European University Institute

MAX WEBER PROGRAMME FOR POSTDOCTORAL STUDIES

EUI WORKING PAPERS

MWP 2013/05
Max Weber Programme

The Capabilities Approach as a Bridge Between Animals and Robots

Migle Laukyte

European University Institute
Max Weber Programme

The Capabilities Approach as a Bridge Between Animals and Robots

Migle Laukyte

EUI Working Paper **MWP** 2013/05

ISSN 1830-7728

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Abstract
In this paper I argue that robots, understood as artificially intelligent and autonomous entities, may one day be regarded as subjects of rights. The argument is built by looking at Martha Nussbaum's capabilities approach, and especially its application to supporting animal entitlements: I draw an analogy between animals and robots by arguing that if what matters is not so much the inherent worth of a (human or non-human) agent, but what an agent is capable of doing so as to live a dignified life, then there is no reason why in the future robots should not have their capabilities recognized and protected as well. The paper unfolds this idea in detail, addressing two main objections to it and arguing that the capabilities approach might be the right way to deal with robots in the future.

Keywords
Artificial intelligence, capabilities approach, robots.

Migle Laukyte
Max Weber Fellow, 2012-2013

Introduction

> There is no doubt whatever that a man could make a machine capable of walking about for some time through a city and of turning exactly at the corners of certain streets. A spirit incomparably more perfect, though still finite, could also foresee and avoid an incomparably greater number of obstacles […and…] it is certain that a finite spirit could be so enlightened as to understand and to foresee demonstratively everything which would occur in a determinate time, so that this spirit could not only construct a ship capable of sailing by itself to a designated port, […] but could also form a body capable of counterfeiting a man. For this involves merely a matter of more or less…"
>
> Leibniz 1976 [1702], 575.

As we can see from the text above, an unquestioned faith in human supremacy was floating over the Europe of the XVIII century. Now, three centuries later, I ask whether this supremacy could accept the idea of robots—artificially intelligent and autonomous machines—as subjects of rights. This is to say that I believe we can, and probably will have to, ascribe rights to robots. To see how this can be done, I propose looking at past examples where new rights have been assigned or new entities have been recognized as right-holders. By doing so, we might gain insights that will help us figure out what new rights can or should be ascribed to robots in the future.

The example I use in this paper is that of animal rights: I consider what experience with animal entitlements can teach us and what part of it we could also apply to robots as potential right-holders. The approach I will use in drawing an analogy between animals and robots is the capabilities approach developed by Martha Nussbaum and applied to supporting both human and animal entitlements. The idea, in a nutshell, is that what matters in considering what rights can be ascribed to what agents (animals or robots) is not so much the agent's inherent worth or dignity but its capabilities, namely, what the agent *can do*. Therefore, instead of thinking about the moral reasons why an agent ought to be protected—as when, paradigmatically, the idea of natural law is coupled with that of human dignity to yield the inalienable rights of most constitutional language—we think about what it means for an agent to be capable of doing something. That is, (*i*) what rights each such ability evokes or calls into play, (*ii*) what consequences the use of those abilities entails, and (*iii*) what duties correspond to those rights or else arise with those consequences.

So as to argue this thesis—that if the capabilities approach can apply to animals, then it (at least theoretically) could also apply to robots—I have organized this paper as follows: in the first part I focus on Martha Nussbaum's version of the capabilities approach. That is, I explain its background and the main ideas related to it, and focus on how this approach has been applied (and this application justified) to support the idea that animals should have similar entitlements to the ones that people have. Then, in the second part, I discuss what exactly the term *robot* means in this paper: I do not give a definition of what a robot is, but instead outline some *sine qua non* characteristics that it must have so as to fall under the umbrella of those robots which I claim might in the future be granted with legal entitlements. In the third part, I put together the first and second parts: I apply the capabilities approach to robots, showing that if it works to justify the idea that we should grant certain rights to animals, similarly it could work to justify similar entitlements to robots. Naturally, the capabilities that animals are recognized as having are different from those that robots could have, yet the first (animal capabilities) can be interpreted so as to include and cover the second (robot capabilities). I dedicate the fourth part of this paper to two main objections to this idea, namely, consciousness and sentience (the ability to feel pain in particular), and argue that these objections are not as sound as they seem, mainly because we simply lack knowledge about both. Then, in the fifth part, I will explain why the capabilities approach could be the best way to deal with future robots, and, as a final part, I close this paper with some concluding remarks and a brief outline for future research.

1. The Capabilities Approach: From Humans to Animals

The capabilities approach may seem self-explanatory, but it would be too limitative to consider it from such a perspective. The approach is much more than just an idea that people and animals have capabilities that we should all respect and ensure. There is much more to it, although for reasons of

space and the purpose of this paper I will not go into a detailed account of the approach,[1] but just briefly describe its main ideas, and then focus on introducing a further development of it, namely the development that the American philosopher Marta Nussbaum matured and applied to explaining why animals should be granted a certain set of rights.[2] Readers who are familiar with Nussbaum's work on this topic can skip this section and go directly to section 2, which is dedicated to clarifying the conceptual mess around the term *robot*.

But before we start, we need to admit that a discussion on capabilities and robot rights could (and should) be looked at in the light of a broader trend in law: namely, since the 20th century in particular we have witnessed the development of a twofold trend concerning human and non-human rights. On the one hand, we have seen an *intensification* of rights: the same social groups are gaining more and more rights. This is the case of immigrants, who just quite recently have obtained healthcare-related entitlements and a right to vote in local elections (at least in Italy), while now the discussion is whether to grant them the right to vote in general elections. Women are also a good example of how the rights of a particular social group can become more *intensive*: once without any legal protection,[3] during the last century they were "discovered" as new subjects of rights, not only to vote, divorce, and own property, but also to such entitlements as education and employment non-discrimination.

On the other hand, we have also been witnesses to the *extension* of rights. That is, rights which once were limited to a certain social group are now recognized not only to numerous social groups but also to corporations and animals. This is the case, for instance, of the right to freedom of speech, which originated as a human right yet in a US court it was recognized that "the First Amendment applies to corporations."[4] This means that the right to freedom of speech was *extended* so as to include not only humans, but also legal persons.

The development of animal rights has, to a certain extent, undergone (and for that matter is still undergoing) both an intensification and an extension: extension in the sense that we, humans, have *extended* to animals some of those rights which once only humans were recognized with, such as the right to be protected from unnecessary torture, which is established in the legal systems of most Western countries; while the intensification of animal rights is supported by the capabilities approach, which I will now focus on. The overall question, reformulated in the light of the last paragraph, is whether we can *extend* some of the animal rights to robots which we might build in the future. The following sections all contribute to supporting a positive answer to this question.

1.1. Human Capabilities

In this subsection I will introduce the main ideas of the capabilities approach, focusing first of all on the human version of it, which was then used as a basis to develop the version for animals (described in subsection 1.2).

---

[1] However, I will give a brief footnote for the reader who would like to learn more about the capabilities approach. The idea of capabilities was first coined by Amartya Sen in his 1979 lecture entitled "Equality of What?", delivered at Stanford University, where he argued that a theory of equality should be built upon "the interpretation of needs in the form of basic capabilities" (1979, 218). For more about Sen's account on the capabilities approach, among other works, see *Inequality Reexamined*. New York: Russell Sage (1992), and *Development as Freedom*. New York: Knoff (1999).

[2] The extension to animals was a later development in Nussbaum's work on capabilities. For more about how this approach was developed, how it was used to deal with gender issues and how it relates to Sen's version of capabilities, see, for instance, Nussbaum, M. C. 2000. *Women and Human Development. The Capabilities Approach*. Cambridge and New York: Cambridge University Press; and Nussbaum, M. C. and A. K. Sen, eds. 1993. *The Quality of Life*. Oxford: Clarendon Press.

[3] As an example, this is how John Stuart Mill (1869, 54) described the women's situation in England: "By the old laws of England, the husband was called the lord of the wife; he was literally regarded as her sovereign, inasmuch that the murder of a man by his wife was called treason (petty as distinguished from high treason), and was more cruelly avenged than was usually the case with high treason, for the penalty was burning to death. […] We are continually told that civilization and Christianity have restored to the woman her just rights. Meanwhile the wife is the actual bond servant of her husband: no less so, as far as legal obligation goes, than slaves commonly so called."

[4] Citizens United v. Federal Election Commission, 558 U.S. 310, 2010.

The capabilities approach,[5] defined by Nussbaum (2011, 18) as an "approach to comparative quality-of-life assessment and to theorizing about basic social justice," tries to give an answer to these fundamental questions: What can people—and each person—actually do and be? What opportunities do persons have? Which of these opportunities should a minimally just society promote? These questions frame the features that Nussbaum highlights as salient in this approach:

- Each person is an end: the approach is not concerned with a general well-being but with the well-being of each person;
- Choice and freedom: a person should be free to choose whether he wants to exercise any of the capabilities; and
- Plurality of value: the value of each capability is different to different people and this multiplicity of meanings that each capability has should be supported, yet all the capabilities aim to ensure human dignity.

Of course, there are more features but for the purposes of this paper these are the most important ones (both to animals and to robots).

So as to make the idea of capabilities more comprehensible and easier to deploy, Nussbaum worked out a list of ten *central* capabilities, that is, a minimal threshold that governments should ensure to their citizens. These capabilities are shown and briefly described in the Table below:

No.	Capability	Which means that one is able to:
1.	Life	live a life of normal length.
2.	Bodily health	have good health.
3.	Bodily integrity	enjoy freedom of movement and decide on one's own reproductive matters.
4.	Senses, imagination, thought	freely use one's mind, cultivate it and express oneself through it; this includes education, free speech, artistic expression, etc.
5.	Emotions	nurture and develop one's emotional world, build such feelings as love, care, friendship, grief, gratitude, etc.
6.	Practical reason	build the notion of good and plan one's own life.
7.	Affiliation	engage in social interactions with other people and be respected as an equal member of society.
8.	Other species	live with and take care not only of people, but also of the vegetative world, fauna and the environment in general.
9.	Play	have fun, enjoy one's activities.
10.	Control over one's environment	participate in political life (the political environment) and own property, seek employment and work, be free from unwarranted search (material environment).

Table 1: Ten central capabilities (Nussbaum 2011).

The Table above only briefly describes the huge content that each of the capabilities covers, yet, brief as they are, the descriptions are defined well enough to reveal the essence of each capability. In the next subsection I will describe how these capabilities were interpreted so as to also apply to animals.

[5] The "capabilities approach" is also known as the "capability approach" and "human development approach" but I use the first version (capabilities approach) because this approach deals with a plurality of capabilities and not only one capability; I do not use the term "human development approach" because I am interested not in human, but in animal and robot capabilities.

1.2 Animal Capabilities

Nussbaum extends this approach to animals[6] by asking the same question: what can animals actually do? She argues that if we observe the capabilities animals already have, and acquire specific knowledge about each of them, then we will be able to create a corresponding—to the human case—capabilities approach for animals: "[…] if we feel wonder looking at a complex organism, that wonder at least suggests the idea that it is good for that being to flourish as the kind of thing it is" (Nussbaum 2004, 306). Indeed, the idea of capabilities goes together with the notions of wellbeing and dignity and, speaking about the latter, Nussbaum refers not only to humans in particular, but also to *agents*, which covers not only persons[7] but includes some species of animals. From this perspective, dignity is a flexible concept which can be extended and attributed to non-human entities.

Concerning the list of ten central capabilities (Table 1), Nussbaum admits that "although this list pertains to human life, its general headings provide a reasonable basis for beginning to think more adequately about what we owe to non-human animals" (2011, 35). This is why, in Table 2 below, I list the ten central capabilities as Nussbaum interprets them in order to apply them to animals as well.

No.	Capability	Which means that the *animal* is able to:
1.	Life	continue its life, with or without conscious interest:[8] no killing for sport, fur, etc.
2.	Bodily health	have good health: no cruel treatment, neglect, etc.
3.	Bodily integrity	enjoy freedom of movement and decide on its own reproductive matters (related to the 2nd capability): no violence, abuse, etc.
4.	Senses, imagination, thought	experience and enjoy different sources of pleasure (species-specific because it depends on the kind of freedom that a particular animal enjoys most).
5.	Emotions	nurture and develop an emotional world, build such feelings as love, care, grief, gratitude, etc.
6.	Practical reason	plan its life, have the opportunity to engage in activities (species-related).
7.	Affiliation	engage in interactions and form attachments (with humans as well), be respected by humans as dignified beings.
8.	Other species	live with other species of animals, humans and the rest of the natural world.
9.	Play	enjoy adequate space, light, sensory stimulation, have company.
10.	Control over one's environment	be respected and treated justly (so that the human guardian is entitled to defend it in court) (political capability); have respect for its habitat and environment (material capability).

Table 2: Ten central capabilities of animals (Nussbaum 2004, 314–317).

[6] Nussbaum's use of the term *animal* includes all animals, with the exception of those with minimal sentience such as amoebas and the like. For the sake of clarity, in this paper the term *animal* refers to mammals only.

[7] "…Human dignity, from the start, is equal in all who are agents in the first place (… excluding those in permanent vegetative state and those who are anencephalic, thus without agency of any kind)" Nussbaum (2011, 31).

[8] For reasons of space, I will not deal with killing for food or predation. Furthermore, all the capabilities described both in this Table and in other Tables will be generic enough to give an idea of what these capabilities are all about, but a detailed description of them will not be included.

The Table above shows that the capabilities we claim for humans and the ones for animals are quite similar. This is why some Western legal systems have translated them into legal tools to protect animals. In the next sections of this paper I will look at whether a similar path could be taken by robots, but first of all I will explain what I mean by the term "robot," as this term can be (and usually is) subject to different interpretations which might lead to misunderstandings.

2. What Are Robots?

In order to continue with the discussion, one particular term needs to be explained: what is meant by the term *robot* in this paper? The term *robot*[9] is used and abused in academia, the fine arts, industry, and other fields, and hence to avoid overlapping and conceptual misunderstandings, this section is dedicated to helping the reader figure out what (who?) should be the new subjects of capabilities.

As a matter of fact, the term—*robot*—is quite vast and applies to almost every tool: we call a robot both a kitchen mixer and an autonomous surgery robot which performs (almost) every task alone with relatively detached control by a human. Furthermore, we also speak about robots when we speak about drones or online chatterbots. And robots are also—just to cite very few—autonomous room cleaners (Yun et al 2012), farmerbots,[10] soccer players[11] and machines assisting autistic children in learning processes.[12]

So, what is a robot in this paper? I would rather not give a definition, because definitions must necessarily be thin and cover a lot of ground, extending across the entire gamut of forms which the thing defined can take. This means that the more we focus on definitions, the more we find ourselves in a situation where we must seek comprehensiveness at the cost of diluteness. It is for this reason that I agree with those computer scientists who find this definitional approach to be vapid, for it divides the world into classes, and says nothing about any of them.[13] A more fruitful approach would be to start from a general intuitive idea of what a robot is by looking at its properties—a robot is able to act, interact, move, perceive, learn, make choices on its own—and see what this means in practice.

For this reason, to explain the term *robot*, I use the idea of action: looking at the kinds of "action" a technology—robot in our case—can perform, makes the particular technology (robot) an agent, that is, an autonomous agent which can act on its own. This is a general idea of action, grounded in the technological context where machines make the transition from passive to active. That is, the more active they are, the more they resemble agents. In Rammert's study (2008, 3) this transition is conceptualized by identifying five basic machine functions, those which consist in moving, acting, sensing, processing information, and communicating; and I should add to this the function of interacting. The change is illustrated in Table 3 below.

Two functions that should be underscored in this Table are those of communicating and sensing, since the transition in these two cases is not from an old way to a new way of doing something, but from an inexistent to a new function. Rammert (2008, 4) expands this list to three, seeing as essential to the transition the capacity for proactive or autonomous action (acting), the capacity for cooperation (communication), and the capacity for responding to context (sensing). This is the so-called "gradual model of agency" and it is also the direction in which current research is heading.

---

[9] I use the term *robot* as a generic term covering a variety of other terms which might be used as (not quite precise) synonyms, such as cyborg, humanoid (gynoid or android), artificial human, artificial agent, etc. In order to avoid confusing the reader, I stick to the term *robot*.

[10] For more about Prospero, a robot farmer, see http://dorhoutrd.com/home/prospero_robot_farmer.

[11] For more about NimbRo-OP, see http://www.ais.uni-bonn.de/nimbro/OP/.

[12] For more about Nao, see http://www.aldebaran-robotics.com/en/Solutions/For-Autism/The-Ask-NAO-initiative.html.

[13] Similarly minded are Saul Amarel, Jerome Feldman, and Randall Davis, three AI researchers for whom there is no need to do research in AI based on an agreed-upon definition of AI. Cf. Bobrow and Hayes 1985, 375–415.

Function	From Passive (machine)	To Active (robot)
Moving	From externally-driven motion, as in the example of a locomotive pulling coaches	To self-driven motion, as in the example of an automobile
Acting	From guided action (based on instructions received)	To autonomous action (where the technology in question is proactive)
Sensing	From blind machines	To machines capable of sensing the environment
Processing information	From devices whose processing is hardwired	To devices that can be programmed in any number of ways
Communicating	From systems whose states are recorded by human observation	To systems capable of observing their own states and communicating them to people and to other systems
Interacting with other machines	From aggregated interaction (based on the combined use of different machines in a single environment)	To integrated interaction (based on the ability of different machines to communicate)

Table 3. Six functions of machines in the transition from passive to active.

What interests us in this Table, however, is that if we look at these six functions and consider the technological transition from passive to active, then we already see that current technologies are *already* becoming more and more human-like: the more a technology becomes capable of "active" action, the more it can be described as an agent technology, and the more the agent (robot) can be understood by analogy to a human agent. This is the robot I have in mind in this paper: intelligent, capable of controlling its own actions, acting on its own initiative, and responding to and interacting with the environment. In other words, this robot is a fully autonomous artificial entity. That is, an entity which can choose not only the means to achieve its goals, but the goals themselves.

Furthermore, the robot of this paper is also capable of distinguishing right from wrong: the ability to make moral judgments is coupled with the competence to justify these judgments, and although these features require consciousness, intentionality and free will—three characteristics that the current state of the art in robotics and artificial intelligence are not yet able to engineer—this does not mean, however, that they will never be built into artificial entities.

One may argue that this is not possible, and yet it seems that not only artificial intelligence research but also research in neuroscience is working in a direction that makes the whole idea of this paper much more realistic. For instance, The Human Brain Project[14] involves approximately 200 institutions around the world in an initiative to build a new ICT infrastructure for neuroscientists, physicians, and computer scientists so as to understand how the human brain works and eventually emulate its computational capabilities. What if they succeed and build an artificial brain? Would it then not be natural to try to build it into a tool and create an artificial being? In fact, the platform of neurorobotics is dedicated to building brain models into virtual robots, so why not work with physical robots as well?[15]

[14] For more about The Human Brain Project, which was recently awarded a billion dollars for 10 years by the European Union, see its website at http://www.humanbrainproject.eu/in_brief.html.

[15] However, the project itself is quite critical in this regard, stating clearly that it is quite improbable that any sentient system or robot will come out of it, yet it is not impossible either.

Furthermore, there are also scientists who think that we are close to building an artificial brain. For instance, Kurzweil (2012) argues that it is possible to engineer the principles of the neocortex—the part of the brain which is responsible for memory, critical thinking and creativity—thanks to the pattern recognition theory of mind—which describes the algorithm governing the neocortex—and the law of accelerating returns, according to which the outcomes of technological and evolutionary processes are more and more complex and sophisticated. This is why he argues that "mathematical techniques that have evolved in the field of artificial intelligence (such as those used in Watson[16] and Siri, the iPhone assistant) are mathematically very similar to the methods that biology evolved in the form of the neocortex. If understanding language and other phenomena through statistical analysis does not count as true understanding, then humans have no understanding either" (Kurzweil 2012, 7). We have succeeded in working out and engineering the key functions of the auditory cortex (this is why, for instance, we can give voice commands to robots), the visual cortex (this is why robots can move around obstacles), and the cerebellum (this is why robots are able to catch a ball): we are still working on the last one—the neocortex—but maybe victory is just around the corner.

What conclusions can be drawn from looking at these examples? My answer to this question is that although we have not yet succeeded in building autonomous and intelligent robots, discussion of applying the capabilities approach to robots could be seen as a foresight discussion: we can accept the possibility of such robots existing at least hypothetically and so it might be worth discussing their (and our) future.

For this reason, in the next section I will apply the capabilities approach to robots and discuss whether this matching is sound.

3. The Capabilities Approach Applied to Robots

Could the capabilities approach apply to robots? I will use Nussbaum's words (2011) when she says that what matters for the success of an approach is not the number of supporters, but the quality of the arguments: in this section of the paper I will try to give valid arguments why the capabilities approach—a theoretical approach created for humans and extended to animals—could in the future be applied to robots as well.

Moving away from an entirely human-oriented vision of the capabilities approach, Nussbaum (2011) suggests five different positions in relation to the interaction between human capabilities and the capabilities of other entities:

> 1. Only human capabilities count: the capabilities of others are valuable as long as they promote and support human capabilities;
> 2. Human capabilities are the main goal, yet the capabilities of others could gain value not only as means to promote human capabilities (1), but also as part of valuable relationships with these capabilities;
> 3. As long as the *other* is a sentient being, its capabilities count equally with the capabilities of other sentient beings;
> 4. Capabilities of all living organisms count as an end;
> 5. Capabilities of (eco-) systems count as an end.

These positions describe a linear progress from the view of non-human capabilities as being means to an end to a vision of them as an end: we move from an extreme position of human supremacy (1) to the inclusion of ecosystems which might not be sentient at all (5). Animals are situated in positions (1) to (3). But where should we place robots?

I would consider this paper successful if at least one reader considered it possible to place robot capabilities within the (1) position, and although this position is purely human-centric, it would

[16] Watson is a computer system created by the IBM DeepQA project which represents the latest advances in natural language and machine processing, knowledge representation and reasoning, and other fields of research. For more about Watson, see http://www-03.ibm.com/innovation/us/watson/.

at least couple the idea of capabilities with robots and admit the possibility that the capabilities of an individual robot could count for something (although only in the light of human interest).

Coming back to animals—my reference point of reasoning why robots might have their capabilities recognized in the future—Nussbaum's suggestion to include animals in the list of entities whose capabilities are worth ensuring is based on the idea that any being capable of "any kind of agency or striving accompanied by sentience" (2011, 88) is worth its capabilities being ensured. It is not even important that animals cannot understand nor assess them. But can the concept of "any kind of agency" be attributed to robots too? And if we speak about animal dignity—which Nussbaum links to the "characteristic form of agency of their kind" (2011, 88)—should we not be able to speak about robot dignity as well? Can we not draw the conclusion that if an entity has agency, then it has dignity, and consequently it is entitled to have its capabilities recognized and protected? The big question then is: are robots agents?[17]

The answer to all the questions in the paragraph above is *yes*: in the section dedicated to robots, I have explained why they are agents in their own right and their capacity for (active) action cannot be questioned. What this agency requires us to do is to come up with a list of species-specific capabilities for animal species and for inanimate species, such as the robots I am concerned with. The list of ten central capabilities (Table 1) is a very general list and, hence, open to interpretation. Could the capabilities identified for humans, suitably broadened for animals, be extended in order to cover robots as well? Table 3 is my answer to this question.

No.	Capability	Which means that a robot is able to:
1.	Life	exist without being destroyed or broken for futile motives.
2.	Bodily health	have full maintenance and updates.
3.	Bodily integrity	have an existence without being subject to violence or any other type of harmful treatment.
4.	Senses, imagination, thought	Learn from experience.
5.	Emotions	_____
6.	Practical reason	have goals and plans for itself, behave rationally.
7.	Affiliation	function with and towards others (humans, animals, robots).
8.	Other species	relate to other species (humans, animals, robots).
9.	Play	_____
10.	Control over one's environment	_____

Table 4: Ten central capabilities applied to robots.

I have deliberately left a few capabilities without explanation—namely, capabilities of emotions, play and control over the environment—because it is not easy to imagine how these capabilities could be translated for robots, and also because for this very first draft of robot capabilities the complete list may appear as an overstatement.

Yet what is already clear from this Table is that some of the capabilities—despite being interpreted differently for humans, animals and robots—could be considered common to all of them:

[17] Furthermore, Nussbaum also argues that "more complex forms of life have more and more complex capabilities to be blighted, so they can suffer more and different types of harm" (2004, 309). Without entering into a discussion over what the term "form of life" should cover, can we not argue that robots could have the necessary level of complexity to be entitled to have the capabilities which would enable them to enjoy this complexity as completely as possible?

harm can be inflicted on all three groups, and all three groups (in different ways and to different extents) interact, learn from experience, and have a purpose. These similarities should be seen as only slight indications of more commonalities among these groups that will be found in the future, as the first two groups—humans and animals—have reached a certain level of evolution, while the third group—robots—is catching up with the first two at high speed.

In the next section I will deal with two main objections to this analogy, arguing that despite their seriousness they can be circumvented.

4. Objections to a Parallel Between Animal and Robot Capabilities

The objections I will consider to test the soundness of an analogy between animals and robots are those of consciousness and of pain, for these are two animal traits which robots are claimed to lack. I reply to these objections, but I also note that each of the capabilities that may be offered as a relevant criterion for an ascription of rights and for an accompanying interspecies comparison is bound to invite scepticism. The point is not really to devise a counter-objection for each possible objection one may have, for the capabilities approach is rather to be understood as a macro-criterion—that of an agent's capabilities as a source for its rights and as a gauge by which to assess points of analogy between different species such as animals and artificial entities. It is then up to us to shape the macro-criterion into a workable and convincing conception.

4. 1 The Criterion of Consciousness

The objection of consciousness is not new, nor are robots the only entities to be denied it: before robots, the same objection was moved against animals by such illuminated thinkers in human history as Descartes.[18] Now we have changed our minds: animals—not only mammals, but birds and reptiles as well—have been recognized as having consciousness (Griffin 2001), although we have not yet managed to build conscious robots. Furthermore, (for the moment) this seems to be the insurmountable obstacle for artificial intelligence and computation, and some thinkers even argue that it will never be possible (Searle 1992, Edelman and Tononi 2000).

What needs to be said about consciousness is that we still know very little about it: it is still a black box for us and we have no agreed-upon criteria or parameters to measure it in a being. Yet although we do not have any fully conscious robots, we should also admit that some—admittedly little—steps forward have been made. Consider for instance the use and understanding of natural language—which Descartes thought of as one of the main reasons for treating animals as not conscious—where robots perform quite well. For instance, Watson—the intelligent system developed by IBM—understands natural language in such a way that it is able to detect double meanings, take hints and comprehend word games. Is this not consciousness in a certain way?

Concerning consciousness as such, Hart and Scassellati (2012) are pursuing research into human self-awareness using a humanoid robot, Nico, which learns about its body and senses by itself (while usually this knowledge is built in by the software engineer). This robot emulates the forms of self-awareness that we develop as children when we discover our body parts, how these parts are related and what changes these parts bring about in objects when we interact with them. If Nico succeeds in learning about itself by observing its body, then it will acquire the very important capability of passing the Mirror Test, which is a proof of self-consciousness in animals.[19] Only dolphins, orcas, elephants, magpies, humans and a few other apes have passed the test so far, and if Nico joins this club of self-conscious entities—for the time being it is only able to recognize its "hand"—an enormous step toward a self-conscious artificial entity will have been taken.

---

[18] Furthermore, Descartes took an opposite direction and compared animals to automata, arguing that they do not have a mind, although he admitted that they have sensations, which however are in no way conscious experiences. For more on this, see Descartes, R. 1985 [1641]. *Meditations on First Philosophy*. Ed. J. Cottingham. Cambridge: Cambridge University Press.

[19] The test is performed as follows: the animal is anesthetized and while it is sleeping, a person puts a red spot on its head. Once the animal is awake, it is put in front of a mirror and if it removes the spot after looking at the mirror it is considered to be self-conscious.

Another example of attempts to understand and engineer consciousness is LIDA (Learning Intelligent Distribution Agent), a computational and conceptual model of cognition based on the Global Workspace Theory (GWT)—one of the most widely acknowledged theories of consciousness—which is being developed to run a machine version of biological algorithms which produce conscious and unconscious aspects of human thinking and perception (Baars and Franklin 2009). No less impressive is the development of the ConsScale project.[20] ConsScale is a consciousness metric to assess the level of consciousness of both robots and virtual agents. It is inspired by human consciousness but it does not build a conscious robot by itself: it only provides a kind of roadmap towards the further design of artificial cognitive machines (Arabales et al. 2009).

What we see in this section is that consciousness is on the one hand the main obstacle to even thinking about granting robots any consideration, and on the other hand is not such an obstacle for computer scientists and software engineers to believe it is impossible to overcome. Consequently, this objection can only be partially accepted, because the examples described above clearly show that we are taking small steps toward conscious machines. Furthermore, how can we use the objection of consciousness if we cannot even say what consciousness means? Given the impossibility of opening the black box of consciousness, we are left with no other choice than to approve the functional consciousness of a robot. As Levy (2009, 211) argues, "if a machine exhibits behaviour of a type normally regarded as a product of human consciousness […], then we should accept that that machine has consciousness." Looking at the above examples, this might only be a question of time.

Having discussed the criterion of consciousness, in the next section I will deal with the criterion of pain, according to which only humans and animals are held to be able to feel pain, giving them entitlements that robots should not be granted.

4. 2 Criterion of Ability to Feel Pain

One of the strongest criticisms against the idea of robot capabilities (and in general against any idea of granting robots any more status than that of a tool) is that robots cannot feel pain in the way we, sentient beings, feel it. This is why we have granted rights to animals: we know that they feel pain, experience hardship and can be hurt, while robots, not being made of flesh and blood, feel nothing.[21]

Firstly, although we have not yet built robots able to feel pain, sensory technologies are at the centre of many studies. For instance, if we describe the feeling of pain in relation to specific neurons which detect noxious stimuli, then research which aims to label certain robot stimuli as *unpleasant* is a small step toward the development of the idea of artificial pain (Wallach and Allen 2009).

However, there is also another side to the objection that robots cannot feel pain and it is related to the question of whether pain is species-specific and whether we, humans, accept only the pain which we recognize. In other words, do we suffer from bias against non–human pain? And if so, are there any other types of pain that are unknown to us yet which might exist?

Furthermore, pain is not the *condicio sine que non* for suffering: we and animals can suffer without feeling pain, so how relevant is pain? If we inflict a malfunction on a robot, could this amount to suffering? This problem is very much related to consciousness: to feel pain is about being sentient, to suffer is about being conscious of it. This brings us back to the discussion in the previous section: we know as little about suffering and pain as we know about consciousness. Consequently, if we interact with a robot which behaves as if it suffers or feels pain, we will consider it to experience (some sort of) these feelings. If Cathexis architecture (Becker-Asaro 2008), for instance, were extended to represent not only emotions, such as fairness, anger, distress, happiness, disgust and surprise, but also such feelings as pain, would we not agree that a robot incorporating this architecture has something that we could consider to be sentience?[22]

---

[20] For more about this project, see http://www.consscale.com.

[21] In fact, it was Bentham (1823, 236) who identified the ability to feel pain as the single most important criterion for ascribing rights to animals: "The question is not, Can they reason? Nor Can they talk? But Can they suffer?"

[22] Furthermore, research into building sentience and emotions into robots could also be seen as beneficial to humans: we know that people with serious brain injuries who are unable to experience emotions (including pain) are dangerous not

Having discussed these two critical facets of my idea, I would like to draw the reader's attention to another critical aspect of my reasoning. Robots may differ from animals in one very substantial way: one day they might claim their entitlements on their own. This is something animals have never been able to do and yet they have succeeded in convincing people to recognize their capabilities without verbal communication, argumentation or discussion. Robots may surpass them in this respect and be spokesmen for their own case: in the end, intelligence is one of the key characteristics that I attribute to future robots and so it is absolutely logical to imagine that they will use this intelligence to promote their wellbeing and to improve their position. However, this need not mean war (forget *Terminator* and the like!), but a dialogue which should lead to a kind of equilibrium between the (natural and artificial) species inhabiting the world of the future: the key challenge for us is to switch from a policy of *exclusion* and focus on the concept of *sharing*. This is another reason why the capabilities approach could be the right approach. The next section will further develop the topic of the suitability of this approach.

5. On Why the Capabilities Approach is the One to Deal with Robots

The capabilities approach is a suitable way to deal with future robots because it proposes a holistic approach to the capabilities of different entities: respecting each others' capabilities would create a single environment where humans, animals and robots could co-exist. Much as my ideas may seem futuristic, science-fiction or (at worst) impossible and bizarre, the capabilities approach could better prepare us to embrace fluidly and coherently what appears to be an inevitable transition toward an environment marked by a closer and closer interaction between humans, animals and robots, an environment of which robots are an integral part rather than an appendage. The premises for the suitability of the capabilities approach have been set out by Nussbaum (2004, 319), who argues that "truly global justice requires not simply that we look across the world for other fellow species members who are entitled to a decent life. It also requires looking around the world at the other sentient beings with whose lives our own are inextricably and complexly intertwined", and although we might not all agree on the sentience of robots, we should all agree on the idea that we, humans, should look around for those with whom we interact and to whom we are related. There is no doubt that the way we interact with machines is changing substantially and that "even though our mind might know that this is just a machine, our behaviour towards another creatures is very intuitive, [and] we aren't even emotionally capable of distinguishing between alive, animated, or stuffed" (Benford and Malartre 2007, 165).

There is also another reason why the capabilities approach could be of use in dealing with robots: the classical theory of the social contract is based on the idea that it is in our own interests to surrender some of our privileges in order to have our other assets ensured. We may have to make a similar contract with robots too: in recognizing and respecting the capabilities of their species, we would ask them to recognize and respect ours. Needless to say, in practice this agreement would involve a series of different and difficult secondary questions, but this is just an underlying idea explaining why the capabilities approach might be useful for finding the right equilibrium among the members of future society, which might include not only humans but also robots.

6. Concluding Remarks and Outline for Future Research

The capabilities approach is not the only theory which could be used to advocate robot rights. John Rawls' theory of justice (Rawls 1971), for instance, does not apply to animals because it only relates to rational persons (this is why it excludes cognitively disabled people), but if we could manage to create rational robots then his theory could (theoretically) apply to robots as well. Similarly, animal rights are not the only parallel with robots: the natural environment could be another example. In fact, Teubner (2007, 16) argues that "the inclusion of ecological rights in political institutions, […], the change in legal language from the semantics of 'protection of nature' via 'ecological interests' to 'rights' of living processes, the slow process of granting standing to ecological associations, the

(Contd.)

only to themselves but also to those around them. If this is so, should not emotion-based engineering in computer science research be supported?

expanding conceptualization of ecological damages without attribution to an individual are indicators that the law is preparing again to create a new breed of actors." He is just one of many scholars who consider nature and the environment next in the queue to become legal subjects.[23] At this point, robots too might have claims for their (legal) consideration. This is clearly a distant possibility, yet it is not as impossible as it may seem, especially if we adopt an optimistic view of the developments and promises of artificial intelligence, neuroscience and other fields.[24]

It goes without saying that before the capabilities approach is seriously taken into account concerning animals—not to mention robots—it must be applied for all human beings. For the time being this is not (yet) the case, as many people are deprived of their capabilities not only in third world countries but also in the most developed societies of the western world. For instance, women and sexual minorities are still fighting for the recognition and respect of their entitlements.

For this reason, the discussion in this paper might seem a distant science fiction topic which is not (and might never be for that matter) a real problem. I agree with this criticism: I do not say that the present is the right time to deal with these hypothetical problems, but I do believe that it could be the right time to think about them.[25] This is especially because enormous investments are being made in both academic and industrial research which should eventually lead to some tangible results in the field of artificial intelligence, robotics, bionics, and the like. For what purpose? Why are we trying to build an artificial mind?

Concerning future research, there is an enormous variety of questions, problems and issues to take into account if this theory of robot capabilities is to stand on its own. For instance, if we accept the capabilities of robots, how would these capabilities interact with the framework of already established capabilities of humans and animals?

I would like to close this paper with another conclusion that I came to while doing research for it: the whole discussion is not about animals or robots, it is about us. We grant rights, recognize entitlements, and ensure capabilities only when we discover something of ourselves in the candidate to these assets: this is what has happened with animals, and this is also what may lie in the background to granting some entitlements to robots as well. If we see in robots some positive human traits which give rise to human empathy, then the way towards recognizing them as members of our society will be easier. Only once we are able to identify with them will they have a chance to become something more than a tool. Consequently, this is less about them but more about us: about what we are able to understand, with whom we are able to build relationships, how we see otherness, how open we are to questioning our assumptions, etc. Nussbaum (2011, 180) observes that concerning the human capabilities approach, what needs to be done is to work out a "political psychology—an account of the emotions and other psychological dispositions that support and impede a program of realizing human capabilities." Needless to say, the same should be done to support a programme of animal capabilities, and consequently those of robots. For the moment, the question of robots is not yet an issue. However, to think about it does no harm; on the contrary, it might be useful to realize that granting capabilities is not a finite action but a process. Who is next on the list, only the future will tell.

---

[23] The first on this topic, as far as I know, was C. D. Stone (1972), who was as ahead of time as to ask whether rights should not one day be granted to such non-living entities as computers. Meanwhile, last summer the Whanganui river, one of the biggest rivers in New Zeland, was recognized as having a legal identity (for more details, see Shuttleworth 2012).

[24] For instance, in the field of neuroscience, according to C. Koch and R. C. Reid (2012, 397) there are more than 10,000 laboratories worldwide which are studying different aspects of the brain "heading away from each other in all directions, in a sort of scientific Big Bang."

[25] Indeed this is another parallel with animals, because animal rights also had to overcome various obstacles before being taken into account, despite their relationship to humans having been an issue not only for the Ancient Greeks and Romans but also for Montaigne, Voltaire, Bentham, Mill, and Shaw (to cite just a few). The same approach could be observed with robot rights as more and more scientists and thinkers (such as Lehman-Wilzig, Solum, Asaro, Chopra, and many others) are discussing and supporting the idea of their entitlements.

References

Arrabales, R., Ledezma A., and A. Sanchis. 2009. Establishing a Roadmap and Metrics For Conscious Machines Development. In *Proceedings of the 8th IEEE International Conference on Cognitive Informatics*. Eds. G. Baciu et al, 94–110. Hong Kong: IEEE Computer Society Press.

Baars, B. J., and S. Franklin. 2009. Consciousness is Computational: The LIDA Model of Global Workspace Theory. *International Journal of Machine Consciousness* 1(1): 23–32.

Becker-Asaro, C. 2008. *WASABI: Affect Simulation for Agents with Believable Interactivity.* Heidelberg: Akademische Verlagsgesellschaft.

Benford, G., and E. Malartre. 2007. *Beyond Human. Living with Robots and Cyborgs.* New York: A Tom Doherty Associate Book.

Bentham, J. 1823. *An Introduction to the Principles of Morals and Legislation*. Vol. 2. London: Pickering.

Bobrow, D. G., and P. J. Hayes. 1985. Artificial Intelligence: Where Are We? *Artificial Intelligence* 25: 375–415.

Citizens United v. Federal Election Commission, 558 U.S. 310, 2010. Available at http://www.supremecourt.gov/opinions/09pdf/08-205.pdf.

Edelman, G. M., and G. Tononi. 2000. *A Universe of Consciousness: How Matter Becomes Imagination.* New York: Basic Books.

Griffin, D. R. 2001. *Animal Minds: Beyond Cognition to Consciousness.* Chicago: Chicago University Press.

Hart, J. W., and B. Scassellati. 2012. Mirror Perspective-Taking with a Humanoid Robot. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, eds. Jörg Hoffmann, Bart Selman, 1990-96. Palo Alto: AAAI Press.

Koch, C., and R. C. Reid. 2012. Observatories of Mind. *Nature* 483: 397–398.

Kurzweil, R. 2012. *How to Create a Mind: The Secret of Human Thought Revealed*. New York: Viking.

Leibniz, G. W. 1976 [1702]. *Philosophical Papers and Letters*. Ed. L. E. Loemker. Dordrecht: Kluwer Academic Publishers.

Levy, D. 2009. The Ethical Treatment of Artificially Conscious Robots. *International Journal of Social Robotics* 1(3): 209–213.

Mill, J. S. 1869. *The Subjection of Women*. London: Longmans, Green, Reader, and Dyer.

Nussbaum, M. C. 2011. *Creating Capabilities. The Human Development Approach.* Cambridge, London: The Belknap Press of Harvard University Press.

Nussbaum, M. C. 2004. Beyond "Compassion and Humanity:" Justice for Non Human Animals. In *Animal Rights. Current Debates and New Directions*. Eds. C. R. Sunstein, and M. C. Nussbaum, 299–320. New York: Oxford University Press.

Rammert, W. 2008. Where the Action Is: Distributed agency between Humans, Machines, and Programs. TUTS: Working Papers, 4. Available at: http://www.soz.tuberlin. deTuts/Wp/TUTS_WP_4_2008.pdf.

Rawls, J. 1971. *The Theory of Justice*. Cambridge, MA: Harvard University Press.

Searle, J. 1992. *The Rediscovery of Mind*. Cambridge, MA: MIT Press.

Sen, A. 1979. Equality of What? *Tanner Lectures, Stanford University*, 195–220. Available at http://www.uv.es/~mperezs/intpoleco/Lecturcomp/Distribucion%20 Crecimiento/Sen%20Equaliy%20of%20what.pdf.

Shuttleworth, K. 2012. Agreement Entitles Whanganui River to Legal Identity. *The New Zealand Herald*, 30 August 2012.

Stone, C. D. 1972. Should Trees Have Standing? Toward Legal Rights for Natural Objects. *Southern California Law Review* 45: 450–501.

Wallach, W., and C. Allen. 2009. *Moral Machines. Teaching Robots Right From Wrong*. New York: Oxford University Press.

Yun J. et al. 2012. Learning to place new objects in a scene. *International Journal of Robotics*, 31(9):1021–43.