Reflections on the EPSRC Principles of Robotics from the New Far-Side of the Law

Voiculescu, A.

This is an Accepted Manuscript of an article published by Taylor & Francis in Connection Science, 29 (2), pp. 160-169. The final definitive version is available online: https://dx.doi.org/10.1080/09540091.2017.1313818

Reflections on the EPSRC Principles of Robotics from the New Far-Side of the Law

Aurora Voiculescu
Law & Theory Lab, Faculty of Social Sciences and Humanities,
University of Westminster, London

The thought-provoking EPSRC Principles of Robotics stem largely from the reflection on the extent to which robots can affect our lives. These comments highlight the fact that, while the Principles may address to a good extent the present technological challenges, they appear to be less immediately suited for future technological and conceptual dares. The first part of the paper is dedicated to the search of the definition of what a robot is. Such a definition should offer the basic conceptual platform on which a normative endeavour, aiming to regulate robots in society, should be based on. Concluding that the Principles offer no clear yet flexible insight into such a (meta-)definition, which would allow one to take into account the parameters of informed technological imagination and of envisaged social transformation, the second half of the paper highlights a number of regulatory points of tension. Such tensions, it is argued, stem largely from the absence of an appropriate conceptual platform, influencing negatively the extent to which the Principles can be effective in guiding social, ethical, legal and scientific conduct.

Keywords: principles of robotics; artificial intelligence; human-robot interaction; human rights and technology; technology and regulation;

1. Introduction

The thought-provoking EPSRC Principles of Robotics initiative (EPSRC, 2010) (hereinafter referred to as EPSRC PRs or simply PRs) stems largely from the reflection on the extent to which robots already affect our lives and on the even greater extent to which it is expected that they will affect it in the relatively near future (Roser, 2016).[1] It is understood that the initiative represents an attempt to address public anxieties related to the future of human-robot interactions. The principles, therefore, aim to reassure the public concerning the extent to which the advantages of using increasingly sophisticated robots, in a variety of contexts, are not going to be outweighed by disadvantages.[2] The agenda to be fulfilled by the principles is thus one that combines both the safety and the acceptability of robotic technology, supporting the governmental policy towards making robotics a technology of critical importance for the UK (Bryson, 2012). The question of whether the initial regulation related to this transformative technology should take the shape of *soft*, guiding principles, of *hard* domestic legal instruments, or even of complex international treaties is a challenging yet, at this point, a secondary issue. The primary issue is, rather, a (legal-) normative one, aimed at delineating clear boundaries of the human-robot co-existence; addressing the normative dynamics of causality, intentionality and responsibility; trying to identify the *locus* of *intention* and *action* in processes

and, dare we say, *relationships* that may well prove to become increasingly complex with the advancement of science and technology (Voiculescu, 2016).

Stemming from this need for normative introspection, this paper is an invitation to reflection on the proposed Principles of Robotics coming out of the multi-disciplinary expert-informed EPSRC and AHRC Robotics Retreat in 2010. The complexity of issues to cover is such that these reflections can only aim to engage with the text of what is proposed, prising out some of the possible meanings or interpretations available there. Such an analysis is proposed as essential for preparing the ground for further discussions, and finally, for embarking on any eventual regulatory processes.

2. Principles in Search of a Definition

Reflecting on the existing principles does invite one, first of all, to address what a robot is and to decide whether the definition that one should settle for and, therefore, the type of entities or phenomena that one should aim to regulate, should be an answer to our state-of-the-art in technology or a reflection of our state-of-the-(technology) in art. In other words, on which point on the spectrum between science and (informed) imagination should we place ourselves when designing norms to govern technological advances and when evaluating these norms' effectiveness? How far into the future should one look, when the future for which we regulate is so far that we can only speculate as to its existence, while at the same time we are galloping towards that very future at an ever greater speed (Roser, 2016)?

Law and regulation, hard or soft, requires, even thrives on definitions. It is the law's way of incorporating a phenomenon into *its* reality, since a definition is never value-neutral, but always establishes the 'in'-s and the 'out'-s, following a more or less declared value-laden path (Norrie, 1991, p. 687). The EPSRC Principles under discussion here do not immediately send to a definition, although they do offer a number of elements of what robots *ought* to be seen as: 'tools' (PR1), 'products' (PR3); 'not responsible agents' (PR2); 'just tools, designed to achieve goals and desires that humans specify' (commentary to PR2); 'simply not people', 'pieces of technology' (commentary to PR3); 'manufactured artefacts' (PR4). Robots, however, cover a rather wide spectrum, the AI that powers them can vary from the very basic to the very complex and the current principles appear to cover only partially this spectrum.

If we look outside of the EPSRC document for an existing definition that may have been taken for granted, we are not going to find much help. A robot is defined in many different ways. Some dictionaries define it as 'a machine capable

of carrying out a complex series of actions automatically' or, differently nuanced, as 'a mechanical or virtual artificial agent, usually an electro-mechanical machine that is guided by a computer program or electronic circuitry' ("Merriam-Webster Dictionary," 2016). Various, more or less workable distinctions are also put forward as part of other definitions; more notably one encounters the distinction between industrial and service machines, between highly autonomous machines and cognitive computer programmes, between embodied and dis-embodied cognition entities, to mention just some of the most widely available distinctions. NASA itself uses a rather mundane and imprecise language, very unhelpful for the regulator, defining robots as 'machines that can be used to do jobs'. Some robots, NASA's formulation goes on, 'can do work by themselves. Other robots must always have a person telling them what to do' (NASA, 2015). Such a variety of formulations creates a regulatory minefield and makes any stemming normative statements difficult to follow and easy to avoid complying with.

While agreeing that there is no agreed definition *per se* (Wilson, 2015), certain other attempts put forth a number of features that a robot would have, features that, from a regulatory (and not only) perspective, are themselves in need of definitions: *sensing* the surroundings (having inbuilt 'awareness' of its environment); *movement*, whether rolling, walking, thrusting, or maybe even *just* data-conveying; *energy*, being able to power itself in ways that will depend on what the purpose of the robot is; *intelligence*: being provided with 'smarts' by its programmer, having the capacity to evaluate surroundings, circumstances, complex information.

A robot is, therefore, defined more specifically as a system, a machine that '*contains* sensors, control systems, manipulators, power supplies and software, all working together to perform a task' (Merzouki, Samantaray, Pathak, & Ould Bouamama, 2013, p. 619). According to such a perspective, '[d]esigning, building, programming and testing a robot is a combination of physics, mechanical engineering, electrical engineering, structural engineering, mathematics and computing. In some cases biology, medicine, chemistry might also be involved'. If the student in robotics may actively engage with all of these disciplines 'in a deeply problem-posing problem-solving environment' (Wilson, 2015), some could rightly say that regulating robots and the robot-human interactions requires a similarly complex interdisciplinary engagement with most if not all of these fields. For the normative discourse, whether hard regulatory or soft-principled, the fact that many of these definitions have a number of points in common is not sufficient for securing consistency and generality in the application of the eventual principles.

One of the main challenges and aims of robotics is to create machines that are capable of independently interacting with a dynamic world.[3] Scaling-up, moving

towards a human-like level of intelligence will take the machine past the easy classification stage, since scientists are already looking into robot intelligence architectures that eschew traditional pre-programming as a solution to meeting the 'human-like' goalpost.

It appears, therefore, that the essence of a robot, so to speak, is either taken for granted – not a very useful platform for a regulator to start from - or is defined through what the robot contains (or not) and how it is made, rather than through what a robot actually *is* and potentially *may be*. If we add to this the tendency towards legal-normative reductionism, that is the tendency of the law to *look* at a phenomenon through pre-existing legal conceptual constructions from within its own environment, the ontological reality of the new technologies are going to struggle to be acknowledged. The new technology has to interact with a social and legal world that is already in place, thus displaying conceptual and normative inertia (Balkin, 2015, p. 50).

A definition that is sufficiently precise, yet dynamic enough to capture the essence of the socio-technological phenomena is, therefore, needed for opening the robotics principles to further development and problematisation and, ultimately, for allowing a more complex conceptualisation of the human-robot interaction (Kahn et al., 2007). This need relates to perspectives such as the one highlighted by Andra Keay,[4] who speaks about robots as '… *an environment*; too large for us to look at as an item'. While inevitably linked to the progresses of technology – '[what] we call a robot today is more sophisticated than what we called a robot in the '80s' says Keay - it is also true that a robot is more than that. 'It has always been *an identity issue*' Keay declares (Brewster, 2014). From such a position, it is understandable that any soft or hard regulatory initiative is bound to be met with a rather steep challenge of capturing the essence of a rapidly changing environment. Unsurprisingly, the concern is that the technological advances are taking place at such a speed and with such opacity that it will take hold of our lives, our privacy, our emotions, 'before our governments and our laws are ready for it' (Barraza, 2016).

While this special medium awaiting normative ordering brings its own challenges, it should be said, however, that identities and classifications encapsulated in definitions – the law's bread and butter - have often been problematic when promoted as part of regulatory initiatives and are almost always problematised when put forth on a dispute resolution platform, whether these definitions have to do with humans or non-humans alike. Law, in particular, always ends up transforming any identity into a legal fiction that often has very little to do with any other physical or scientific dimension of that entity (Fagundes, 2001). Yet, law - in its widest sense, as socially backed normative imperatives - has always thrived on definitions. The absence of a working

definition of a robot, conveying flexible yet coherent complexity, to be eventually incorporated into a preamble of the normative document, appears to be both a witness to the challenges of pinning down technology in its rush, and a reflection of a possible weakness to be addressed in the proposed document. Of course, once formulated and placed at the heart of the normative initiative, a definition may be contested and problematized. However, this does not make the definition any less of a necessity as a key normative coagulation point.

Last but not least, in addition to the absence of an agreed working definition, there is also the acknowledgement that, as mentioned earlier, definitions are never neutral and, it could be argued, nor is their absence. Scholars, already decades ago, brought this idea to bear on defining the responsibility of corporate organisations as collective agencies (Voiculescu, 2017). Since this normative and legal innovation takes the notion of responsibility beyond the sphere of the discrete individual human being, it also touches on our discussion here. While the absence of key definitions leaves the normative space prey to being populated by the hegemonic discourse-setting voices, the provision of definitions advances *pseudo-unities* that are proposed as facts. Such pseudo-unities set up oppositions that 'arbitrarily separate those who are included and those who are excluded from a shared conceptualisation or practice' (May, 1996, p. 171ff). In the case of robotics, such arbitrariness of definitions will become more and more evident with the passing of time, once the AI spectrum is enlarged (Grady, 2014). Treading between a definitional vacuum and pseudo-unities, the reflection on definitions ought, therefore, to be at the heart of any normative principles for robotics.

3. Principles in Practice

Whether anticipated with dread or with excitement, the challenges of regulating the multiple dimensions of the human-robot interactions are significant. A number of issues are put forward here for reflection in relation to the five EPSRC Principles of Robotics, issues deriving more or less directly from the absence of clearer definitions and a clearer regulatory conceptual framework.

In an attempt to maybe promote their commonness and enhance their acceptability, the PRs declare robots to be (simply) multi-use tools, (PR1): *Robots are multi-use tools. Robots should not be designed solely or primarily to kill or harm humans, except in the interest of national security.* While the principle aims to be simple and clear, it should also be effective in guiding conduct, and this is where the principle may be deemed to be less than ideal. First of all, as a safety-centred principle, PR1 appears rather confusing. Declaring robots as '*multi-use tools*' is virtually a dimension that does not serve an actual purpose. It is unclear why a robot has to be multi-use in order to be safe or, conversely, in what way an

otherwise deadly robot may become any less deadly if designed as multi-use. This relates to the next part of the principle: *'robots should not be designed solely or primarily to kill or harm humans*'. Were one to actually aim to build 'killer robots', in order to bring such an aim to compliance with the letter of the above principle, it would suffice to teach such robots not only to kill but also to make pancakes or knit woolly socks. In the absence of over-arching clear definitions, the principle allows for the introduction of what one would call a 'creative-compliance loophole'. In order to identify and use such a loophole, a legal eye needs to look no further than to a *literal* interpretation of the text. Such an interpretation is one of the primary rules of interpretation in law, in particular when the interpretation enabled by the *spirit* of the set rule may not bring the desired outcome. Since they fail to acknowledge the distinction between what robots *can be* and what robots *ought to be*, the explanations given to this principle in the 2010 original document do not seem to really address this rather basic approach to interpreting rules and its potential consequences. This, unfortunately, is bound to render the principle ineffective as a normative rule destined to guide conduct.

The commentary published at the time in connection to this principle points to other potential pitfalls for normative reasoning. As mentioned with respect to 'multi-use tools', there is, for instance, an attempt to put forward the idea that robots are tools like any others. In order to pursue this logic, equivalences are sought at any cost, by comparing a robot with a knife or a gun used for different, both relatively benign and criminal purposes. However, all indications are that robots *are not* tools like any other and this is precisely why our conceptual frameworks about agency, intentionality, responsibility, humanity even, are challenged and, ultimately, this is also the source for our social unease about robots.

Moreover, very dangerously, the principle accepts the use of robots for military purposes as an inevitability. As a public relations exercise, this is not meant to be very reassuring. Allowing robots to be developed and used '*in the interest of national security'* appears as an indirect acknowledgement that regulating the use of robots created for violence against humans is very difficult, even impossible, to achieve. One can only imagine the consequence of such loophole-ridden reasoning in a normative context such as, for instance, the one supported by the First Amendment in the US or on the wider international arena. While one can acknowledge the drafters' decision to remain realistic about governments' decisions on the use of robots, especially in the context in which robots are already being used by the military, it would have been highly desirable for the draft to display more courage in addressing the potential dangers brought about by advanced highly autonomous robots used for military purposes.

However, a clearer definition, addressing the spectrum of AI-endowed robots, would have been more adequate in order to highlight the extreme potential for harm that such robots may bring, and this might have been instrumental in emphasising their use as weapons as being highly undesirable, based on the precautionary principle (Gevel & Noussair, 2013, p. 40; Nemetz, 2013, p. 185). The danger of using sophisticated robots for violent, aggressive goals is so great (Knight, 2015) that, like the use of nuclear technology for weaponry, or the use of chemical and biological weapons, it should be strongly discouraged as a matter of principle. Even if this use were not to be totally forbidden, a clearer definition of what machines one is talking about would, again, be very helpful. The 'national security' argument, for instance, should be more precisely and narrowly defined, as should the types of 'harm' that one can allow/design robots to do. In this context, implying the need for strict international regulation would have been equally desirable.

From the second principle (PR2) we learn, quite appropriately, that robots are not responsible agents: *Humans not robots are responsible agents. Robots should be designed; operated as far as is practicable to comply with existing laws and fundamental rights and freedoms, including privacy.* In as much as this principle corresponds *for now* to the existing technological AI achievements, the absence of a clear working definition that would reach further into the not so distant future - as a number of AI writers purport (Adam, 2008; Ashrafian, 2015) - leaves this second principle ill-adapted to attend to the expected speed of technological progress, and even more so to a singularity threshold that would have AI surpassing the human intellect and potentially be assigned increasing autonomy.

'Advanced learning robots', 'intelligent machines', AGIs, etc., are all concerned with these principles and their parameters. Their 'mechanics', however, may well be more complex than the law or the normative discourse can handle in the absence of clear definitions. Robots and AI machines may well learn to deal with 'exceptions' before the law learns to deal with 'differences'.[5] Equally, other disciplines seem to indicate that numbers (in this particular case, 'programming') may well be more than just that, numbers being inalienably paired with a *narrative* propensity that, one could say, may well be construed differently by the machine than by the human, yet may still be *construed by it* (du Sautoy, 2015).[6] In this rather complex context, two aspects are certain: firstly, the AI scientists are working hard at blurring the distinctions between coding and consciousness and, secondly, the law has a long tradition of working with both narratives and fictions and of ascribing responsibility to both acquired and assigned (moral) agency. An improved second principle should acknowledge both of these dimensions.

An additional element refers to the comments accompanying the PR2 that seem to add more confusion than clarity. First of all, the comments voice the presumption that 'no one is likely deliberately to set out to build a robot which breaks the law', the only reason why breaking the law might happen being the ignorance of the law due to the lack of complex legal expertise. This presumption, of course, has no foundation in the real world of deviance and defiance, as revealed by socio-legal studies in the population at large as well as among the white collars (Fisher, 2015; Mathews & Wacker, 2010). Secondly, and more importantly, the way responsibility is assigned through these comments appears to ignore both the way law *thinks* (Teubner, 1989) about social phenomena as well as the way robots may fail to 'achieve goals and desires that humans specify', and thus may contribute to bringing about novel types of transgression (Bostrom, 2012).

An equal need for a definitional framework stems from PR3: *Robots are products. They should be designed using processes which assure their safety and security.* This principle raises issues of legal distinctions with respect to self-defence; tort issues, assimilation of responsibility, to name just a few, issues that depend very much on the context and on the extent to which, as mentioned above, the regulator will be willing to construe equivalences between robots and other types of tools or between robots and other types of property items (Balkin, 2015).

PR4 endeavours to protect humans from being sold or presented with counterfeited emotions and interactions: *Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent*. Again, this principle raises several issues. The ability of people to communicate with sociable machines is one of the avenues pursued intensely in the AI scientific community (Breazeal, 2003, p. 129). While the aim is not one of deceiving, creating the context for this type of communication to take place will raise many regulatory problems. The borderline between a crafty illusion and a deceptive interface can easily be filled by the human emotions. Part of what we are after in creating 'service robots' that will read bedtime stories to our children or care for our elderly disabled father, is emotional engagement. The scientific evidence appears to show that creating the conditions for emotional engagement may well transcend the 'behind the veil' appearance, even when that appearance remains, in fact, the one of a 'tin' (Dautenhahn, 2007; Goodrich & Schultz, 2007). From here to the marketing manipulation of the emotional engagement there remain just a few steps that the regulatory environment will find challenging to accompany. The comments published with this principle rightly welcome further reflection.

Finally, PR5 addresses the thorny issue of actual legal responsibility for harm or damage produced by a robot: *The person with legal responsibility for a robot should be attributed.* This principle appears clear enough in the context in which we are a long way away from AI capable of technically fulfilling the requirements for such responsibility. At the same time, taking into account the complex components and fields of knowledge that contribute to the make-up of a robot - as indicated above when discussing the definition - some challenges to the regulatory approach will persist, such as the ones related to prising apart the various degrees of responsibility when things go wrong. Having a 'registered keeper', bearer of responsibility, is only part of the solution and does not address the problem of 'many hands', for instance. The responsibility-bearing entity will require further reflection, as will the type of harm(s) that may be attributed to such entities, taking into account, as some suggest, that lately, for the first time, 'the promiscuity of data' can be shown to be combined with 'the capacity of doing physical harm' (Calo, 2015).

4. Conclusions

Insofar as AI developments are concerned, it is quite likely that in a context in which market powers and scientific capability deregulation share substantial synergies, we will most surely, in the words of Roy Bradbury, first 'jump off the cliff and build our wings on the way down'. In such circumstances, what one can do is to make sure of being relatively prepared for this jump by already interrogating and problematizing our relationship with science and technology, and by intensifying our reflection on living with robots. Addressing 'tools', 'products', 'artefacts' and 'agents' with respect to their normative implications, one may want to pay heed to St Augustine's reflection on the intricate connection between language and interpretation as a path to revealing a deeper, existential level of *self-understanding*. The way we think normatively about human-robot *interaction*[7] will say as much about the robot as about the human. To borrow from Gianbattista Vico's 1725 *New Science* (Vico, 1744), we need to bear in mind that our thinking about robots is rooted in a given cultural context. This means that, in reflecting about the normative parameters of leaving in a world populated with robots as much as with humans, the social scientist will not deal with a field of idealised and putatively 'subject-independent objects', but will investigate a world that is, fundamentally, her own. The process of regulating robots is, therefore, a process of self-understanding, a process rooted in a given historical context and practice. Such an understanding does not automatically culminate in neat normative legal propositions as the five principles may suggest. This being said, we acknowledge the contribution that the EPSRC Principles of Robotics bring to the debate, as an initiative that is to be understood as a heroic, normative 'jumping off the cliff' moment. It remains, of

course, for the initiative to build up some strong regulatory wings. As indicated in these reflections, part of the required strength is to be provided through a conceptual framework and definitions that bring clarity and flexibility.

References

Adam, A. (2008). Ethics for Things. *Ethics and Information Technology*, *10*(2–3), 149–154.

Ashrafian, H. (2015). Artificial Intelligence and Robot Responsibilities: Innovating Beyond Rights. *Science and Engineering Ethics*, *21*(2), 317–326.

Balkin, J. M. (2015). The Path of Robotics Law. *California Law Review Circuit*, *6*, 45–60.

Barraza, E. (2016, January 23). How Amnesty International Aims to Bring Technology and Human Rights Together. *SingularityHUB*. Retrieved from http://singularityhub.com/2016/01/23/how-amnesty-international-aims-to-bring-technology-and-human-rights-together/

Bostrom, N. (2012). The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents. *Minds and Machines*, *22*(2), 71–85.

Breazeal, C. (2003). Emotion and Sociable Humanoid Robots. *International Journal of Human-Computer Studies*, *59*(1–2), 119–155.

Brewster, S. (2014, July 5). What Is a Robot? The Answer Is Constantly Evolving. *GIGAOM*. Retrieved from https://gigaom.com/2014/07/05/what-is-a-robot-the-answer-is-constantly-evolving/

Bryson, J. J. (2012). The Making of the EPSRC Principles of Robotics. *The AISB Quarterly*, *Spring 2012*(133), 14–15.

Calo, R. (2015). Robotics and the Lessons of Cyberlaw. *California Law Review*, *103*, 513.

Dautenhahn, K. (2007). Socially Intelligent Robots: Dimensions of Human–Robot Interaction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1480), 679–704.

Deleuze, G. (2006). *The Fold* (New Edition). London ; New York: Continuum.

Dominey, P. F. (2016, May). *Neural Basis of Semantic Representations: Grounding Meaning in Sensory-Motor Experience and Narrative with the iCub (Robot Cognition Lab, Stem Cell & Brain Research Institute, Lyon, France)*. Presented at the AISB Annual Convention, Sheffield, UK. Retrieved from http://webcache.googleusercontent.com/search?q=cache:aOSXPCuXv7YJ:neuroinformatics.usc.edu/wp-content/uploads/2014/02/Dominey_Peter.pdf+&cd=3&hl=en&ct=clnk&gl=fr&client=firefox-b

du Sautoy, M. (2015). *Narrative and Proof: Two Sides of the Same Equation?* Oxford: TORCH, The Oxford Research Centre in the Humanities. Retrieved from http://www.torch.ox.ac.uk/narrative-and-proof-two-sides-same-equation-0

EPSRC. (2010, September). Principles of robotics: Regulating Robots in the Real World. Retrieved June 19, 2016, from https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/

Fagundes, D. (2001). Note, What We Talk About When We Talk About Persons: The Language of a Legal Fiction. *Harvard Law Review*, *114*(6), 1745–1768.

Fisher, K. (2015). The Psychology of Fraud: What Motivates Fraudsters to Commit Crime? Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2596825

Gevel, A. J. W. van de, & Noussair, C. N. (2013). *The Nexus between Artificial Intelligence and Economics*. Springer Science & Business Media.

Goodrich, M. A., & Schultz, A. C. (2007). Human-Robot Interaction: A Survey. *Foundations and Trends in Human-Computer Interaction*, *1*(3), 203–275.

Grady, K. (2014, December 31). Artificial Intelligence: Be Afraid, Be Very, Very Afraid (Or Not). *SeytLines: Changing the Practice of Law*. Retrieved from http://www.seytlines.com/2014/12/artificial-intelligence-be-afraid-be-very-very-afraid-or-not/

Kahn, P. H., Ishiguro, H., Kanda, T., Freier, N. G., Severson, R. L., & Miller, J. (2007). What Is a Human? Toward Psychological Benchmarks in the Field of Human–Robot Interaction. *Interaction Studies*, *8*(3), 363–390.

Knight, W. (2015, August 3). Military Robots: Armed, but How Dangerous? *MIT Technology Review*. Retrieved from https://www.technologyreview.com/s/539876/military-robots-armed-but-how-dangerous/

Latour, B. (2002). Body, Cyborgs and the Politcs of Incarnation. In S. Sweeney & I. Hodder (Eds.), *The Body: Darwin College Lectures* (pp. 127–141).

Latour, B. (2004). How to Talk About the Body? the Normative Dimension of Science Studies. *Body & Society*, *10*(2–3), 205–229.

Mathews, R., & Wacker, W. (2010). *The Deviant's Advantage: How Fringe Ideas Create Mass Markets*. New York: Random House.

May, L. (1996). *Sharing Responsibility* (New edition 1996). Chicago: University Of Chicago Press.

Merriam-Webster Dictionary. (2016). Retrieved February 10, 2016, from http://www.merriam-webster.com/dictionary/robot

Merzouki, R., Samantaray, A. K., Pathak, P. M., & Ould Bouamama, B. (2013). *Intelligent Mechatronic Systems*. London: Springer London.

NASA. (2015, May 18). What Is Robotics? Retrieved February 10, 2016, from http://www.nasa.gov/audience/forstudents/k-4/stories/nasa-knows/what_is_robotics_k4.html

Nemetz, P. N. (2013). *Business and the Sustainability Challenge: An Integrated Perspective*. London and New York: Routledge.

Norrie, A. (1991). A Critique of Criminal Causation. *Modern Law Review*, *54*, 685–701.

Roser, M. (2016). Technological Progress. Published online at OurWorldInData.com. Institute for New Economic Thinking at the Oxford Martin School & The Nuffield Foundation. Retrieved from https://ourworldindata.org/technological-progress/

Teubner, G. (1989). How the Law Thinks: Toward a Constructivist Epistemology of Law. *Law and Society Review*, *23*(5), 727–757.

Vico, G. (1744). *The New Science of Giambattista Vico*. Cornell University Press.

Voiculescu, A. (2016). *Human Rights Beyond the Human: Hermeneutics and Normativity in the Age of the Unknown* (SSRN Scholarly Paper No. ID 2798430) (pp. 1–24). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=2798430

Voiculescu, A. (2017). Business Responsibility for Human Rights Violations from a Theoretical Perspective: Towards a Moral Division of Labour. In J. D. Rendtorff (Ed.), *Perspectives on Philosophy of Management and Business: Including a Special Section on Business and Human Rights* (Vol. 51, pp. 1–22). London and New York: Springer.

Williams, B. (2005). Cognitive Robotics: Syllabus. Aeronautics and Astronautics | MIT OpenCourseWare. Retrieved from http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-412j-cognitive-robotics-spring-2005/syllabus/

Wilson, H. J. (2015, April 15). What Is a Robot, Anyway? *Harvard Business Review*. Retrieved from https://hbr.org/2015/04/what-is-a-robot-anyway

---

[1] The author would like to thank the two anonymous reviewers for their comments, the Organising Committee of the AISB 2016 Symposium on the Principles of Robotics for the invitation to comment on the EPSRC initiative and the Symposium participants for their comments and for sharing their thoughts and ideas. Any remaining errors are entirely the author's responsibility.

[2] The resulting EPSRC document was in fact made up of five Principles of Robotics, each of them with a 'legal' as well as a 'general audience' version, of seven 'High-Level Messages', as well as of commentaries to each of the five Principles.

[3] The examples of hope and confidence in the cognitive potential of robots are more than numerous. Illustrative, for instance is the opening sentence of the MIT Cognitive Robotics Course Syllabus: 'Modern and future robots will have enough computational horse power to be able to reason continuously and deeply about themselves and their environments' (Williams, 2005).

[4] See interview with Andra Keay, founder of Robot Launchpad and Managing Director of Silicon Valley Robotics in (Brewster, 2014).

[5] This is built on the ideas of identity and difference, from Leibniz to Kant (Deleuze, 2006, p. 95)

[6] For an interesting quest of AI-created narratives, see (Dominey, 2016)

[7] This interaction can be understood on Latour's spectrum of facts and agency (Latour, 2002, 2004)