Informatica e diritto, XXXIX annata, Vol. XXII, 2013, n. 1, pp. 39-45

Robotic Societies and Law: A Plea for a Robotic and Simulation Science of Legal Phenomena

Domenico Parisi*

Today, computers are everywhere and they are increasingly invading the life of human beings and the functioning of their societies. This has both positive and negative aspects but one certainly positive aspect is that computers can be scientific tools that let us better understand all aspects of reality, including human beings and human societies. Computers can collect enormous quantities of data and find all sorts of regularities in the data but it is not clear that these "big data", although useful, do not really explain reality. A more interesting use of computers as scientific tools is that they can be used to construct artifacts that behave like human beings and live in artificial societies which are like human societies. The artifacts are theories of human behaviour and human societies. Scientific theories of human behaviour and human societies are traditionally formulated by using words but words have unclear and ambiguous meaning, and their meaning remains unclear and ambiguous even when they are defined or re-defined by using other words. Theories as computer-based artifacts are entirely clear and unambiguous because, otherwise, the artifact cannot be constructed and one can always "open" the artifact and see how it is structured and functions. The behaviour of the artifacts are the empirical predictions which are derived from the theory which has been used to construct the artifacts. If the artifacts behave like human beings and reproduce human societies, the theory is confirmed and the artifacts capture what underlies human behaviour and human societies and explains them.

To exploit the advantages of this approach, one should follow the principle "one artifact (one theory)/many phenomena". The same artifact should be able to reproduce (explain) as many different human phenomena as possible. In fact, computers make it possible to develop a non-disciplinary science of human beings and human societies. Reality is a very large ensemble of different phenomena but all these phenomena are connected together and, often, to explain the phenomena which are studied by one discipline, it is necessary to look at the phenomena studied by other disciplines. Hence, to

* The Author is former director of the Institute of Cognitive Sciences and Technologies, National Research Council of Italy (ISTC-CNR), Rome and associate researcher at the same Institute.

really understand human beings and their societies, the same artifact should be able to reproduce phenomena which are studied by the biological sciences, in particular evolutionary biology and the neurosciences, psychology, anthropology, sociology, economics, political science, legal science, and the historical sciences. To explain human behaviour one cannot ignore its biological and, ultimately, physical bases, and therefore the artifacts that reproduce human behaviour and human societies must be robots and collections of robots. Today, most robots are constructed with practical applications in mind. These are robots as technology. What we are talking about here are robots as scientific theories, and the two different goals for constructing robots should be kept separate because they dictate different research agendas.

One significant class of human behaviours and human social institutions are those which are studied by the legal sciences. The system of laws and regulations existing in a society is an important component of the society and we cannot claim to have understood a society unless we have understood and explained its system of laws and regulations. The legal sciences, perhaps more than other scientific disciplines who study human beings and human societies, are both an attempt at knowing and understanding reality and an attempt at identifying how human societies should be regulated in order to function properly and what should be done with respect to behaviours which violate these regulations and therefore damage others - either single individuals or the entire society. Science can be useful to human beings both for understanding reality and for solving their problems but the two goals should not be confused together because science can only know and understand reality if it looks at reality with complete detachment and ignores values and ideologies. One advantage of a robotic science of human beings is that it forces scientists to ignore their values and ideologies because, once you have constructed a robot, the theory incorporated in the robot will generate all the predictions that can be derived from the theory, both those that you like and those that you do not like. But robots as scientific theories should also help human beings to deal with the very difficult problems they face today and will increasingly face in the future and to design what can be called "non-utopian utopias": artificial worlds which can be analyzed and manipulated like the real world and which, although they are not real, can be made real.

A robotic science of legal phenomena still does not exist, although there are attempts in the field of artificial intelligence and agent-based social simulations, and even in the collection and analysis of "big data", that produce

interesting results and offer useful suggestions. What we will do here is outline some of the phenomena that this science should be able to reproduce and some of the questions that it should be able to answer.

Robot as theories of human behaviour must have a body and a brain, they must have genes which evolve in a succession of generations and that they inherit at birth, they must live in an environment which includes other robots, they must learn in this environment, and they must mostly learn by imitating or by being told by other robots. Furthermore, there must be all sorts of interactions among the robots and all sorts of social organizations and institutions which constrain their behaviour and shape their society. This clearly is a difficult objective to reach but our principle "one artifact/many phenomena" requires that we progressively realize this objective because only if we construct robots that have all these characteristics and exhibit all these behaviours, we can reasonably claim that we have reproduced and, therefore, understood and explained human behaviour and human societies. Which phenomena should these robots reproduce that are studied by students of laws and regulations? We will answer this question by posing a number of research questions.

Human beings sometimes behave selfishly, by doing their personal good while ignoring the good of others, and sometimes they behave altruistically, by doing the good of others even at the expense of their own good. Is selfishness genetically inherited and altruism culturally learned, or the other way round? Are they both genetically inherited and culturally learned? To the extent that selfishness and altruism are culturally learned, what kind of culture and what kind of society will favour selfishness or altruism? Can we construct robots that help us to answer these questions?

Sometimes human beings behave in ways that damage others and, if a society does not find how to contain these other-damaging behaviours, it may collapse because for its members the advantages of living together may be overridden by the disadvantages. Can robots help us explain other-damaging behaviours? Can robots help us distinguish among different classes of other-damaging behaviours? Can robots help us to identify how to contain these different classes of other-damaging behaviours?

Punishment is one way in which other-damaging behaviours can be contained. Punishment is a very important aspect of the behaviour of all animals. If one behaviour is followed by some stimulus which decreases the probability that the behaviour will be executed in the future, the stimulus is a punishing stimulus. (Rewards are stimuli that increase the probability that the behaviour which has been followed by them will be executed in

the future). It is evolution which gives punishment or reward value to stimuli and therefore the system of punishment and reward is inscribed in the genes. But, especially in human beings, whose behaviour is mainly learned during life, initially neutral stimuli can become punishing or rewarding because they are associated with stimuli which already are punishing or rewarding. Robots that learn based on punishments and rewards already exist but legal phenomena pose many new problems. Is being punished because of one's other-damaging behaviours different from being punished for behaviours which only damages oneself? Is being punished by "nature" (like touching a very hot object with one's finger) different from being punished by another individual? Some social animals are punished by other individuals. Is being punished by another individual (a conspecific) the same for nonhuman animals and for human beings? Is being punished by another individual in the presence of other individuals the same or different from being punished by another individual in the absence of other individuals - or without that other individuals can be informed of the punishment?

Punishment can have different effects as a function of the nature of the other-damaging behaviour, for example, as a function of whether the other-damaging behaviour was "willed" or "not willed". Can we construct not only robots which, like current robots, do X but also robots that want to do X? How should the two types of robots behave so that we are justified in saying that they are different? How should the brains (artificial neural networks) that control the behaviour of the two types of robots be different? At which age human beings become able to want to do X? Are there pathologies that cause human beings to become unable to want to do X so that they can only do X? What is the role of the ability to predict the consequences of one's actions in wanting to do X - or in being "conscious" of what one is doing or will do? What is the role of being able to linguistically describe to oneself these consequences? What is the role of the ability to evaluate the predicted consequences of one's actions before executing the actions? Does this evaluation takes into consideration only one's advantages and disadvantages or also the advantages and disadvantages (damage) for others? (Robots that predict the consequences of their actions and evaluate these consequences and robots that talk to themselves have already been constructed).

Are there nonhuman animals that want to do what they do? Is there a continuum, which can be illustrated by examining different animals, or a neat separation between simply doing X (for example, in worms) and doing X because one wants to do X (for example, in some nonhuman primates)?

If doing X is followed by punishment (especially punishment from others), what are the consequences for robots that only do X and for robots that do not only do X but also want to do X? Should we construct other types of robots, for example, robots that have the intention to do X and robots that are thinking of doing X?

As we have said, most behaviours of human beings are learned from others and human beings tend to imitate the behaviour of the other individuals with whom they interact. Therefore, we should construct communities of robots that include sub-communities made up of robots that interact among themselves more than with robots belonging to other sub-communities. This may lead to the existence to sub-communities of robots who tend to behave without damaging others and other sub-communities made up of robots which tend to damage others. Does the existence of sub-communities of robots that tend to damage others reduce the efficacy of punishment? (Some robots that try to answer this question have been constructed).

Punishment may not be the only mechanism for containing other-damaging behaviours. Other-damaging behaviours can be exhibited because they are the only way to obtain what is needed for one's survival. A community of robots can organize itself so that all members of the community have what is needed for their survival and this may reduce the incidence of other-damaging behaviour. And the community of robots must compare the costs of other-damaging behaviour and of discovering and punishing other-damaging behaviours with the costs of insuring that every member of the community can survive without damaging others. (Robots that answer this question also have been constructed). Other interesting questions are the following. Can we construct robots that can be said to have "rules of behaviour"? Can only robots that want to do X because they predict the consequences of their planned but not executed actions and evaluate these consequences, have rules of behaviour? Is language and talking to oneself necessary? Is being punished not by a single other robot but by the entire community of robots, or by a representative of the entire community, necessary? How rules of behaviour emerge and how are they established? Does violating rules of behaviour automatically damage others because it makes behaviour less predictable and being able to predict the behaviour of others is crucial for living together? Why are some rules of behaviours written? What difference does it makes to have written rules of behaviour? In what types of robotic societies are rules of behaviour written?

As we have said, a robotic science of human beings and human societies should construct robots that progressively reproduce all the different phenomena which are studied by different disciplines. Human beings live in different societies and different cultures and their societies and cultures have different histories. Therefore, a robotic legal science should construct robotic societies with different cultures and compare how other-damaging behaviours are dealt with in these different cultures and how legal systems have changed during history. And it should find how legal systems correlate with the economic, political, and religious institutions of different robotic societies. Furthermore, legal systems are not the only mechanism for containing other-damaging behaviours. Another mechanism is self-punishment which can take the form of religious beliefs or lay morality. A robot punishes itself if it does some other-damaging behaviour or even if it thinks of doing some other-damaging behaviour, and this capacity/tendency to self-punishment can be innate or acquired during life. To the extent that it is acquired during life, we should construct different robotic societies in which the capacity/tendency to punish oneself for other-damaging behaviours is more or less developed.

But, as we have also said, a robotic science of human beings should also help human beings to better understand the many difficult problems they face today and will face in the future and, perhaps, suggest how to deal with these problems. We should construct robotic societies which are like today's societies and we should reproduce how today's societies deal - or should deal - with other-damaging behaviours. Some of the phenomena which characterize today's societies and which are relevant for a socially useful legal science are the following.

Due to advances in the technologies for transporting people, goods, information, and money, human beings are increasingly living in a globalised world. This globalisation is economic and cultural but political sovereignty remains to the states. Since states are responsible for defining the laws and regulations which are to be applied to their citizens, this creates a conflict between globalisation and local laws and regulations which will have to be solved, especially because other-damaging behaviours themselves become globalised. A plurality of robotic states with different legal systems and a variety of economic and cultural relations can be of help in finding and testing possible solutions.

The list of other-damaging behaviours must be constantly up-dated as a function of the changes that occur in societies and ways to contain new types

of other-damaging behaviours must be found. Today's societies change very rapidly and, for example, the financial economy, marketing at all levels, and digital technologies have an increasing impact on human life and on the organization of human societies. Simulating financial economies, marketing, and the impact of digital technologies by constructing robotic societies may help legal systems to identify their potential other-damaging impact and how to reduce this impact.

Western culture leads to a decrease in the strength of self-punishment as a way to contain other-damaging behaviours. Religion plays a decreasing role in Western culture and economic and cultural reasons, the reduced role of the family and the fact that an increasing number of people live in cities, lead to forms of extreme individualism in which the consequences of one's behaviour for others are less taken into consideration, unless they are punished by the law. Robotic societies which reproduce Western societies and non-Western societies which are adopting Western culture should help us to examine this phenomenon and its consequences and, if these consequences are negative for the well-being of human beings, to find ways to correct it.