Robots

Technical Individuals and Systems

Paul Dumouchel
Ritsumeikan University

Japan is considered to be the most highly robotized society in the world, which at the industrial level at least is certainly true. There are presently some 250,000 industrial robots employed in Japan, which is more than in any other country, and this number is expected to double in less than five years, and quadruple in ten. Furthermore, according to a claim that has been around at least since the early 1960s, there is a 'love story' between Japan and robots (Koestler 1960). The Japanese are said to love robots, while Westerners tend to fear or at least to be wary of them, and many studies show that social acceptance of robots is higher in Japan than in Europe (Hornyak 2006). Japanese newspapers are always eager to report that someone invented a robot for making sushi or a robotic bed for elderly patients that transforms into a wheelchair. Robots in everyday life are seen as useful helpers rather than potentially dangerous rivals.

However, for those of us who live there, the presence of robots in everyday life is not so evident. It is not clear to me that there are more robots in my normal environment in Japan, than what I find when I go to France or Canada. The love story between Japan and robots may be true, but it seems to be taking place at the level of cultural representations more than that of daily experience. Part of the explanation for the apparent discrepancy between the number of robots in Japan and the daily experience of robots, may simply be the lack of social visibility of most robots. Unless you are a factory worker, encounters with industrial robots are rare events. Most robots remain hidden in industrial plants, where most of us rarely see them, or inside hospitals or other health care service facilities, where their use is reserved for specialized personnel and particular populations, i.e. elderly persons, sick children or patients with particular diseases. Robots may be numerous in Japan, but they usually fulfill specific functions in private locations and rarely directly interact or interfere with the general public.

What Is a Robot?

There are other reasons, I believe, why robots tend to remain invisible in our societies. The first is that the answer to the question 'What is a robot?' is far from clear. The second closely related reason is that in spite of the fact that the answer is not clear, or perhaps because it is not clear, we, the general public, tend to have pretty specific ideas about what is and what is not a robot. It follows that very often, because it does not fit our preconceived idea of what a robot is, we do not recognize a robot when we see one! For example, in French we call a food processor a 'kitchen robot' (un robot de cuisine). At first sight this is a bit surprising; a food processor does not really correspond to what most of us imagine when we think of a robot. What first comes to mind is usually something like Asimo, the Honda-built humanoid robot, or Aibo, the robotic dog from Sony. Both seem better candidates for what we expect a robot to be. Perhaps a Roomba vacuum cleaner or a robotic lawn mower would also do the trick. However, when you think of it, a food processor satisfies pretty well the original definition of the word 'robot': a worker, conceived as an automatic, self-powered device that replaces a human worker and that is to some extent autonomous. A food processor works for you in the kitchen. It cuts things up, or purees them, or whips up a mayonnaise.

The fact is that most robots do not have a humanoid shape. Drones and other types of unmanned land, marine, or airborne vehicles are robots. *Resyone*, the robotic healthcare device made by Panasonic that transforms from a bed into an electric reclining chair, is a robot. Industrial and medical robots come in forms and shapes determined by their particular use or function, and more often than not they look like just any other machine. That is to say there is nothing particular in their appearance that picks them out as a robot.

The question which this diversity raises is, are all automated devices robots? Is an automatic door opener a robot? Is the automatic pilot in a plane that can keep it flying for hours with little human intervention a robot? Is an escalator a robot? What about a moving sidewalk or an automatic vending machine? Is the sensor that turns the light on at night when someone approaches a robot or not? Is the automatic wicket that reads your magnetic card and lets you in at a subway station a robot? Is a driverless subway train a robot? Is an automated teller machine a robot? Is a food processor a robot? How about your dishwasher? Is the printer attached to your computer a robot? The list can easily continue, so where do we stop? And if we do, how do we draw the line between robots and other types of automated device? And if we don't stop, then it seems that robots are everywhere in the modern world, so present that we don't even see them anymore, and the word 'robot' would then refer to so many things, that it would not mean anything in particular.

The answer to the question, 'is this or that object a robot?' is far from clear, because among other things the concept of robot itself is not very clear. The word was invented—or rediscovered—for a play, as the generic name of fictional characters and not as a scientific category. In Karel Čapek's *R.U.R. (Rossum's universal robots)*, robots are artificial biological–rather than mechanical–creatures: they are androids made of synthetic biological material. They are indistinguishable from humans and work for us, as secretaries, gardeners, servants, factory workers or whatever we may need, until of course the day inevitably comes when they rebel and destroy the human race. Hence robots are artificial agents that work, that perform some function in our stead and that enjoy some autonomy in doing so. This last point, relative autonomy, is what distinguishes robots from ordinary machines or first generation automated tools, that is, automated tools that are entirely dependent on human workers to accomplish their task, like an electric saw or lawn mower. The word 'robot' and the idea remained, but the synthetic biological origin of robots was forgotten. Robots came to be understood as automated mechanical devices that replace human workers, or accomplish their function, something which by definition requires at least a minimum of autonomy. The difficulty with this definition and way of understanding what a robot is, comes from the fact that it mingles two very different types of criteria. First, an engineering mechanical criterion: an autonomous automated device; and second, a social functional criterion: that works in our place. If 'work' is a well-defined scientific concept, in physics for example, fulfilling a human task is not.

The fact is that there are many ways of doing the same thing, especially when doing the same thing is understood as obtaining the same or a similar result. Imagine someone who is hired to carve floral patterns on soaps in a factory. The owner can replace him or her, or more precisely dispense with this worker, simply by having molds made such that the soaps come out with the patterns already on. Another way of reaching a similar result is to buy and install an automated soap carver which fulfills the function of the human carver. The first solution may be understood as a way of automating the production line, in the sense that the desired result is automatically obtained through the production process without any particular agent's intervention. The second consists in replacing a human agent by an artificial agent, a 'robot soap carver'. The artificial agent must be able to do some of the things the human agent was doing; to recognize a bar of soap when there is one in its field of perception, to pick it up and carve it, probably to be able to determine the size of the soap and modulate its carving or choice of pattern in consequence, perhaps also to recognize the density of different soaps and adjust its activity to obtain better results, and so on. The machine must also be able to do this by itself (i.e. autonomously, without constant human supervision).

Why Robots?

Now robots, unlike human soap carvers, do not get tired (though they do break down), they do not complain, they do not get distracted and they do not go on strike. These are some of the reasons why we want robots and resort to them in various circumstances. Yet only some of the reasons; robots are also cheaper, and often more efficient or precise than human workers, they do not require retirement plans, and they do not have legal rights. That is to say, we want robots to have all the qualities (and more) that masters look for in slaves, or factory owners in workers, or commanders in soldiers (Arkin 2009); but none of what they see as their failings, weaknesses or insubordination. In other words, there are many dimensions of human autonomy which *we do not want* robots to have. Thus, we want robots *both to be and not to be autonomous*. It is this contradiction which is at the heart of Čapek's early fable, where robots are just like us, but different, until they reveal themselves *to be too much like us* and decide to wage war upon their masters! This contradictory project is also, I believe, what explains the permanence of the theme of the robots' rebellion. Roboticists, and others who write or reflect upon robots, for example Wendell Wallach and Colin Allen in *Moral Machines*, often repeat, and sometimes lament, that at the present time and for the foreseeable future, we are unable to create really autonomous machines, for example, artificial agents that are sufficiently autonomous to be morally responsible agents (Wallach & Allen 2009; see also Lin, Abney & Bekey 2012). This is certainly the case. However, the truth also is that *we do not want to create* truly autonomous artificial agents; we are afraid of them.

Many examples from popular Western culture bear witness to this refusal and fear. One of our greatest fears, since Čapek first introduced the term 'robot', and when the project of creating artificial agents became culturally plausible, is that if robots someday become 'really' or 'completely' autonomous, if they become 'like us', they will destroy us and take over the world. Short stories like Algis Budrys's 'First to serve' (1960: 73–86), films like *Blade Runner*, *The Matrix*, *Terminator*, and more recently *Transcendence*, essays like Bill Joy's 'Why the future doesn't need us' (Joy 2000), or the popular idea of 'the singularity' all explore, exploit and construct this apocalyptic fear as a central theme in contemporary Western culture: when robots become conscious and autonomous, it will be the end of the world as we know it.

Autonomous robots have long existed in Japanese manga and anime. Beginning with *Astro Boy* in the early 1950s, autonomous robots have generally been construed as good and helpful, even as heroes and saviors. In fact, Astro Boy is different from the other robots that are present in the story, because he has a soul. In consequence he is more human-like than other machines. This gives him greater insights into reality and makes him superior to other robots—though his incredible strength and power also help—but his human likeness also makes Astro Boy vulnerable to mistakes, to doubt and inner suffering. In spite of all his power, this robot, this artificial creature, is a lot like us. He shares many of our failings and weaknesses, and his being like us is what allows him to become a real hero and a successful ambassador for peace. There also are evil robots in *Astro Boy*, but their evilness usually comes from the bad intentions and goals of their creators. Robots can be evil, but they are not evil in themselves, and certainly not as a consequence of becoming autonomous. To the contrary, Astro Boy's greater autonomy makes him more reflexive, complex and sensitive to moral issues. A somewhat similar position is adopted by Akira Toriyama in his famous manga *Dragon Ball* where some robots are evil because of who created them, but robots are not evil in themselves.

The theme of the danger posed by autonomous robots because they are autonomous is less present in Japanese popular culture. It is however not entirely absent, for example at the beginning of the film *Patlabor 2,* the opening scene where the robotic suit of the captured rebel *labor* turns out to be empty suggests that *labors*, robotic external skeletons and armor can become autonomous, take action on their own even when there is no driver inside, and in consequence become dangerous.[1] Yet, in this case at least this theme is not followed up and the central argument of the film is quite different. The idea that the world will be taken over by machines, and that robots are the vanguard, or at least important agents of that invasion, though present in Japanese manga, films and anime, does not constitute a central theme in many or most of them. Popular science fiction

[1] See for example Bolton (2007).

stories that involve robots have generally been more attentive to the psychological and ethical issues related to the ability to wield formidable power that robots confer upon human beings, than to intrinsic danger of technology or of robots as such. This said, the point is not to oppose Japanese popular culture to Western popular culture (many counter-examples could be found on both sides), but to distinguish between two different attitudes towards robots and technology.

While Astro Boy is an autonomous robot, the robots in *PatLabor*, *Gundam* or *Neon Genesis Evangelion*, are vehicles and weapons, semi-autonomous machines whose common characteristic is that they are inhabited (driven) by humans—usually adolescents—who use them to fight evil and protect humanity. The human driver constitutes the brain and soul of the machine and the robot in return transforms the personality of its driver, makes him or her into someone different, occasioning in consequence severe psychological conflicts for the driver. As noted by many critics, problems of identity among adolescents often constitute one of the major themes of these stories (Napier 2007). Sometimes, as in *Evangelion,* the machine itself, the Eva01, has a soul of its own. In this case, the machine's soul is that of the mother of the driver, Shinji, and for him the experience of driving the robot and combat is also that of becoming one with himself and with the machine's soul. The important point is that in all these cases either the robot itself (as in *Astro Boy*) or the coupling of the driver and the machine (as in *Gundam* or *Evangelion*) constitutes an individual and the story describes the progressive individuation of the agent; him or her gaining, or failing to gain, a strong identity. The experience—either of the autonomous system itself, or that of living in a close, nearly symbiotic relationship with a technical object—constitutes a learning and growing process. In many ways these stories resemble apprenticeship novels (*Bildungsroman*).

To the contrary, in films like *The Matrix* or *Terminator* humans are originally the victims of their excessive confidence in and dependence on technology. When the film begins they are already suffering the disastrous consequences of their lack of foresight, and the gist of the story is their efforts at trying to escape if possible the terrible disasters their shortsightedness brought upon them. Instead of learning from their encounter with technology and becoming better through it, humans have as a consequence of their faith in and dependence on technology lost sight of what is important. Now that they have been punished for it, they must try to recover what has been lost. Rather than an apprenticeship novel, what we are given here is a cautionary tale. Beyond that difference in the literary genre there is also a fundamental difference in the way technology and robots are presented. In these films, robots do not constitute individuals. The enemy against which the hero fights is no so much the individual robots he encounters as the ‘system’, the matrix or Sky net. The evil that befalls us is carried out by individual agents, but it springs from the ‘system’, from technology as a whole which has become autonomous and has taken over the world. ‘Agent Smith’ and the ‘terminator’ are only foot soldiers of a much more powerful entity, omnipresent and quasi-omnipotent, that directs their smallest moves. These machines may be autonomous in a sense, but not in another; they do not evolve towards autonomy or gain an identity,[2] but remain remote controlled devices, even if like agent Smith, defective ones.

This failure of autonomy and absence of individuality in many popular representations of robots and technology is also manifest in another way, through the incident that constitutes the beginning argument of the film. On some fateful day Sky Net became conscious and the same more or less happened for the Matrix. How did it happen? We don’t know, but by what we are told, it just took place by itself. At some point, the system reached a threshold of complexity and interconnections, and that was it! Nobody did it! In contrast, Astro Boy is created by a solitary scientist who has lost his son in a traffic accident. In *Gundam* the first mechanical suit is built in a secret underground laboratory by the father of the main protagonist, who dies in the attack that opens the story and gives his son the plans for the machine. In *Evangelion* the Evas are created by the father and mother of Shinji, the main protagonist, and Eva01 which he operates is inhabited by his dead mother’s soul. Apart from the importance of the parent/child relationship, in all these stories the fact is that the machine—the robot, combat suit or the bio-mechanical entity ‘Eva’—is the defining technological breakthrough which makes the story possible. Morever, it is always created by someone, an individual with a name, who often also plays an important role in the plot beyond that of the inventor of the machine.

Technological systems like the Matrix or Sky Net may be ‘autonomous’ in a sense, but they are anonymous and they are not individuals. Rather, they constitute the environment within which individuals act, but an environment whose main characteristic is apparently to eradicate all traces of individuality, either by replacing humans with machines that the system directly controls, or by enslaving humans to its own purposes. On the contrary, in the Japanese manga and anime mentioned above, technology tends to constitute the triumph of individuality in at least three complementary but related senses. First, the central technology is the technical triumph of an individual, a major technical success

[2] This is particularly true of Agent Smith, who when he escapes the control of the Matrix appears in endless copies of himself like a computer virus. Only the recycled terminator from *Terminator1* gives in *Terminator 2* some signs of developing an identity, but precisely because he is now free from the control of Skynet.

achieved by someone. Second, it constitutes the occasion for the main protagonist—and often for others also—to triumph over him or herself, to face his or her fears, to resolve inner conflicts and to become a truly autonomous individual. Third and finally, it allows him or her to triumph in combat, to become a hero.

These two popular images and evaluations of robots and technology could hardly be more different: robots are evil as opposed to good or morally neutral; dangerous products of failure worthy of moral censure as opposed to technologies that occasion moral growth and provide opportunities to become better people; alienation and impersonality, as opposed to the triumph of individuality.

There is another fundamental difference between these two popular images of robots and technology. The first image of robots in Japan, found in many popular Japanese manga and anime, provides answers to such questions as: why do we want robots? What is good in technology? Overall that answer is: 'Robots and technology can make us better humans'. On the other hand, the image that is conveyed by many popular Western films and stories, which repeat and expand on Čapek's original fear,[3] does not answer those questions. In fact, the second image I listed, from the Western perspective, gives us many reasons for why we should not want to have robots, and why we should be wary of technological growth. Simultaneously, it paradoxically suggests that these questions do not really arise because this transformation will happen by itself. The bleak future which these films describe is not necessarily presented as inevitable, but technological growth or the 'rise of the machines' is viewed as an impersonal, autonomous (in the sense of automatic) process for which no one is responsible, and which no one can control or radically change.

Individual Artificial Agents and Systemic Agents

These different cultural representations of robots and of technology do not simply correspond to different ways of understanding technology and our relation with artificial agents. They also illustrate—in a dramatic and obscure way—differences in researchers' goals and objectives, and further reflect real differences in existing technologies. They reflect different *social* technologies, in the sense that these different types of technologies correspond to different strategies relative to the role of modern technology in social control.

[3] Probably not quite Čapek's 'original fear'. The play was written in the 1920s and it seems pretty clear that in *R.U.R.* robots and their rebellion are metaphors for the working class. However, this meaning was rapidly forgotten and the play was understood as a precautionary tales concerning modern technology and the 'rise of the machines'.

One possible definition of an autonomous robot is that it is a three-dimensional physical system—something which distinguishes and separates robots from virtual agents on computer screens—that can sense and respond to its environment and whose responses are (to some extent) under its own control, in the sense that the agent can learn and is adaptable, hence the system is not and cannot be entirely pre-programmed. Depending on the technical characteristics of the system and/or on the social and technical environment in which it is active, the robot's margin of autonomy can be larger or smaller.

Given this understanding of an autonomous robot, is the autopilot in a modern plane an autonomous robot? It certainly is a three dimensional physical system; it can also sense its environment and respond to it, adjusting the altitude, trajectory, tilt and yaw of the plane in response to changes in the environment. Its responses are not entirely predetermined: there is no complete map of the states of world where every situation the plane may encounter is identified and the proper response determined. So it seems that under this definition, a plane's autopilot must be an autonomous robot.

Perhaps, however consider the following difficulty. There is an object in the plane which is such that if you take it out, the plane does not have an autopilot anymore. Is this object an autopilot? In a sense 'yes', yet in itself this is just a computer that regulates its outputs in function of its inputs. By itself it cannot do anything, certainly not fly a plane. In order to do that, this device has to be properly attached, integrated into the complex physical, chemical, electric, electronic and hydraulic system that constitutes a modern plane. In order to sense the world and react to changes in the environment it has to be connected to, or coupled with, or better it has to become part of the whole system which is the plane itself. That is why, I submit, an autopilot cannot be considered to be either an autonomous robot or even simply a robot, because it can only act and sense its environment to the extent that it is inseparably united to the plane, with which it forms a unique system. Is the plane itself a robot, autonomous or otherwise? Answering that question sends us back to the perplexities mentioned earlier. What is a robot? The category of robot is too ill-defined, as argued above, to allow a clear answer to the question: 'Is this or that object a robot?' The autopilot is nonetheless an embodied intelligent artificial agent, but it is embodied in a particular way.

Using that same criteria to define an autonomous robot, is a drone that takes off, flies and lands by itself, that accomplishes its mission independently and then returns to base, an autonomous robot? It has on board a complex sophisticated autopilot, which is one of the systems that allow it to be autonomous. Is it a robot? If so, for what reason should we consider it an autonomous robot when the autopilot taken by itself is not? Essentially, because the drone satisfies the second socio-functional criterion of a robot: the drone does what a plane with a human pilot, and perhaps some other member of an on-board crew, would do. That second criterion however, as argued earlier, is not well-defined, and the claim that such a drone is an autonomous robot reflects its social role and the way it fulfills it, rather than intrinsic characteristics of the machine that it is.

Yet there is a central difference between the drone and the plane's autopilot. The drone is an individual, while an autopilot—either the one which the drone has on board or one that we find in modern plane—is not. It is just 'part of', an element in a larger system. My suspicion is that, faithful to Čapek's original use of the term, we spontaneously tend to consider as robots, artificial agents which are individuals, but that our intuitions leave us relatively in the dark whenever we are faced with artificial agents that are differently embodied. What is it then that makes an autonomous drone an individual?

Many years ago Francisco Varela argued that one of the defining characteristic of an autopoietic system is that it has a border in physical space that topologically delimits where the processes that define the system take place (Varela 1989). An autonomous drone is not, by far, an autopoietic system, but *its border in physical space* plays a somewhat related role, something which is not the case for an autopilot. More precisely, what the autopilot lacks is not a border in physical space; it is after all a physical object which you can move, install or remove. Rather, it is that where it acts, how it acts, what it senses, what it can and cannot do, is completely independent of the limits in space that determine it as the physical object that it is. To put it otherwise, what determines it as an object in physical space has no relation whatsoever to what it can do,[4] and, as mentioned earlier, by itself it cannot do very much![5] To the contrary, what the drone can and cannot do is inseparable from what determines it as the physical object that it is. Of course, what it can do also depends on its characteristics as an artificial intelligent agent; however, the drone's characteristics as an intelligent agent are not independent of its physical characteristics, while the characteristics of the autopilot as an intelligent agent are independent of its characteristic as a physical object, though they are not independent of the physical characteristics of the plane. There is more to this however: an autonomous drone is only autonomous to the extent that ground control allows it to be. At any point in time controllers can take control of the drone and direct it as they want. When that happens the drone

[4] This does not mean that its physical characteristics have no relation whatsoever to what it *could* do, *were* it to be installed as part of a larger system.

[5] Which also means, a point to which we will return, that considered as a physical object it is not an agent.

remains the same physical object that it is, though in the proper sense it becomes a different physical system. Does the drone cease to be an autonomous robot as a result? Clearly its margin of autonomy has been reduced or constrained. However, as long as it still exists, it is autonomous and further it seems that the drone continues to satisfy the second criteria. More to the point however, does the drone continue to be an individual agent? In this case, the answer seems to be a straightforward 'yes'. No matter the extent to which its margin of autonomy has been reduced, the drone, just as a soldier, private or officer, remains an individual agent even though it is embedded in a hierarchical command structure.

What we have here are two different types of technical objects, two different types of artificial intelligent agents. The distinction between these two types of artificial agents is more important to understand the social impact of these technical objects than knowing whether they are robots, autonomous or not. One type of artificial agent is embodied as an individual and will generally be spontaneously considered as a robot. What it can do in the world is closely related to what it is as a physical object. The other is embodied as part of a larger system, what it can do in the world depends on the characteristics of that system. From here on I will refer to these two types of artificial agents as individual artificial agents and systemic artificial agents.[6]

Autonomous individual artificial agents and systemic intelligent agents correspond to two different directions in robotics research. They also correspond to different research interests and motivations. These two different directions of research are highly interrelated since building an autonomous individual agent requires embedding in it numerous systemic intelligent agents that react autonomously to changes in the environment. They constitute preconditions for the robotic agent to be able to do the many things it does, i.e. walking, following its human partner's gaze, picking up a ball, extending its arm, etc. Whatever this robot does however, whether it is painting a car fender or being a receptionist, also depends on its physical characteristics and not only on the intelligent agents it contains. Because of this, these artificial agents are identifiable physical objects which can be individualized by human observers. Systemic intelligent agents, on the other hand, are invisible, not only because as physical objects, as elements of a larger system, they are usually hidden inside the machine, beneath its visible surface, but also because they essentially are what may be called *analytic agents*. They are analytic agents in the sense that *they do not do what they do*. That is to

[6] Of course individuals can be embedded as parts of larger systems. The characteristics of these higher-level systems will depend, among other things, on whether or not, and the extent to which, their parts retain their individuality.

say, whatever action they are responsible for takes place at a higher hierarchical level in the system in which they are embedded.

Japanese anime and manga like *Astro Boy*, *Gundam*, *Patlabor*, or *Neon Genesis Evangelion* focus on autonomous or semi-autonomous robots, on individual artificial agents, as do some films in the United States, for examples, those in the *Iron Man* series. These machines are tremendously powerful and dangerous physical objects that can be either good or evil, or that can be put either to good or bad use, but with which, in a sense, we can deal just as we deal with other individuals that populate our common world. A central characteristic of such agents or robots is that they are visible, identifiable agents. We can transparently attribute to them their actions in the world. Actions for which they may be or may not be morally responsible, but which we can in any case clearly attribute to them. Whether it is autonomous or controlled from somewhere in the United States, it is the drone which one can hear and see in Pakistan that shoots the Hellfire missile and in that sense it is the agent.

Popular films, like *The Matrix*, *Terminator* or more recently *Transcendence*, to the contrary, focus on invisible mystical logical entities, the mythical personification of systemic artificial intelligent agents that, interestingly enough, it seems can only be evil. A central characteristic of systemic intelligent artificial agents is that they are invisible, or if you prefer, they are essentially analytic agents, and in consequence they give rise to actions that cannot be attributed to any agent in particular, that cannot be attributed to anyone.

A common example can illustrate this. Sometimes when using your credit card online or in an ATM, the operation is refused. Since you know who you are and know that there are sufficient funds in your account, you wonder why this takes place. An intelligent artificial agent is to blame, which essentially is an algorithm that takes into account different kinds of information, i.e. your past traveling and buying habits, whether you made any mistakes punching in your security code, etc. Based upon this information, it calculates the likelihood that you could now be in France spending such a large amount of money, and in view of certain objectives (for example, to allow customers access to their money, but also to prevent theft and fraud), decides either to authorize the transaction, or not. To make things worse, suppose that the machine now refuses to give you back your credit card. Who has done that? The artificial agent? But it is only a piece of code which cannot do anything by itself and now, with cloud computing, which is probably properly nowhere! In order to do anything whatsoever, to accept or reject the transaction, to keep your credit card or to give it back, the artificial agent needs to be part of a much larger system that comprises identifiable physical objects with which you interact (the ATM or computer screen), numerous other physical systems which most likely you will never see (data centers), many other systemic intelligent artificial agents in complex interrelation, the internet, banking procedures, and so on. In other words, it needs to be part of a complex system, of which this particular intelligent artificial agent constitutes a minute, though essential part. In order to act, an analytical agent needs to be embodied in some way, but as a systemic agent, the way it is embodied does not individuate it as a particular identifiable physical agent.

So who has done this? Who has refused your transaction and decided to keep your bank card? The answer is 'no one in particular!' What did it was an invisible, mythical entity, the 'system' that is omnipresent because it is nowhere in particular. In this case, it is true that a 'system' did it, but this claim here cannot be reduced to the confused expression of a paranoid conspiracy theory. It is not 'the system' but *a system* that has specifiable characteristic and that has been constructed to achieve particular goals and objectives.

Systemic artificial intelligent agents work for humans, just as individual artificial agents or robots do, but not quite in the same way. The one we have been talking about replaces a bank teller or store clerk in some of his or her functions. It would be more adequate to say that it does away with them, because what was previously done by a person is now not done by anyone at all, but nonetheless happens. The main danger and consequence of the growing number of systemic artificial agents in everyday life transactions is not that they threaten to become too intelligent and to take over the world, but that the more things they do for us, the more tasks they accomplish in our stead, the more they curtail, rather than augment, our ability to act.[7] The way in which they do this is not so much by constraining us directly. Rather, they tend to pre-empt our actions. As a consequence, they often impose that whatever has to be done is either done in the 'one right way' which they provide or not done at all![8] The above statement is not entirely correct. It is not *they*, the artificial systemic agents, who ultimately curtail rather than enhance our freedom of action, because systemic intelligent agents are analytical agents who can only act or do something within a much broader system that ultimately is socially determined. What curtails our ability to act—to the extent to which this is the case—ultimately depends on the reasons why we resort to the socio-technical systems of which they are part, and on the goals which we pursue by institutionalizing these systems.

This transformation of our capability for acting raises fundamental political questions, that we should be asking and that very often we do not ask,

[7] For an interesting and early analysis of the dangers of such agents see Lessig (2006).

[8] For example, if you are living in Japan, try inputting your real address in a webpage which rules that street addresses must always start with numbers!

because we fail to see these socio-technical transformations as the result of social and political choices. Rather we tend to see them as a form of technological determinism and tend to see the problems as merely technological. Because individual artificial agents are social agents, somewhat in the way in which human beings are social agents, individual artificial agents pose quite different types of problems than do systemic artificial agents. Unlike systemic agents, individual artificial agents are not invisible and they are not anonymous, at least not anonymous in the sense that it is always possible to attribute the action to the agent. This difference is fundamental and it suggests a limit to actor-network theory, for in its desire not to erase distinctions between human agents and technical objects, as part of networks of action, actor-network theory overlooks the difference between different types of agents, whether artificial or natural.

References

Arkin, R. 2009. *Governing lethal behaviour in autonomous robots*. New York: CRC Press.

Bolton, C. 2007. The mecha's blind spot: Patlabor 2 and the phenomenology of anime. In *Robot ghosts and wired dreams* (eds) C. Bolton, I. Csicsery-Ronay & T. Tatsumi, 123–47. Minneapolis: University of Minnesota Press.

Budrys, A. 1960. *The unexpected dimension*. New York: Ballantine Books.

Hornyak, T.N. 2006. *Loving the machine: the art and science of Japanese robots*. New York: Kodansha International.

Joy, B. 2000. Why the future doesn't need us. In *Wired 8.04*. (http://archive.wired.com/wired/archive/8.04/joy.html).

Koestler, A. 1960. *The lotus and the robot*. London: Hutchinson.

Lessig, L. 2006. *CodeV2*. New York: Basic Books.

Lin, P., K. Abney & G.A. Bekey (eds) 2012. *Robot ethics: the ethical and social implications of robotics*. Cambridge: MIT Press.

Napier, S. 2007. When the machines stop, fantasy, reality and terminal identity in Neon Genesis Evangelion and Serial Experiments Lain. In *Robot ghosts and wired dreams* (eds) C. Bolton, I. Csicsery-Ronay & T. Tatsumi, 101–22. Minneapolis: University of Minnesota Press.

Varela, F. 1989. *Autonomie et connaissance*. (trans.) P. Dumouchel & P. Bourgine. Paris: Seuil.

Wallach, W. & C. Allen 2009. *Moral machines: teaching robots right from wrong*. New York: Oxford University Press.