Editorial: Ethical Principles of Robotics

Tony Prescott and Michael Szollosy

This Connection Science special issue, published in two parts as volume 29/2 and 29/3, addresses ethical and societal issues in robotics. In this editorial we explain the background to the special issue and consider its position within the on-going and increasingly high profile debate about the future impacts of advanced robotics and artificial intelligence (AI) and the ethical obligations of the community involved in robotics research and development.

Background

In 2010 the UK's *Engineering and Physical Science* and *Arts and Humanities Research Councils* (EPSRC and AHRC) organised a retreat to consider ethical issues in robotics to which they invited a pool of experts drawn from the worlds of technology, industry, the arts, law and social sciences. This meeting resulted in a set of ethical "Principles of Robotics" (henceforth "the Principles"), that were published online by the EPSRC (Boden *et al*., 2011), that aimed at "regulating robots in the real world", and were stated in the form of five "rules" and seven "high-level messages".

The years since the publication of Principles have seen substantial advances in research in robotics, and in related fields of artificial intelligence and cognitive science, along with broader changes in society resulting from the real-world impact of these technological developments. There is increasing awareness that the "fourth industrial revolution"—brought about through advances in robotics and artificial intelligence— will have very substantive societal and economic impacts (see, e.g. Ford, 2015) and this is reflected in frequent media reports and surveys.

Owing to the increasing interest in, and pertinence of, ethical issues around robotics, *The Society for Study of Artificial Intelligence and Simulation of Behaviour* (AISB), the UK's leading society for research in AI, hosted a workshop on the 4th of April 2016 to re-visit and re-evaluate the Principles. The workshop was hosted at the AISB's annual convention in Sheffield and was chaired by Tony Prescott. The organising committee comprised Alan Winfield, Madeleine de Cock Buning, Joanna Bryson, and Noel Sharkey, two of whom (Winfield and Bryson) had participated in the original EPSRC/AHRC retreat and co-authored the Principles. In advance of the workshop the organisers published an open call for commentaries on the Principles that then provided the starting point for the workshop discussions. Commentators were specifically asked to assess the continued relevance of the Principles according to three criteria:

> *Validity*—are the Principles correct as statements about the nature of robots (for instance that they are tools and products), robot developers, and the relationship between robots and people (for instance that robots should have a transparent design), or are they ontologically flawed, inaccurate, out-dated, or misleading?
>
> *Sufficiency/generality*—are the Principles sufficient and broad enough cover all of the important issues that might arise in the regulation of the robotics in the real-world or are significant concerns overlooked?

Utility—are the Principles of practical use for robot developers, users, or law-makers, in determining strategies for best practice in robotics, or legal standards or frameworks, or are they limited in their use by lack of specificity or through allowing critical exceptions (such as the use of robots as weapons for the purpose of national security)?

The call resulted in fourteen submissions from a diverse range of contributors including experts in robotics and artificial intelligence, law, social science and the humanities; all were deemed to be within scope and accepted for presentation at the meeting. Further details of the workshop and its outcomes are provided in the report (Szollosy, Rapporteur's Report, issue 29/3) that forms the final article of our special issue. As summarised in that report the workshop noted some of valuable qualities the original Principles, however, several contributors also pointed to what they considered to be significant limitations. The workshop then discussed and voted on fourteen specific "amendments, additions, or reflections" on the Principles. Of these, 9 out 14 were adopted by majority vote, six receiving strong support (67% or more in favour), one majority support (53%), with several of the remaining receiving mixed support of between 33% and 47%. An important and unanimous conclusion of the workshop was that "the Principles should be amended through a thorough and inclusive process".

Content and organisation of the special issue

This *Connection Science* special issue brings together all of the commentaries on the Principles that were submitted to the Sheffield workshop. These commentaries were published in draft form in the proceedings of the workshop but have been peer-reviewed, revised and, in some cases, extended for publication here. One additional commentary has been added that was authored after the workshop (N. Sharkey, this issue).

The first part of this special issue (29/2) includes the full text of the Principles (Boden et al., this issue) published for the first time in print form. This is followed by a commentary (Bryson, this issue) by one the authors of the original Principles on their meaning and the context in which they were developed. Subsequent contributions address general issues arising from the Principles particularly with regard to their philosophical and ontological commitments (what is a robot and what is a human?) and to their possible use in law. The final commentary in part one addresses ethical issues concerning the use of robots as weapons (principle 1). The second part of the special issue (29/3) contains contributions that focus on specific issues related to principles 2-5, and concludes with the report of the meeting (Szollosy, issue 29/3).

We hope that these commentaries, together with the published Principles and workshop report, will form an important contribution to the on-going debate around the societal role of robotics and will be useful in framing any future revision of the Principles or other attempts to codify ethical regulation of robotics research and development.

The wider context

Less than a year has passed since our meeting our in Sheffield, and yet a great deal has already happened that is relevant to the debate around robot ethics. Soon after our meeting in 2016 we saw populist victories for Brexit in the British referendum in June and for Donald

Trump in the US elections in November. Both elections, and wider movements in the West as a whole, suggest broad societal concern both around globalisation and the technologies that are driving it—internet, big data, artificial intelligence, automation, and robotics.

Furthermore, it is clear that academic and expert opinion on the important issues of the day is increasingly challenged. In the words of British politician Michael Gove, people have "had enough of experts", whilst in the US voters are offered "alternative facts". Where research becomes the topic of public controversy, for example in case of genetically modified organisms (GMOs), or climate change, scientists (and in the case of robotics, engineers, psychologists, sociologists, etc.) cannot assume that their qualifications and authority will be enough to convince a deeply sceptical, wary public of the inherent goodness of their research. Today, across all areas of science and technology a blogger with a large number of Twitter followers can have greater impact on public opinion than a Nobel Prize winner or a scientist with a high H-factor. The reality is that it is no longer sufficient, if it ever was, to find consensus within a research community, and it has become more important than ever to reach out and engage with wider stakeholders and to understand public concerns in order to advance science and technology in a consensual way (Prescott & Verschure, 2016).

In this context, it is important to recognise that cultural representations of robotics already have an existence that is well beyond the control of the experts that would seek to reassure the public. It is worth recalling that GM foods were frequently portrayed in the popular media as "Frankenstein foods", and yet GMOs have had nowhere near the same coverage as robots and AI in mainstream media such as films, literature, and video games. Robotics, in the popular imagination, is also constructed through narratives that adhere to the Frankenstein archetype, and to its 20th century re-invention as "The Terminator". In terms of attempts to rewrite this narrative we are already playing a game of catch-up.

The period 2015-2016 may be remembered as the time in which the world woke up to imminent and likely impacts of developments in artificial intelligence and robotics. In 2015 noted scientists and entrepreneurs—Elon Musk, Stephen Hawking, Bill Gates—voiced concerns about the future threat of AI, some of an existential nature, and an international group of experts spearheaded by the Institute for the Future of Humanity in Oxford, UK publish an open letter stating their commitment to develop AI for the benefit of humankind that to date has been signed by more than 8,000 people (Future of Humanity Institute, 2017). The topic of AI and Robotics has become a regular focus of debate about global risks at meetings of the World Economic Forum (see World Economic Forum, 2015; 2017); national Parliament's, including the UK's *Parliamentary Science and Technology Committee,* have established their own investigations and published preliminary findings (Science and Technology Committee, 2016). The IEEE, the world's largest technical professional organization dedicated to advancing technology, has established a global initiative to develop ethical standards for the design of AI and autonomous systems (IEEE, 2016). The European Union (EU) has also looked to address and advance the debate. In January 2017 the EU parliament published a draft report with recommendations for new rules governing robots and AI (Delvaux 2016). In February, members of the Legal Affairs Committee of the European Parliament passed a resolution accepting the recommendations in the report. This outcome could mean changes in European law that would require a code of conduct for roboticists and engineers, new laws on insurance and corporate governance of robots and AI and, most controversially in the popular media, new categories for robots and AI, one of

which might see autonomous robots defined as 'electronic persons'. This controversy was fore-shadowed in our own workshop where the ontological status of robots was hotly debated—whether, and it what circumstances, robots should be considered to be more than simply machines or tools.

We are heartened to see these issues being discussed and taken seriously by national governments and by international organisations and federations. The heightened level of interest demonstrates the prescience of the organisers of 2010 meeting, and the importance to revisit and build on their foundations. Research in robotics is global, and is conducted by many different types of organisations, including traditional bodies, such as universities and corporations, as well as newer and more informal players, such as online and open-source "maker" communities. Whilst the recent European initiative is an important milestone, governance and regulation will need to be international if it is to be effective and not simply promote competitive advantages for less regulated countries. It is our view that the research community in robotics and artificial intelligence can and should provide global leadership in this domain, attending to and engaging with critical voices, and helping to shape regulatory frameworks as they emerge. We present this special issue as a contribution towards this challenge.

Acknowledgements
The organisers of the AISB 2016 Workshop are grateful to the UK RAS Network for sponsoring some of the workshop costs, and to the EPSRC for allowing us to republish "the Principles" in this theme issue.

References

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., ... & Sorell, T. (2011). Principles of robotics. *The United Kingdom's Engineering and Physical Sciences Research Council (EPSRC). web publication*.

Delvaux, Mady. (2016). Draft Report with recommendations to the Commission on Civil Law Rules on Robotics. http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML%2BCOMPARL%2BPE-582.443%2B01%2BDOC%2BPDF%2BV0//EN. Accessed 24th March 2017.

Ford, Martin (2015) Rise of the Robots: Technology and the Threat of Mass Unemployment. London: Oneworld publications.

Future of Humanities Institute (2017). An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence. https://futureoflife.org/ai-open-letter/. Accessed 24th March 2017.

IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (2016). *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing*

with Artificial Intelligence and Autonomous Systems (AI/AS). http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html. Accessed 24th March 2017.

Prescott, T. J., & Verschure, P. F. M. J. (2016). Action-oriented cognition and its implications: Contextualising the new science of mind. In A. K. Engel, K. Friston, & D. Kragic (Eds.), Where's the Action? The Pragmatic Turn in Cognitive Science (pp. 321-331). Cambridge, MA: MIT Press for the Ernst Strüngmann Foundation.

Science and Technology Committee (2016). Report on Robotics and Artificial Intelligence. October 2016. https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/14502.htm. Accessed 24th March 2017.

World Economic Forum (2015). Deep shift technology tipping points and societal impact. September 2015. https://www.weforum.org/reports/deep-shift-technology-tipping-points-and-societal-impact. Accessed 24th March 2017.

World Economic Forum (2017). The Global Risks Report 2017. January 2017. http://www3.weforum.org/docs/GRR17_Report_web.pdf. Accessed 24th March 2017.