'inductive turkey'.[108] Secondly, the producer ought to have assumed that exceptional circumstances may occur where it is necessary to be able to quickly and safely override the program and call the elevator at need. Such a 'safety device' should be embedded in the application and the producer ought to be held liable for not having conceived it.

From a philosophical perspective, we may say that the ability to learn is a choice the agent made for the machine; the subsequent behaviour is therefore heteronomously determined by the 'creator' who caused the robot to be what it is, and have those abilities, which were originally allowed or conceived, irrespective of how they have evolved. From a legal point of view, we need to stress that foreseeability is a broad concept which can adapt over time, through technological evolution and the assessment of standards by competent technical bodies and agencies, and can thus effectively accommodate such kinds of applications.[109] Therefore even the ability to learn does not suffice *per se* to justify a change in perspective when addressing the regulation of robotic applications.

VIII. THE ADEQUACY OF (PRODUCT) LIABILITY RULES: CONTROL

It has been shown that so long as robots do not achieve self-consciousness they cannot be deemed moral agents or autonomous—in a strong sense—beings. Short of that capacity there is no logical, moral or philosophical—and thus not even legal—necessity to consider them subjects of law and bestow individual rights on them. Therefore, all existing robots up to that point are to be deemed objects—more precisely, artefacts created by human design and labour, for the purpose of serving identifiable human needs, otherwise known as products.

Even the abilities to intelligently interact and to learn do not identify a subject whose actions could be considered the consequence of self-determination and awareness, despite——at least in some cases—being autonomous and not predetermined in their

108 This is the typical example derived from Bertrand Russell, *The Problems of Philosophy* (Oxford University Press, Kindle edn 2001) pos 879, who used a chicken to criticise inductivism. The turkey, having been fed on a daily basis, infers that it will continue to be fed in the future, yet the day comes for it to be butchered. The sample of data the turkey bases its judgment upon is in fact limited and does not take into account the possibility of unexpected events, also known as black swans: see Nassim Nicholas Taleb, *The Black Swan* (Random House, 2007) 40 ff.

109 See Karnow (n 26) 17 ff.

actual content by the agent.[110] From a legal point of view, this forces us to conclude that the natural paradigm within which to frame issues of liability involving robotic applications is that of product liability rules.

Some authors find those rules to be inadequate and identify an ever widening 'responsibility gap'[111] due to the absence of control on the part of the producer or programmer[112] over the actions of sophisticated machines, which should encourage the attribution of legal personhood to robots as has been done to corporations.[113] Before moving on to evaluate the effectiveness of the suggested alternative—namely the awarding of legal personhood to the robot—it is necessary to tackle the *pars destruens* of the claim. Said conclusion is in fact based on the assumption that 'control' is the essential requirement for the attribution of liability, this being understood as the possibility to supervise or directly determine the behaviour of the party causing harm. Yet modern tort law theory, as well as existing legislation, shows that the duty to compensate is not always grounded in the ability of the individual to directly determine or prevent the harmful consequences that may occur.

Law and economics literature has long pointed out that the legislator's decision to acknowledge an entitlement of a party through the adoption of a liability rule ought to depend on the assessment of different criteria:[114] the ability of the tortfeasor to prevent a given harm is taken into account, together with other conditions such as the ability of the same party to lower transaction costs,[115] or his ability to insure against said damages, as well as to reduce administrative costs.[116] At the same time, each legal system knows some forms of vicarious liability where the party that is held responsible for the acts of someone else does not necessarily exert a direct control[117]—neither on the facts which led to the damage nor on the actions of the tortfeasor himself. The liability of the employer for the acts of his employees is only at times explained through the *fictio* of a

110 See also Koops, Hildebrandt and Jaquet-Chiffelle (n 40) 515, who distinguish between 'Automatic agents', automated non creative applications, 'Autonomic agents', which present the capacity to initiate a change in their own program to achieve a goal, and 'Autonomous agents', being 'capable of living up to [their] own law'. The 'middle kind' could still show a high degree of autonomy and yet would not amount to a being whose constitutional rights should be acknowledged (532).

111 Matthias (n 26) *passim*.

112 It is sometimes discussed how liability should be apportioned between producer and programmer. Such a distinction appears to be trivial for the sake of the current analysis. In the first place, the two subjects may coincide; otherwise their internal relationship may be regulated pursuant to a contract that the two parties entered into. Since the issue addressed here is rather whether it is the 'human behind the machine' or the robot itself that ought to be held liable, it suffices to refer to the one or the other—which of the two would be called to compensate is a matter of fact to be addressed in the particular case.

113 See Leroux *et al* (n 25) 60 ff.

114 Ronald Harry Coase, 'The Problem of Social Cost' in Ronald Harry Coase (ed), *The Firm, the Market, and the Law* (University of Chicago Press, 1990) 95 ff.

115 See Guido Calabresi and Douglas A Melamed, 'Property Rules, Liability Rules, and Inalienability: One View of the Cathedral' (1972) 85 *Harvard Law Review* 1089, 1096–7.

116 See Steven Shavell, 'Liability for Accidents' in Mitchell A Polinsky and Steven Shavell (eds), *Handbook of Law and Economics*, vol I (North-Holland, 2007) 149–50.

117 At least not in the terms identified by Matthias (n 26) 175.

culpa in eligendo—namely the fault of selecting the given collaborator—but does not really depend on the assessment of negligence in any form, since the employer cannot be reprehended for doing something that caused the specific negative outcome. Instead, because he is gaining a return from employing other individuals who contribute to the running of his enterprise, he is also called upon to bear the costs that those persons impose on third parties when operating in his interest. The way he exerts control over his personnel is certainly mediated: he may impose the adoption of safety measures, which are required by the law or deemed necessary in light of the specific activity being carried out, and he may in some cases supervise performance and enforce the observation of necessary procedures; yet proving that he did all that was required would not set him free of liability if harm was caused by the employee acting within the scope of his employment—and, it should be said, the level of autonomy of the actions performed by an adult is certainly greater than those of existing or reasonably foreseeable robots. Similarly, the liability of parents for the acts of children, where admitted, is not justifiable by reference to the direct control they exert on their offspring. Parents may in fact educate their children and thus influence their character, but they surely do not supervise them constantly, nor could that be demanded of them; at the same time, children's actions are almost by definition unforeseeable and difficult to anticipate, much more so than those of a sophisticated robot. So even if it is required that the parent be living with the child for him to have a chance to influence his behaviour, the fact that at the moment the damage occurred he was on holiday in a different location does not exclude the adult's vicarious liability.[118]

Therefore if, on the one hand, there are different considerations which are implied in the choice of attributing liability to a given subject, which go far beyond the appreciation of a form of control, on the other hand there are very different notions of that term which may be inferred by reading existing law. Control may be direct, in the sense that it is the person called upon to compensate for harm that can avoid the negative consequence by intervening at the moment harm is caused or at an earlier point in time along the causal line which led to the event; or indirect, as in cases of vicarious liability, where the party may only in a mediated and more remote way intervene in order to lower the chances of harm—say by training his employees or providing them with adequate equipment, or by educating children to behave according to society's rules. In other words, the attribution of liability from a private law perspective is merely the shifting of a cost from one side to another, and moral considerations do not always coincide with the outcome produced by the application of existing norms.

Finally, despite it not being possible to conduct an in-depth analysis of product liability rules, some fundamental aspects of it will be sketched, starting with its implied rationale. David Owen has defined the object of this field as the study of the relationship between the maker and the victim of a product, since

118 So held the Italian Corte di Cassazione in decision Cass, 9 June 1976, n° 2115.

> [b]y choosing to expose product users and others to certain types and degrees of risk, manufacturers appropriate to themselves certain interests in safety and bodily integrity that may belong to those other persons. Similarly by choosing to make claims against manufacturers for harm resulting from such risks or uses victims of product accidents seek to appropriate to themselves economic interests that may belong to manufacturers and other consumers.[119]

So conceived, product liability rules aim at balancing opposite interests: having safe products[120] and actually distributing them in the market for profit. The theory of recovery is often claimed to be strict liability,[121] but a more careful reading actually shows a mismatch with the daily application of those principles under American common law, and to some extent even in European law.[122]

For the purposes of the current analysis it could be stressed that with a strict liability rule such as the one affirmed by Article 1 of the EU Directive[123] or section 402A of the Restatement (Second) of Torts, the fundamental ground for holding the manufacturer liable could not be identified with the ability to increase the product's safety to a desirable standard; control over the design and production would therefore not fully explain the attribution of liability.[124] Instead, such a choice could be justified based on the ability of the producer to (better) insure himself and therefore handle costs—including transaction costs—associated with the distribution of the product in an aggregated, and thus more efficient, fashion.[125]

The actual nature of product liability rules is, however, extremely complex. Provisions such as those holding retailers liable under EU law[126] as well as case law affirming

119 David G Owen, *Products Liability Law* (Thompson West, 2nd edn 2008) 7.

120 This is also the declared purpose of EU Directive 85/374/EEC, Liability for defective products, as subsequently modified ('the EU Directive' or 'the Directive').

121 See s 402A Restatement (Second) of Torts, as well as Art 1 of the EU Directive; see also Carlo Castronovo, *La nuova responsabilità civile* (Giuffrè, 2006) 687.

122 Trying to summarise the two regimes in a comparative perspective is an impossible task, and to some extent lies beyond the purposes of the current analysis. It would be necessary to consider not only the case law of every North American state, as well as its regulation, but also the different ways in which the EU directive has been enacted within each legal system of the member states as well as its current application in national courts. Here, instead, the aim is rather to show that existing norms present a certain degree of elasticity, which could accommodate most of the problems and issues arising from the introduction of robotic applications in our society.

123 For discussion of this issue, see Jürgen Oechsler, *Staudinger Kommentar BGB §§823–829 ProdHaft.G.* (2003), 826 ff.

124 If the producer is held liable even in cases where it could not be deemed negligent, then such cases could not provide an additional incentive in improving the overall product's safety. Instead, they would simply increase the market price of the product, by imposing upon the producer the requirement to buy additional insurance for harm which may arise from the normal use of its device. See Richard Posner, *Economic Analysis of Law* (Wolters Kluwer, 7th edn 2007) 182.

125 See the considerations of Castronovo (n 121) 687 ff. Who is actually called upon to bear the costs of such measures is ultimately determined pursuant to the elasticity of the demand curve of the given product. The producer may in fact transfer partially or integrally such costs to the end user through the product's price.

126 See Art 3 EU Directive, stating: 'where the producer of the product cannot be identified, each supplier of the product shall be treated as its producer unless he informs the injured person, within a reasonable time, of the identity of the producer or of the person who supplied him with the product.'

producers' strict liability for manufacturing defects under US law[127] point to a pro-consumer perspective, where the ability to influence the harmful outcome is quite secondary as opposed to the intent to compensate all harm suffered by using the product. At the same time, the most frequent application of the criterion of foreseeability in design and warning defects[128] takes into account technological development and actual knowledge at the moment of production in order to determine whether the party is to be held liable. Indeed, the required level of safety is determined through an evaluation which closely recalls a standard of care judgment[129] in most US courts and to some extent in EU courts as well. While control cannot be identified as the major or only criterion for the attribution of liability for defective products, it of course remains important that any allocation of responsibility should not be unjust; this implies that the effective, efficient and desirable nature of the incentive to produce only safe products needs to be assessed in light of the specific circumstances. In order to understand the capability of existing norms and criteria to accommodate scientific and technological development, the criterion of foreseeability and the development risk defence need to be briefly described.

IX. FORESEEABILITY AND THE DEVELOPMENT RISK DEFENCE: THE ELASTICITY OF THE EXISTING NORMATIVE FRAMEWORK

Traditionally, US product liability law developed around three different concepts of defect: (i) manufacturing, (ii) design and (iii) warning defect, one of which needs to be present in order to ground an action for the recovery of a suffered harm. Defectiveness is understood as an intrinsically excessive risk arising from the use—and to some extent misuse as well—of the product,[130] and ought to be proved through expert testimony, be it that of an engineer or any other sort of technical expert.[131] While manufacturing as well as warning defects do not pose any problem peculiar to robotic applications, since they could be successfully tackled as with any other product, a closer reading is required for design defect, which entails that the given risk could have been reduced through 'a reasonable alternative design',[132] often leading to the application of a cost-benefit analysis such as that set out by the so-called Learned Hand formula.[133] The point, though, is

127 Pursuant to the Restatement (Third) of Torts: Products Liability a product contains a manufacturing defect if it 'departs from its intended design even though all possible care was exercised in the preparation and marketing of the product'.

128 The distinction is of essential relevance for US law and has more of a descriptive relevance under EU law (see Art 117 of the Italian Consumers Code).

129 See Owen (n 119) 34 and 71, and below, section IX.

130 *Ibid*, 34.

131 The leading case under US law is *Daubert v Merrel Dow Pharmaceuticals Inc*, 509 US 579 (1993).

132 Restatement (Third) of Torts: Products Liability, s 2(b). For a brief discussion see David G Owen, John E Montgomery and Mary J Davis, *Products Liability and Safety,* Statutory Supplement (Foundation Press, 5th edn 2007) 40 ff.

133 It is the so-called BPL analysis: see Posner (n 124) 184.

to determine the risk in respect of which a deviation from a demandable standard of care ought to be measured, ultimately the only true limit to the imposition of liability on the producer. The criterion to be used for this purpose is that of foreseeability.[134]

The abovementioned theories consider the behaviour of the robot to be unforeseeable if it was autonomous or due to a learning capacity. Yet it has been shown that a *weakly* autonomous behaviour is not *per se* unforeseeable since a robot performing a program—even when that leaves a wide variety of choices and alternatives among which to choose—is still following the instructions it was loaded with, and assessing the variables it was allowed to estimate, through the sensors and actuators it was provided with. Similarly, the ability to learn does not change the overall paradigm since such a capacity can either pertain to the design phase, and thus be denied to the machine at the later stage when the product is introduced into the market, or—even when allowed thereafter—it can be limited in such a way as to keep it under control, or within safe boundaries. In any case such a notion of unpredictability does not necessarily equate to that of unforeseeability for the purposes of product liability rules. The latter in fact refers to those risks which, because of existing knowledge and available scientific data, cannot even be perceived as such (we may call them unknown unknowns) as well as those uses which are absolutely remote and cannot be anticipated; everything else is by definition foreseeable. That of course does not suffice to ground liability since the plaintiff will have to prove that an alternative design could have been imposed to make the product safer.

In the examples above, the fact that the elevator may induce wrong conclusions based on a temporarily altered flow of users is realistic and easy to foresee, and likewise the fact that a driverless vehicle may find a pedestrian jaywalking in the traffic is very easily imagined. Whether or not the safety conceived suffices in order to exculpate the producer is simply a matter of fact, where opposing criteria need to be assessed, among which is the cost of designing the machine otherwise and safer. Such a judgment, being determined *ex post* and in the specific circumstances, allows a higher degree of flexibility than that of a specific normative standard; moreover, evolving over time,[135] it could more effectively accommodate *ex ante* unanticipated technical advancements.

Overall, it should be stressed that the ability of the producer of robotic applications to be held liable in cases where the potential harm was foreseeable simply forces this party to internalise the costs of its business choices. Therefore, when designing a robot, if a specific risk in the usage of the machine could be anticipated, the producer would be bound to conceive a safety device to prevent or reduce the risk of actual harm resulting from it, and when that was not currently possible the decision to provide the robot nonetheless with the same capacity should lead to the assessment of the duty to compensate for damage.

134 *Feldman v Lederle Laboratories*, 479 A2d 374 (NJ 1984).
135 Karnow (n 26) 17–18.

The same conclusions may be reached where European regulation sets the standard. Despite the Directive not presenting the same distinctions among different kind of defects,[136] Article 6, when enumerating the circumstances that need to be taken into account, specifically mentions '(a) the presentation of the product; (b) the use to which it could reasonably be expected that the product would be put; (c) the time when the product was put into circulation'. Foreseeability therefore matters, even if the way each national court interprets the notion of reasonable expectations may vary, often emphasising the objective aspect (close to a risk-utility test) over the subjective one (consumer expectation test).

Finally, it should be stressed that the different standards of safety required from products can—at least in part—be determined by decisions of specific bodies, called upon to define the characteristic a given object needs to present. Such criteria are normally taken into account by the courts[137] as relevant evidence of the intrinsic quality of the product; yet compliance with said standards does not *per se* exclude liability,[138] since those are usually intended as minimal requirements merely allowing the distribution of the product on the market.[139] A compliance defence (such as the one set forth by Article 7(d) of the Directive) would in fact normally exclude liability only if the damage occurred because of the specific feature set forth by the legal rule, and not simply because of the (mal)functioning of the good, which otherwise conformed to all legal standards provided for by technical regulations.At least as relevant for the purposes of the current analysis is the so-called 'development risk defence' admitted by Article 7(e) of the Directive, which states that the producer shall not be held liable when 'the state of scientific and technical knowledge at the time when he put the product into circulation was not such as to enable the existence of the defect to be discovered'. The application of the rule may vary in each member state but still softens the strict standard set forth by Article 4, thus completing the overall picture: only reasonably foreseeable uses need to be taken into account when devising the product, and dangers falling beyond existing scientific knowledge may not be imposed on the business.

This very limited survey of the fundamental principles of product liability legislation aims to show how, from a legal point of view, existing norms may successfully accommo-

136 The very notion of defect appears to be devoid of any additional meaning: the duty to compensate arises whenever the product does not provide the demandable degree of safety, and thus is a mere duplicate of the notion of risk (compensatory damages cannot in fact be awarded when a defect arises which simply makes the thing purchased unsuitable for the desired use). See Castronovo (n 121) 692–3.

137 See *Doyle v Volkswagenwerk AK*, 481 SE 2d 518, 521 (Ga 1997).

138 For an analysis of the grounds which militate against the general recognition of a compliance defence, see Owen (n 119) 930 and 936.

139 See also Restatement (Second) of Torts, s 288C, which states: 'Compliance with a legislative enactment or an administrative regulation does not prevent a finding of negligence where a reasonable man would take additional precautions.' But see, for instance, *Southland Mower Co v Consumer Prod Safety Comm'n*, 619 F2d 499 (5th Cir 1980), in which the technical standard adopted by the regulation rendered it immaterial that a different design would have been safer. This, however, was a case where the standard set was extremely narrowly tailored (number of seconds before the blade of a mower was to stop).

date the diffusion of robotic applications in society, in particular thanks to the elasticity of criteria such as foreseeability and the development risk defence. At the same time, this does not necessarily mean that reasons could not be found to justify, with respect to a single application (or set of applications), the preferable nature of another rule or even an exemption; yet such a choice ought to be specifically justified. In other words, we may conclude that there is no general and common reason to relax safety rules and principles when discussing robotic applications, and neither their capacity to operate autonomously nor their ability to learn provide sufficient grounds: quite the contrary.[140]

X. ROBOTS AS LEGAL PERSONS

Stating that robots do not amount to autonomous beings and thus should not be recognised as subjects of law does not otherwise imply that legal personhood could not be awarded for functional reasons as it is to corporations. In such a perspective, though, a specific end needs to be identified, and alternative tools ought to be taken into account before concluding that such would be the preferable way to achieve the desired result. It may indeed prove useful to attribute legal personhood to a software agent,[141] which would then be registered, so as to identify the limits of its ability to validly conclude contracts, the maximum amount of the obligations it could assume, and eventually the (physical or legal) person it is representing.

With respect to liability issues, the recognition of personhood would mainly serve as a liability capping method; yet it would neither necessarily change the person bearing the costs of its functioning nor the cases when compensation is awarded. In fact, unless the robot was capable of earning a revenue from its operation, its capital would have to be provided by a human, or a corporation, standing behind it, thus not necessarily shifting the burden from the party that would bear it pursuant to existing product liability rules.[142] A similar if not identical result could be achieved with an—eventually even compulsory third-party—insurance mechanism or with a simple damages cap such as

140 The standard of safety that a robotic application should be bound to observe is the same as that applicable to a non-robotic application.

141 See Tom Allen and Robin Widdison, 'Can Computers Make Contracts?' (1996) 9 *Harvard Journal of Law and Technology* 26; Steffen Wetting and Eberhard Zehendner, 'A Legal Analysis of Human and Electronic Agents' [2004] *Artificial Intelligence and Law* 111. Gunther Teubner, 'Rights of Non-Humans? Electronic Agents and Animals as New Actors in Politics and Law' (2006) 33 *Journal of Law and Society* 497 justifies the attribution of personhood to software agents through the creation of hybrids of such entities with humans. The purpose, though, still seems to be functional, thus to simplify economic interactions and the adaptation of contract rules.

142 Quite to the contrary, if the robot was responsible for all damage it caused in an objective fashion—since it would be purposeless to impose a simple standard of care on the machine, as it would not modify at all the end result achieved today through the application of a rule of negligence to the producer—then the person financing the robot's fund would be correspondingly worse off. He would in fact be held liable objectively, without any possibility of freeing himself.

that admitted by Article 16 of the Directive.[143] But, even if the robot was allowed to earn a fee for its performance, this would only constitute a tax on the user, producing an overall risk-spreading effect which could be effectively achieved otherwise, for instance through the adoption of a no-fault scheme funded by the product's users in various fashions.[144] Which of the different alternatives is preferable is still a matter of correctly specifying particular circumstances, among which are the size of the market for the given application and the existence of evident failures which could be designed around through *ad hoc* regulation; much less would depend on the machine being weakly autonomous or even able to learn.

XI. A FUNCTIONAL APPROACH TO THE ISSUES OF LIABILITY IN ROBOTS

The conclusion that robots ought to be deemed products and that existing rules are not altogether inadequate to address issues of liability involving the operation of sophisticated machines does not imply that said rules provide the correct and desirable incentives in every case. On the contrary, product liability rules do appear to be ineffective in some cases, for diverse and even opposite reasons according to each legal system.[145] But, of course, most of the law's inadequacies are not peculiar to robotic applications, since similar claims can be upheld when addressing other kinds of devices and technologies. Therefore, even when it comes to robotics, other criteria, besides technical aspects, need to be taken into consideration: market failures, the limited number of potential users, the desirability of a specific application, and existing legal—and mostly constitutionally relevant—principles may lead to proposals for alternative compensation methods. For a merely descriptive purpose some examples can be sketched.

143 Stating: 'Any Member State may provide that a producer's total liability for damage resulting from a death or personal injury and caused by identical items with the same defect shall be limited to an amount which may not be less than 70 million ECU.'

144 A no-fault scheme is an automatic compensation system which does not require the ascertainment of fault or a defect, but simply that harm occurred in a specific circumstance involving, in the case in point, the use of a robot. Such a system reduces the administrative costs associated with compensation, yet may trigger moral hazard—though the system could be conceived in a way that might compensate for such side effects. For discussion and further references see Giovanni Comandè, *Risarcimeno del danno alla persona e alternative istituzionali* (Giappichelli, 1999) 333 ff.

145 Under American common law it is normally the high litigation costs as well as excessive awards of damages which are mentioned as hindering the system, allowing distortions and—at times excessively—burdening the defendant in product liability cases: see Owen (n 119) 25 ff. Under EU law, the very limited number of cases decided pursuant to such regulation is at times mentioned to draw the conclusion that it is ineffective and fails to provide adequate protection for the consumer. See eg Sara Biglieri, Andrea Pupeschi and Christian Di Mauro, 'The Italian Product Liability Experience' in Dennis Campbell (ed), *Liability for Products in a Global Economy* (Kluwer, 2005) 21.

Robotic prostheses could actually be of interest, despite them being neither autonomous nor capable of learning. Indeed, the argument for the adoption of a different compensation system would take account of the limitless situations and ways in which the prosthesis could be used,[146] and the excessive burden that a strict liability rule would impose on the producer, also considering the very limited market for potential users.[147] Grounds for favourable discrimination could be found in fundamental rights and constitutional principles, such as those set forth in national constitutions,[148] the Lisbon Treaties,[149] and the UN Convention on the Rights of People with Disabilities, which may be read as requiring the adoption of affirmative measures in this respect.[150] The development of such devices should in fact be favoured because of their intrinsic value for the purpose of helping the disabled integrate into society and achieve a higher quality of life, increasing their independence.

A driverless vehicle, by contrast, having a potentially large market[151] and performing a very specific task, may be deemed to be like any other product and not so substantially different from its non fully autonomous alternative; at the same time, it could certainly be considered a desirable technology from a public policy perspective, capable of improving mobility for a large share of the population, ranging from children to the elderly and disabled. For those reasons, ad hoc liability schemes may be conceived so as to favour their development and diffusion,[152] such as a compulsory third-party insurance imposed on the owner, who may be called upon to compensate, irrespective of whether

146 Given that a prosthetic limb is attached to the person, it accompanies the individual everywhere he goes. Patients show a great capacity to adapt to the use of such limbs and—for instance in the case of a hand—may actually learn to use it in a way not previously conceived of in order to complete additional tasks and improve their quality of life. Yet the malfunctioning of the limb may bring about very different consequences. If a prosthetic hand malfunctions while the person is carrying a shopping bag, the eggs inside may break; if the same occurs while it is lifting a weight in the gym he might be seriously harmed; if while driving, third parties may be injured or killed.

147 There are fortunately fewer than two million amputees in the US, and not all of them would qualify for the use of robotic prostheses: see www.amputee-coalition.org/fact_sheets/amp_stats_cause.html.

148 For instance Arts 2, 3 and 32 of the Italian Constitution, the latter granting the right to health and free medical assistance for those who cannot afford it.

149 In particular Arts 1, 3 and 26, the latter stating, 'The Union recognizes and respects the right of persons with disabilities to benefit from measures designed to ensure their independence, social and occupational integration and participation in the life of the community', as well as Art 35, affirming that 'Everyone has the right of access to preventive health care and the right to benefit from medical treatment under the conditions established by national laws and practices. A high level of human health protection shall be ensured in the definition and implementation of all the Union's policies and activities.'

150 The Convention even affirms at Art 4(g) that in order to promote the full realisation of all human rights and fundamental freedoms, the signing parties 'undertake … to undertake or promote research and development of, and to promote the availability and use of new technologies, including information and communications technologies, mobility aids, devices and assistive technologies, suitable for persons with disabilities, giving priority to technologies at an affordable cost'.

151 There are more than 800 million vehicles on the planet, and this figure is expected to double in the next 10 years: see Broggi *et al* (n 79) 1176.

152 See *ibid*, 1178, claiming that, because the current legal system prevents the creation of truly autonomous vehicles, man is necessarily left in control.

or not he was personally using the vehicle. It may then be discussed whether, at a later moment in time, such a party ought to be allowed to sue the producer in recourse—and under what conditions—where harm derives from the malfunctioning of the robot. A compliance defence could in fact be specifically granted, which would otherwise neither exist nor be justifiable as a general principle for all other applications indistinctively.[153] The choice among possible solutions should, however, be carefully determined through clear policy consideration, as well as an analysis aimed at weighing the efficiency of each measure to be adopted.

At the same time an autonomous vacuum cleaner could effectively be dealt with like any other non-robotic household device, leaving the legislator indifferent to it.

XII. CONCLUSIONS

A meaningful legal analysis of robotic applications should not rely on science fiction. Neither inevitable uncertainty nor serendipity, which causes future developments in robotics to be to some extent unforeseeable, should be compensated by looking at literary depictions. There are two essential grounds that induce such a conclusion, and one is its major consequence.

First, the disastrous accounts where a programmer sets up 'an evolutionary system whose limitations are to him unclear and possibly incomprehensible'[154] are not only pessimistic—denoting very little faith in the abilities and skills of current and future engineers and scientists—but clearly unjustified. Such an outcome may be precisely averted through regulation—and liability rules do play a central role to this end—which ought to provide criteria for what is desirable and what is not, what can be allowed and to some extent facilitated, and what should rather be opposed. More broadly, should we ever end up in a world ruled by robots who submit human beings to their desires it will be because of actual choices made all along the way, which are definitely not inevitable. In other words, there is an opportunity and a need for regulation of technological development. Such regulation may in some cases prevent the development of undesired technologies,[155] or simply pose limits to its usage, and in other cases it may provide needed incentives for those applications to come into existence.

153 As a condition, though, the *ex ante* conforming standard relative to which the defectiveness of the vehicle is to be measured should be narrowly specified by a technically competent and independent authority, should impose a high level of safety, and should be periodically updated according to technological developments. Such a solution, coupled with compulsory periodical check-ups for each machine, may prove a plausible alternative to the single kind of product, trading *ex post* (un)certainty for higher *ex ante* costs. See also above, n 138.

154 Lehman-Wilzig (n 5) 446.

155 The question of whether or not to allow the development of completely autonomous robots is first of all a political decision, which national as well as supranational authorities may be called upon to take. Decisions in that respect may also vary from one country to another as energy policies currently do, for instance with respect to the use of nuclear power. A similar analogy can be found in Stefano Rodotà, *Il diritto di avere*

Secondly, science fiction attracts attention to very remote issues, triggering feelings of uneasiness and fear that actually harm the possibility of developing useful applications today.

As a consequence, then, given the a-technical nature of the term 'robot', legal analysis should address issues relating to specific robotic applications—or in some cases classes thereof—separately, identifying the fundamental aspects that trigger the need for a change of existing regulation, in a *de iure condendo* perspective.

For the more limited scope of the issue at stake—namely liability rules—neither autonomy nor the ability to learn suffices *per se*. Indeed, if machines became autonomous in a strong sense, then they would stop being objects and become necessarily subjects of law. Such a possibility, though, is not only remote because of the constraints of existing technological and scientific knowledge, but does not appear to be desirable for the reasons sketched, and should therefore be opposed. In any case, truly autonomous applications appear to be the upper limit, which should only be considered for the purpose of drawing a line where the perspective would actually shift. Yet, taking such an extreme example off the table would benefit legal analysis, separating science fiction from reality.

Below such a line, though, there are no alternative technical grounds which force a change in the assessment of liability for damages involving the use of robots. On the one hand, a weakly autonomous robot is in fact merely behaving, thus executing a program, no matter how sophisticated, and performing tasks it was designed to perform by its creator; on the other hand, the ability to learn is often conceived as an alternative method for elaborating solutions that would otherwise not have been conceived.

Finally, if a robot was allowed to modify itself through interaction with the environment after the moment it was released onto the market it would still be exerting a capacity it was given by its creator, who should only allow this when sufficient safety features that prevent foreseeable harm can be embedded.

At the same time, those aspects do not provide a compelling argument for the relaxation of the burdens imposed on producers by existing norms. Innovation and safety need to be balanced out, and there is no one one-size-fits-all rule[156] that could be indifferently applied to machines so diverse from one another as a driverless vehicle, a prosthetic hand, a vacuum cleaner, and a softbot.

diritti (Editori Laterza, 2012) 369–70. Moreover, different rules may be applied to research in sensitive areas (eg potentially leading to the development of certain kinds of technologies), as is the case today with animal experimentation, denied in some cases, or allowed upon the meeting of certain conditions in others.

156 For a similar view, see Calo (n 1) 603 ff. The author then suggests that robot manufacturers be offered a limited immunity for the potential misuse of their products, in particular in order to foster the development of an open kind of robotics. If the claim can be generally shared, each single kind of application needs to be specifically considered in order to ponder whether the proposed solution is suitable and desirable. After all, the example addressed by Calo, namely the aviation industry, does not necessarily correspond with the entire industry for robotic applications, but maybe with a part thereof.

Applicable liability rules, both negligence based and strict, potentially have the necessary level of elasticity to accommodate existing and reasonably foreseeable applications, operating such balancing *ex post*. Case law over time may help to clarify required safety standards, and soft law criteria adopted by national and international authorities could play an essential role in such a perspective by providing some technical guidance and a higher degree of *ex ante* certainty for researchers and developers of applications. It is clear, however, that the degree of safety demanded of robots should be equal to that demanded of corresponding, if existing, non-robotic technology.

While mere technological aspects might not suffice to justify a change in existing rules, other reasons can be found through constitutional law and public policy considerations. As and when deemed desirable, liability may be limited or shifted in different ways, according to what is more efficient and effective in the specific case: a damage cap, a no-fault plan, a normative exemption, or the awarding of legal personhood. The analysis of robotic technologies should thus concentrate precisely on identifying such grounds, and on devising those alternative compensation methods, which would provide correct incentives for the specific kind, or class, of applications.