CONNECTION SCIENCE, 2017
VOL. 29, NO. 2, 124–129
http://dx.doi.org/10.1080/09540091.2016.1271400

Principles of robotics: regulating robots in the real world

Margaret Boden[a], Joanna Bryson[b], Darwin Caldwell[c], Kerstin Dautenhahn[d], Lilian Edwards[e], Sarah Kember[f], Paul Newman[g], Vivienne Parry[h], Geoff Pegman[i], Tom Rodden[j], Tom Sorrell[k]*, Mick Wallis[l], Blay Whitby[a] and Alan Winfield[m]

[a]Department of Informatics, University of Sussex, Brighton, UK; [b]Department of Computer Science, University of Bath, Bath, UK; [c]Italian Institute of Technology, Genova, Italy; [d]School of Computer Science, University of Hertfordshire, Herts, UK; [e]Law School, University of Strathclyde, Glasgow, UK; [f]Department of Media and Communications, Goldsmiths, University of London, London, UK; [g]Department of Engineering Science, University of Oxford, Oxford, UK; [h]Independent Scholar; [i]RU Robotics Ltd, Manchester, UK; [j]School of Computer Science, University of Nottingham, Nottingham, UK; [k]Department of Philosophy, University of Birmingham, Birmingham, UK; [l]School of Performance and Cultural Industries, University of Leeds, Leeds, UK; [m]Bristol Robotics Laboratory, University of the West of England, Bristol, UK

ABSTRACT

This paper proposes a set of five ethical principles, together with seven high-level messages, as a basis for responsible robotics. The Principles of Robotics were drafted in 2010 and published online in 2011. Since then the principles have influenced, and continue to influence, a number of initiatives in robot ethics but have not, to date, been formally published. This paper remedies that omission.

ARTICLE HISTORY
Received 17 August 2016
Accepted 9 October 2016

KEYWORDS
Robot ethics; principles of robotics; responsible innovation

1. Introduction

In September 2010, a group drawn from the worlds of technology, industry, the arts, law and social sciences met at the joint EPSRC and AHRC Robotics Retreat to discuss robotics, its applications in the real world and the huge promise robotics offers to society. Robots have left the research lab and are now in use all over the world, in homes and in industry. We expect robots in the short, medium and long term to impact our lives at home, our experience in institutions, our national and our global economy, and possibly our global security. However, the realities of robotics are still relatively little known to the public where science fiction and media images of robots have dominated. One of the aims of the meeting was to explore what steps should be taken to ensure that robotics research engages with the public to ensure this technology is integrated into our society to the maximum benefit of all of its citizens. As with all technological innovation, we need to try to ensure that robots are introduced from the beginning in a way that is likely to engage public trust and confidence; maximise the gains for the public and commerce; and proactively head off any potential unintended consequences.

Given their prominence it is impossible to address the governance of robotics without considering Asimov's famous three laws of robotics (Asimov, 1950). (Asimov's laws state

CONTACT Alan Winfield alan.winfield@uwe.ac.uk
*Present address: Department of Politics and International Studies, University of Warwick, Coventry, UK

that 1 – a robot may not injure a human being or, through inaction, allow a human being to come to harm; 2 – a robot must obey the orders given it by human beings except where such orders would conflict with the first law, and 3 – a robot must protect its own existence as long as such protection does not conflict with the first or second laws.)

Although they provide a useful departure point for discussion Asimov's rules are fictional devices. They were not written to be used in real life and it would not be practical to do so, not least because they simply don't work in practice. (For example, how can a robot know all the possible ways a human might come to harm? How can a robot understand and obey all human orders, when even people get confused about what instructions mean?) Asimov's stories also showed that even in a world of intelligent robots, his laws could always be evaded and loopholes found. But finally, and most importantly, Asimov's laws are inappropriate because they try to insist that robots behave in certain ways, as if they were people, when in real life it is the humans who design and use robots who must be the actual subjects of any law.

As we consider the ethical implications of having robots in our society, it becomes obvious that robots themselves are not where responsibility lies. Robots are simply tools of various kinds, albeit very special tools, and the responsibility of making sure they behave well must always lie with human beings. Accordingly, rules for real robots in real life, must be transformed into rules advising those who design, sell and use robots about how they should act. The meeting delegates devised such a set of "rules" with the aim of provoking a wider, more open discussion of the issues. They highlight the general principles of concern expressed by the group with the intent that they could inform designers and users of robots in specific situations. These new rules for robotics (not robots) are outlined below. The five ethical rules for robotics are intended as a living document. They are not intended as hard-and-fast laws, but rather to inform debate and for future reference. Obviously a great deal of thinking has been done around these issues and this document does not seek to undermine any of that work but to serve as a focal point for useful discussion.

2. Principles for designers, builders and users of robots

The five rules are presented in a semi-legal version together with a looser, but easier to express, version that captures the sense for a non-specialist audience. Each rule is followed by a commentary of the issues being addressed and why the rule is important.

Rule	Semi-legal	General Audience
1	Robots are multi-use tools. Robots should not be designed solely or primarily to kill or harm humans, except in the interests of national security	Robots should not be designed as weapons, except for national security reasons

Commentary. Tools have more than one use. We allow guns to be designed which farmers use to kill pests and vermin, but killing human beings with them (outside warfare) is clearly wrong. Knives can be used to spread butter or to stab people. In most societies, neither guns nor knives are banned but controls may be imposed if necessary (e.g. gun laws) to secure public safety. Robots also have multiple uses. Although a creative end-user could probably use any robot for violent ends, just as with a blunt instrument, we are saying that robots should never be *designed* solely or even principally, to be used as weapons with

deadly or other offensive capability. This rule, if adopted, limits the commercial capacities of robots, but we view it as an essential principle for their acceptance as safe in civil society.

Rule	Semi-legal	General Audience
2	Humans, not robots, are responsible agents. Robots should be designed; operated as far as is practicable to comply with existing laws, fundamental rights & freedoms, including privacy	Robots should be designed and operated to comply with existing law, including privacy

Commentary. We *can* make sure that robot actions are designed to obey the *laws* humans have made.

There are two important points here. First, of course no one is likely to deliberately set out to build a robot which breaks the law. But designers are not lawyers and need to be reminded that building robots which do their tasks as well as possible will sometimes need to be balanced against protective laws and accepted human rights standards. Privacy is a particularly difficult issue, which is why it is mentioned. For example, a robot used in the care of a vulnerable individual may well be usefully designed to collect information about that person 24/7 and transmit it to hospitals for medical purposes. But the benefit of this must be balanced against that person's right to privacy and to control their own life e.g. refusing treatment. Data collected should only be kept for a limited time; again the rule puts certain safeguards in place. Robot designers have to think about how rules like these can be respected during the design process (e.g. by providing off-switches).

Secondly, this rule is designed to make it clear that robots are just tools, designed to achieve goals and desires that *humans* specify. Users and owners have responsibilities as well as designers and manufacturers. Sometimes it is up to designers to think ahead because robots may have the ability to learn and adapt their behaviour. But users may also make robots do things their designers did not foresee. Sometimes it is the owner's job to supervise the user (e.g. if a parent bought a robot to play with a child). But if a robot's actions do turn out to break the law, it will always be the responsibility, legal and moral, of one or more human beings, not of the robot (we consider how to find out who is responsible in rule 5, below).

Rule	Semi-legal	General Audience
3	Robots are products. They should be designed using processes which assure their safety and security	Robots are products: as with other products, they should be designed to be safe and secure

Commentary. Robots are simply not people. They are pieces of technology their owners may certainly want to protect (just as we have alarms for our houses and cars, and security guards for our factories), but we will always value human safety over that of machines. Our principal aim here was to make sure that the safety and security of robots in society would be assured so that people can trust and have confidence in them.

This is not a new problem in technology. We already have rules and processes that guarantee that, e.g. household appliances and children's toys are safe to buy and use. There are well worked out existing consumer safety regimes to assure this: e.g. industry kite-marks, British and international standards, testing methodologies for software to make sure the bugs are out, etc. We are also aware that the public knows that software and computers can be "hacked" by outsiders, and processes also need to be developed to show that robots are secure as far as possible from such attacks. We think that such rules, standards and tests

should be publicly adopted or developed for the robotics industry as soon as possible to assure the public that every safeguard has been taken before a robot is ever released to market. Such a process will also clarify for industry exactly what they have to do.

This still leaves a debate open about how far those who own or operate robots should be allowed to protect them from e.g. theft or vandalism, say by built-in taser shocks. The group chose to delete a phrase that had ensured the right of manufacturers or owners to build "self-defence" capabilities into a robot. In other words we do not think a robot should ever be "armed" to protect itself. This actually goes further than existing law, where the general question would be whether the owner of the appliance had committed a criminal act like assault without reasonable excuse.

Rule	Semi-legal	General Audience
4	Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent	Robots are manufactured artefacts: the illusion of emotions and intent should not be used to exploit vulnerable users

Commentary. One of the great promises of robotics is that robot toys may give pleasure, comfort and even a form of companionship to people who are not able to care for pets, whether due to restrictions in their homes, physical capacity, time or money. However, once a user becomes attached to such a toy, it would be possible for manufacturers to claim the robot has needs or desires that could unfairly cost the owners or their families more money. The legal version of this rule was designed to say that although it is permissible and even sometimes desirable for a robot to sometimes give the impression of real intelligence, anyone who owns or interacts with a robot should be able to find out what it really is and perhaps what it was really manufactured to do. Robot intelligence is artificial, and we thought that the best way to protect consumers was to remind them of that by guaranteeing a way for them to "lift the curtain" (to use the metaphor from The Wizard of Oz).

This was the most difficult rule to express clearly and we spent a great deal of time debating the phrasing used. Achieving it in practice will need still more thought. Should all robots have visible bar-codes or similar? Should the user or owner (e.g. a parent who buys a robot for a child) always be able to look up a database or register where the robot's functionality is specified? See also rule 5 below.

Rule	Semi-legal	General Audience
5	The person with legal responsibility for a robot should be attributed	It should be possible to find out who is responsible for any robot

Commentary. In this rule we try to provide a practical framework for what all the rules above already implicitly depend on: a robot is never legally responsible for anything. It is a tool. If it malfunctions and causes damage, a human will be to blame. Finding out who the responsible person is may not however be easy. In the UK, a register of who is responsible for a car (the "registered keeper") is held by DVLA; by contrast no one needs to register as the official owner of a dog or cat. We felt the first model was more appropriate for robots, as there will be an interest not just to stop a robot whose actions are causing harm, but people affected may also wish to seek financial compensation from the person responsible.

Responsibility might be practically addressed in a number of ways. For example, one way forward would be a licence and register (just as there is for cars) that records who is responsible for any robot. This might apply to all or only operate where that ownership is not obvious (e.g. for a robot that might roam outside a house or operate in a public institution such as a school or hospital). Alternately, every robot could be released with a searchable online licence which records the name of the designer/manufacturer and the responsible human who acquired it (such a licence could also specify the details we talked about in rule 4 above). There is clearly more debate and consultation required.

Importantly, it should still remain possible for legal liability to be shared or transferred e.g. both designer and user might share fault where a robot malfunctions during use due to a mixture of design problems and user modifications. In such circumstances, legal rules already exist to allocate liability (although we might wish to clarify these, or require insurance). But a register would always allow an aggrieved person a place to start, by finding out who was, on first principles, responsible for the robot in question.

3. Seven high-level messages

In addition to the above principles, the group also developed an overarching set of messages designed to encourage responsibility within the robotics research and industrial community, and thereby gain trust in the work it does. The spirit of responsible innovation is, for the most part, already out there but we felt it worthwhile to make this explicit. The following table sets out the messages alongside explanatory commentaries.

	Message	Commentary
1	We believe robots have the potential to provide immense positive impact to society. We want to encourage responsible robot research	This was originally the "0th" rule, which we came up with midway through. But we want to emphasise that the entire point of this exercise is positive, though some of the rules above can be seen as negative, restricting or even fear-mongering. We think fear-mongering has already happened, and further that there are legitimate concerns about the use of robots. We think the work here is the best way to ensure the potential of robotics for all is realised while avoiding the pitfalls
2	Bad practice hurts us all	It's easy to overlook the work of people who seem determined to be extremist or irresponsible, but doing this could easily put us in the position that GM scientists are in now, where nothing they say in the press has any consequence. We need to engage with the public and take responsibility for our public image
3	Addressing obvious public concerns will help us all make progress	The previous note applies also to concerns raised by the general public and science fiction writers, not only our colleagues
4	It is important to demonstrate that we, as roboticists, are committed to the best possible standards of practice	As above
5	To understand the context and consequences of our research, we should work with experts from other disciplines including social sciences, law, philosophy and the arts	We should understand how others perceive our work, and what the legal and social consequences of our work may be. We must figure out how to best integrate our robots into the social, legal and cultural framework of our society. We need to figure out how to engage in conversation about the real abilities of our research with people from a variety of cultural backgrounds who will be looking at our work with a wide range of assumptions, myths and narratives behind them

	Message	Commentary
6	We should consider the ethics of transparency: are there limits to what should be openly available	This point was illustrated by an interesting discussion about open-source software and operating systems in the context where the systems that can exploit this software have the additional capacities that robots have. What do you get when you give "script kiddies" robots? We were all very much in favour of the open-source movement, but we think we should get help thinking about this particular issue and the broader issues around open science generally
7	When we see erroneous accounts in the press, we commit to take the time to contact the reporting journalists	Many people are frustrated when they see outrageous claims in the press. But in fact science reporters do not really want to be made fools of, and in general such claims can be corrected and sources discredited by a quiet and simple word to the reporters on the byline. A campaign like this was already run successfully once in the late 1990s

4. Afterword

The introduction, principles and high-level messages in Sections 1–3, are presented as originally published in 2011 (Boden et al., 2011), with only minor editorial corrections for both grammar and consistency. The purpose of this paper is not to revise or extend the principles and messages, which remain here unchanged.

Since publication online in 2011, the principles of robotics have been disseminated in various ways and media, including in New Scientist (Winfield, 2011). Subsequently the principles have been cited in Wikipedia (2016), and in an influential paper in AI Magazine setting out research priorities for 'Robust and Beneficial Artificial Intelligence' (Russell, Dewey, and Tegmark, 2015). They are also incorporated into British Standard BS 8611, Guide to the ethical design and application of robots and robotic systems (BS 8611:2016, 2016).

Acknowledgements

The workshop which drafted the principles was expertly chaired by Vivienne Parry. We are also grateful to Ian Baldwin, Ann Grand and Paul O'Dowd who facilitated the discussion, and research council leads Shearer West (AHRC) and Stephen Kemp (EPSRC). The principles of robotics are reproduced here with the kind permission of the Engineering and Physical Sciences Research Council (EPSRC).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Asimov I. (1950). *I, Robot*. New York: Gnome Press.

Boden M., Bryson J., Caldwell D., Dautenhahn K., Edwards L., Kember S., ... Winfield A. (2011). *Principles of robotics*. Swindon, UK: Engineering and Physical Sciences Research Council. Retrieved from www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/. Accessed 16 August 2016.

BS 8611:2016 (2016). *Robots and robotic devices: Guide to the ethical design and application of robots and robotic systems*. London: British Standards Institution.

Russell S., Dewey D., & Tegmark M. (2015). Research priorities for robust and beneficial artificial intelligence. *AI Magazine*, *36*(4), 105–114. Association for the Advancement of Artificial Intelligence.

Wikipedia. (2016). *Three laws of robotics*. Retrieved from en.wikipedia.org/wiki/Three_Laws_of_Robotics. Accessed 16 August 2016.

Winfield A. (2011). Roboethics – for humans. *New Scientist*, *210*(2811), 32–33. 7 May 2011.