Robotics: the attribution problem

Agustín Varela

Abstract

This paper examines the theoretical debate between professors Ryan Calo[1] and Jack M. Balkin[2]. The former wrote Robotics and the Lessons of Cyberlaw[3], while the latter published The Path of Robotics[4]. In his article, Balkin aims to refute some of Calo's ideas, while at the same time he places value on others. The focus will be on robotics considered as a new social and legal phenomenon, primarily regarding the liability for harmful results. Within this subject, I will emphasize on criminal liability in order to try to answer some of the questions from the specialized literature.

1. Introduction

Robots are not a new concern for humanity. Since the Industrial Revolution, and even before, humans have created machines and integrated them to the productive system. The recent interest in robotics can be explained because of the particular moment we are in: robots now interact with other technologies and are being used increasingly for more tasks. Robots are no longer confined to the factories but involved in people's daily life; therefore, they will constitute an intermediate category between an object and a social actor[5]. Humanity will enjoy the robots' creations, and at a certain moment their level of cooperation may be indistinguishable: thus, society will evolve by the common effort among human beings and robots. In this regard philosophers, ethics professors, law scholars, law makers, producers, engineers and programmers ask themselves for the new challenges raised by the robotics phenomenon.

Certainly, the task of defining what a "robot" is, is not easy. Ryan Calo says that they are "*mechanical objects that take the world in, process what they sense, and turn it act upon the world*"[6]. Others precise the term as a "*physical machine which is aware of and able to act*

---

[1] Lane Powell and D. Wayne Gittinger Associate Professor of Law, University of Washington School of Law

[2] Knight Professor of Constitutional Law and the First Amendment at Yale Law School

[3] Ryan Calo, "Robotics and the Lessons of Cyberlaw", *California Law Review*, Vol. 103:513 2015

[4] Jack M. Balkin, *The Path of Robotics Law*, 6 CALIF. L. REV. CIRCUIT 45, 59 (2015)

[5] This concept is taken from Ryan Calo, who talks about a new ontological category for robots in Calo, *op. cit.* 2015, p. 532

[6] Calo, *op. cit.* 2015, p. 529

upon it surroundings and which can make decisions"[7]. Some authors consider autonomy as a synonym for robots[8]; nevertheless, I personally think that autonomy is a category within the larger class of robotics. Autonomous robots, also called *smart robots*, are those that can take and execute decisions in the real world without an external input. Meanwhile, Hilgendorf defines autonomous system as a *"technical system that can cope with problems intelligently in various situations, without having to rely on human input*"[9].

This paper focuses on smart or autonomous robots as considered by the ISO standard definition 837 (2012): "*robot capable of performing tasks by sensing its environment and/or interacting with external sources and adapting its behavior*".

2. Calo and Balkin. Conceptual differences

Ryan Calo wrote Robotics and the Lessons of Cyberlaw with the purpose of finding the common ground and the differences between robotics and cyberlaw. His goal was to treat their parallels in the same way and to seek solutions for their distinctive features.

This author talks about the *sense-think-act* paradigm[10], according to which robots are machines with three distinct characteristics: they can sense the external world, process the information they sense and act in a direct way upon the world. Naturally, the Washington professor draws from the conviction that the most salient feature of the robots, as compared to earlier machines, is their embodiment[11]. This notion is precisely what gives robots the capacity to "act" upon the world.

Following Calo, robots are embodied objects with the ability to proceed in a tangible way. This characteristic is specially relevant for the author, since he considers it enables them to cause physical harm[12]. The author lists their embodiment next to other two salient characteristics: their emergent behavior and their social valence.

---

[7] That is the definition brought by a legal study commissioned by the European Parliament's Legal Committee, p. 12, available at http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf

[8] That is the reach that both Calo and Balkin seems to propose for the term "robot"

[9] Hilgendorf, Eric, "Automated Driving and the Law", in Hilgendorf, Eric and Seidel, Uwe (eds.) *Robotics, Autonomics, and the Law*, Nomos Verlagsgesellschaft, 2017, p. 172

[10] Calo, *op. cit*. 2015, p. 529

[11] Calo, *Id*. at p. 532, Cf. Bottalico, Barbara and Santosuosso, Amedeo, "Autonomous Systems and the Law: Why Intelligence Matters" in Hilgendorf, Eric and Seidel, Uwe (eds), *op. cit.* 2017 (Fn. 9), p. 33

[12] Calo, *op. cit.* 2015 (Fn. 3), p. 53

By the concept of "emergent behavior", Calo refers to the robots' capacity to learn from previous behavior[13]. This implies that they will have a behavioral pattern determined by their programming, but at some point they will start integrating these aptitudes with the "experience" acquired in the real world and, as a consequence, their conduct will differ from the initial programming. This may be the robots' most important feature in legal terms; and it will be further analyzed.

"Social valence", on the other side, refers to a human characteristic: the cognitive bias that makes humans to feel that robots are like people. We perceive robots as if they were people, or at least we feel them more people alike than we do with other objects. Robots blur the line between animate and inanimate and that's why Calo says we need a new ontological category for them:[14] in the same way that Internet invoked a new sense of place, robots fill a position never occupied before[15]. They are in between the categories subject and object.

Professor Balkin's answer to Calo can be found in *The Path of Robotics Law*. Both scholars are American and deeply pragmatic. However, while Calo focuses on the essential qualities of robotics, Yale´s professor has a social insight on the subject: his concern is how society uses robots and integrates them into daily life, transforming the social relations and generating new experiences[16]. The author emphasizes the interaction between product and users in new technologies: the use of "generative" technologies[17] is shaped by users through their necessities and personal tastes.

When it comes to embodiment, Balkin's perspective differs radically from Calo's: he thinks it's not a robotics' salient quality, as he understands that earlier technological devices has also had embodiment and could cause physical harm through it. As an example, he mentions the case in which a laptop causes physical harm if it is thrown to a person. Balkin considers that the key is that physical harm can be caused *because* of the programming that regulates the behavior of new machines[18]. This creates deep concerns related to the assignment of responsibility for injuries, both from a civil and from a criminal perspective. But he does not think that the ability to cause physical harm is an essential quality of robots:

---

[13] *Id.* at p. 538

[14] *Id.* at p. 532

[15] Further on, we will see that we could place it in the same position as animals, even it's doubtful that they will be perceived in the exact same way. Although, animals are not as a close to the subject category as the robots.

[16] Balkin, *op. cit.* 2015 (Fn. 4), p. 45

[17] Concept used by Jonathan Zittrain and cited by Balkin, in Balkin, *op. cit.* 2015 (Fn. 4), p. 47

[18] *Id.* at p. 49

instead, he considers this is just an aspect that can (must) be analyzed by legal experts and lawyers.

Both authors come to an agreement regarding emergent behavior. This feature poses the question of who will be responsible for any harm caused by robots; self-learning systems, says Balkin, can be unpredictable[19]. This strikes as a serious problem when it comes to the assignment of liability. The breaking point that emergent behavior produces in respect of programming is notorious; and it will be also discussed in the following section.

Balkin is quite clear about the social valence assigned to robots: he says that what Calo is describing might be either anthropomorphism or zoomorphism, and that the projection of human or animal emotions onto inanimate objects is as old as history itself[20]. Yale´s professor thinks that what Calo attempts to describe is something deeper, that he calls the "substitution effect"[21]. According to this, people treat robots and AI systems as a human or animal, but always for a specific purpose. When they are not being used for that purpose, they are considered again as objects. (This distinction brings deep discrepancies when it comes to the way each author treats physical harm caused by robots.)

3- Problems related to liability for harmful results

The issues that appear when it comes to liability for harmful results are multiple and can be combined. Nevertheless, they will be analyzed as schematically as possible, treating them firstly within the scope of Civil Law and secondly within Criminal Law. Finally, problems in the field of Ethics will be posed. Civil law does not present big challenges, neither does liability for intentional crimes. In the negligence field, on the contrary, things might get more difficult. Ethics, the trickiest issue in robotics, will be treated in a less comprehensive manner as it exceeds the scope of this work.

Taking the Civil Law into perspective, I should agree with Hilgendorf, who refers to the strict liability provision contained in the German Road Traffic Act[22]. In Argentina, the 1757 article of the Civil and Commercial Code contains a similar rule[23], which repeats the content of the previous 1113 article of the Civil Code. Traditionally, that was the rule that informed the

[19] Unpredictable not only for programmers but also for users, *Ibid.*, p. 52
[20] *Ibid., p. 56*
[21] *Ibid*., p. 57
[22] Hilgendorf, *op. cit.* 2017 (Fn. 11), p. 180
[23] It establishes that the damage resulting from risks or defects of the thing or from the risk or danger of the activity, as the liability is of an objective nature, shall make the owner or guardian liable

liability for the drivers that cause a harm even in a full respect of the driving provisions. This solves further problems that could be presented by robotics in users' liability.

Let us focus on criminal liability for intentional conducts. Imagine the situation where a person uses an autonomous robot to break-and-enter another's house[24]. As soon as we realize that the human person is using the robot as an instrument, our opinion about the intentional meaning of the conduct does not change. The burglary example presents yet a little problem: the fact that, in order to commit burglary, law claims for the human person to be the one that breaks and enters into the building. Even though the figure of the person behind does not fit perfectly, the main principle that founds it does: the control over the act. It should not matter that the natural persons do not execute by themselves the elements of the criminal provision: if the conduct of the machine is directed by them, they control the outcome. Perhaps it will be necessary, in order to preserve the fair warning standard, to add another clause in the General Part of the Criminal Code stating that when a crime is committed using a machine as an instrument, the natural person is to be held liable.

Liability in negligence presents more serious problems. The first one is about the attribution of the result: who would be held accountable? Suppliers[25], users, or both? The second one, deeper, is about the causes of the result[26]: it could be a system failure, inadequate programming[27], or misuse. Finally, all these causes can converge, or it could be even possible that the self-learning capacities of the device have resulted in an unlawful outcome. Balkin adds the problem of demonstration of responsibility[28], but I think that proof problems should not be mixed with the substantial analysis of liability.

At this point, I agree again with Hilgendorf, who talks about the special nature of negligence to discard the need for legislative action[29]: these requirements are decided according to the specific case. Whenever the level of duty of care is not accomplished, the court will establish which duty was violated and declare liability. This makes things easier, but not in the case of convergent risks. Perhaps it should be necessary to lighten the reliance

[24] We can even imagine a situation in which the robot isn't even entering the property, but is already inside and is remotely controlled, ex: hacking., See Calo, Ryan, "Robots in American Law", in Hilgendorf, Eric and Seidel, Uwe (eds), op. cit 2017 (Fn. 9), p.82

[25] We understand the term supplier in the widest sense: producers, programmers, developers, etc.

[26] See Balkin, op. cit. 2015 (Fn. 4), p. 52

[27] Let us think these two last cases as separate ones. "System failure" refers to a situation in which the machine was well programmed, but there was a system failure. When we say "inadequate programming" we are referring to cases in which the system is in perfect conditions.

[28] Balkin, *op. cit.* 2015 (Fn. 4), p. 53

[29] Hilgendorf, *op. cit.* 2017 (Fn. 11), p. 181

between the negligent conduct and the result, in order to punish the forbidden risks. Of course, that would imply at a certain point to submit the market and the technological development to a constant criminal threat, but it might be necessary. This could be done by the creation of endangerment crimes.

The real problem within negligence reveals when considering harms caused by an emergent behavior. Who will be held accountable for the harm or damages caused by a robot as a result of a behavior learned by its own experience in the real world? Answers must go from strict liability for the users[30] to the personal liability of the robots[31]. This last option should be discarded given the fact that law regulates only human behavior. Nevertheless, it was proposed in the context of the European Parliament, and it will be analyzed further on.

The concept of emergent behavior, as defined by Calo and Balkin, adds confusion to the topic. The first thing that we should analyze is foreseeability. For users, this can lead to liability in negligence for a fault at the duty of care over a risky object. In the case of producers and programmers, things get more difficult and could be seen from the perspective of the quality standards and the sate-of-the-art at the moment[32].

But what happens when it comes to unforeseeable results? Here we need to bring on Balkin's reflections related to the substitution effect: humans create robots with the objective of putting them in charge of tasks that historically have been made by humans. This necessarily means that, when a robot assumes a task, humans are delegating it. This should impact in the users liability, liberating them from it. In autonomous cars, some authors think about the need for a fallback solution, by giving the control back to the user in critical situations. I agree with Jochen Feldle, who judges this solution unsatisfying[33]. He even quotes experiments that suggest that human drivers need up to 40 seconds to gain awareness of the road's situation.

---

[30] That seems to suggest Ryan Calo in *Calo, op. cit.* 2015 (Fn. 3), p. 554

[31] That is the consequence of positions that consider robots as electronic persons, exposed in the point 59 f of the Motion for a European Parliament Resolution of Robotics, available at http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2017-0005+0+DOC+XML+V0//EN#top. This position is also commented on Takayuki Matsuo, "The Current Status of Japanese Robotics Law: Focusing on Automated Vehicles", in Hilgendorf, Eric and Seidel, Uwe (eds), *op. cit.* 2017 (Fn. 9), p. 170

[32] That seems to suggest Uwe Seidel in Seidel, Uwe "Industry 4.0 and Law- Experiences from AUTONOMICS", in Hilgendorf, Eric and Seidel, Uwe (eds), *op. cit.* 2017 (Fn. 9), p. 21

[33] Feldle, Jochen, "Delicate Decisions: Legally Compliant Emergency Algorithms for Autonomous Cars, in Hilgendorf, Eric and Seidel, Uwe (eds), *op. cit.* 2017 (Fn. 9), p. 196

This poses scenarios of harmful results without attribution, with the respective keen feeling of injustice generated. However, I think that it does not represent a real problem: the need for attribution of harmful results is a relic of the ancient objectivism. If we are facing a result that is contained in a criminal provision and was neither caused by a negligent conduct of the producer, the user, nor the programmer, we will not be able to punish any person. And, from the basis of the pragmatism revealed by the American authors, society will have to assume the risks establishing an acceptable level for them, related to the development degree that is expected to be achieved[34]. We should also have in mind the phrase that is repeated in industry as a mantra: autonomous vehicles could reduce traffic accidents in a 90%, as they eliminate the human error factor. So, it would not be desirable but mandatory to create autonomous vehicles. In the future, it will not be crazy to consider manual driving as a forbidden risk.

Basically, I propose to think in emergent behavior and the substitution effect as legal limitations on the grounds of the proximate cause theory[35], in order to liberate the programmer and the user, respectively. Thus, we would not fall into the temptations of strict liability for natural persons or subjective liability for robots. On the first case, I agree with Balkin, who explains that liability without at least negligence can endanger the industrial development[36], besides that it is not an adequate solution within criminal law. The second possibility, as previously said, is not satisfying: law cannot regulate machines conducts but the human behavior related to them. Yet again, Balkin is right: technology is important because of the new practices and social relations that creates.

The tendency to consider robots as liable entities is an effort to avoid the analogy to animals and recognize them in their intelligence. It is a dangerous path, by which we could witness sci-fi scenarios. Let's use the *Copy or Vote Paradox* as an example, a hypothetical case created by James Boyle and modified by Calo in his article[37]: an artificial intelligence announces it has achieved self-awareness and, after reading *Skinner v. Oklahoma*, claims the right to make copies of itself. Eventually, these copies demand a pathway to suffrage in order to get representation in Congress.

---

[34] This seems to be Balkin's thought in Balkin, *op. cit.* 2015 (Fn. 4), p. 59

[35] I refer to the *Theorie der objektiven Zurechnung*

[36] Balkin, *op. cit.* 2017 (Fn. 4), p. 52

[37] Calo, *op. cit.* 2017 (Fn. 4), p. 529

Of course, the example is not reasonable, but reflects the spirit of the incomprehensible motion to consider robots as electronic persons, taken from the European Parliament[38]. It must not be understood as a consequence of the new ontological category for robots that Calo claims. The existence of an ontological category does not necessarily lead to consider robots as persons. The particularities in robotics do not impose the attribution of rights and obligations, but to treat them differently. In the document *European Civil Law Rules in Robotics*[39] was considered "useless and inappropriate". They ask themselves for the inconvenience of establishing that autonomous robots have rights and obligations: the field of obligations (duties) is understandable, but not the rights sphere; would it be the right to life? The right to receive remuneration? It would simply be absurd.

Yet again, these positions try to avoid the non-punishment solution for harmful events. It is obvious that in many cases it would be possible to establish that, besides the emergent behavior of the robot, the result was foreseeable and a human person can be held liable[40]. In other cases, we could think of liability determined by the guarantor role in the supervision of a risky element, generally used to assign responsibility to the owners of dangerous animals. In other cases, it will not be so easy. Unfortunately, the only way to build duty of care rules related to robots is taking note of the harmful results as they occur. Think of one basic criteria to assign liability for a harm within the risk: the rules of experience. If there is not any experience, there are not rules. This would imply, again, to accept harmful events by which anyone can be held liable for. Balkin's insight would simply talk about the equation between benefits and risks.

4- Ethical issues

The relativity of how autonomous systems should be programmed goes even deeper when taking into consideration the cases in which there will necessarily occur a harmful event, therefore having to choose which one will take place.

There is a debate between the American utilitarianism and the German Human Dignity Principle: for the Americans, it is possible to sacrifice one life in order to save many[41]. For the

[38] UE, *op. cit.* 2015 (Fn. 31) point 59 f

[39] *op. cit.* 2016 (Fn. 7), p. 12

[40] This is the same situation that was previously mentioned. In these cases, the analogy with animals is valid. In a certain way, animals also have an emergent behavior, and as long as it's predictable, it generates criminal liability to the owners.

[41] We should check if the utilitarian reflection is also valid for scenarios where we have to sacrifice an old man to save a young one, a sick person to save a healthy one, a person who has many children instead of a single one, etc.

Germans, on the other side, influenced by Kant, human life is the most important legal interest and its sacrifice is not allowed to save more lives: one is just as equal as many. If we take the American perspective into consideration, we would say that the car should swerve to the least harmful result. Should we take the German perspective instead, would inform that the car will not be allowed to swerve and risk people that was not endangered before[42]. In short, the American philosophy enables to consider it under the necessity defence, while the German principles consider that act unlawful. As an unlawful act, programming cannot be done in such way.

Even if we took the American perspective as valid, this would not solve the situations in which we have to decide whether to choose between two harmful results in the same quality and quantity of people. Let us suppose equal variables in age[43], people in charge of, economic incomes[44], etc. In these cases, perhaps we should apply the German principle that does not permit to risk people that were not endangered before.

Nevertheless, it does not seem to be so easy to solve the conflictive situations with an objective parameter. In other words, the number of victims can be used as a value, but we have to take into consideration other aspects such as the already mentioned people in charge of, age, profession, etc. It gets even thornier when we put into the variables the possibility of harming the passengers.

There are some quite interesting articles related to this, where authors ask themselves for the context in which it would be allowed- or not- to endanger the passengers[45]. We have to take into account that, in some cases, it would be more rightful to risk passengers rather than passerby, based on the fact that they were those who introduced the risky element into society. The problem poses the next perspective: which enterprise would advertise a car saying that, in the case of various risks, it could sacrifice its passengers?

This is precisely the reason why the MIT has launched the so-called *Moral Machine* experiment[46]. It consists in a website that shows to the visitor different scenarios where a

---

[42] This equals to programming the car not to swerve in a situation like that. It will have to kill the person that was in the car's original direction. See Feldle, Jochen, *op. cit.* (Fn. 33), in Hilgendorf, Eric and Seidel, Uwe (eds.), *op. cit.* 2017 (Fn. 9), pp. 202-203.

[43] Age is an ambiguous value. While some people might consider that an old person has lived for a longer period of time and therefore must be sacrificed to save a young person, others may think that the oldest life is worth more for its acquired experience.

[44] Apart from the fact that it attempts against Human Dignity

[45] Marchiori, Samuela (2016) (When) should self-driving cars be allowed to endanger their passengers? 10.13140/RG.2.2.15997.77281.

[46] See http://moralmachine.mit.edu/

harmful event will inevitably happen. This visitor should choose between the variables mentioned before so they can create a result for each scenario. In this way, they would charge their decisions by creating, as an outcome, a "social moral". It all seems to point that there have been attempts to achieve a relative consensus over the ethics rules that should be applied in each case, as proposed by Iyad Rahwan, MIT professor in a video[47]. At the end, he analyses Asimov's *zeroth law of robotics*[48], suggesting that the need for preserving the humanity *as a whole* requires to find that consensus.

5-Conclusion

Legal conflicts posed by robotics are not related to embodiment, but to emergent behavior. This feature is the one that allows unforeseeable robot behaviors for producers, programmers and users. Also, a substitution effect takes place: robots are taking over tasks that once were done by humans, therefore producing a task delegation. Both circumstances must block the attribution of the results by the grounds of the proximate cause theory.

The need for blaming someone for any undesired outcome is a flaw from ancient objectivism. For this reason, robotics must be analyzed by the personal doctrine of the unlawful act. Every actor should answer to what he could avoid. There will be harmful results that cannot be attributed to anybody, and such is the price that has to be paid for the increase of life quality brought by robots. This has happened with every mass-produced inventions, such as cars themselves.

Finally, it is true that in a country such as Argentina, where there is no development of robotics industry, the analysis of these matters could seem useless. Notwithstanding, even as consumers, results eventually may be produced and we should be prepared to face them. On the other hand, the possibility of crime committing through autonomous systems also requires legislative efforts in order to avoid impunity situations.

[47] See https://www.youtube.com/watch?v=nhCh1pBsS80
[48] "*A robot may not harm humanity, or, by inaction, allow humanity to come to harm*"