

For the ‘Ethical Principles of Robotics’ special issue

EPSRC Principles of Robotics: Commentary on safety, robots as products, and responsibility

Paula Boddington

Department of Computer Science, University of Oxford, Oxford, UK

Wolfson Building, Parks Road, Oxford, OX1 3QD

paula.boddington@cs.ox.ac.uk

Acknowledgements: This work is supported by the Future of Life Institute.

ESPCR Principles of Robotics: Commentary on safety, robots as products, and responsibility

The EPSRC Principles of Robotics refer to safety. How safety is understood is relative to how tasks are characterised and identified. But the exact task(s) a robot plays within a complex system of agency may be hard to identify. If robots are seen as products, it is nonetheless vital that the safety and other implications of their use in situ must also be considered carefully, and they must be fit for purpose. The Principles identifies humans as responsible, rather than robots. We must thus understand how the replacement of human agency by robotic agency may impact upon attributions of responsibility. The Principles seek to fit into existing systems of law and ethics. But these may need development, and in certain context, attention to more local regulations is also needed. A distinction between ethical issues related to the design of robotics, and to their use, may be needed in the Principles.

Keywords: robot ethics; principles of robotics; safety; responsibility

1. Introduction

This commentary focuses on rules 2 and 3 of the EPSRC Principles of Robotics (Boden et al, 2011), with brief consideration of rule 5. The notion of safety which appears in the Principles needs to be elaborated and specified, in conjunction with the implications of regarding robots as products. In order to clarify issues relating to safety, issues relating to the design and manufacture of robots, and also to their use in situ, will both need to be considered explicitly. In these tasks, we will also need to consider carefully how the Principles treat the concept of responsibility. If robots are not responsible, and humans are, as the Principles state, then the use of robots replaces responsible agents with non-

responsible machines, and the implications of this must be analysed. Responsibility is a multifaceted notion, which operates within a complex and often nested system of agency and accountabilities. In the social world, such systems may often be only partially understood.

2. Safety in the EPSRC Principles of Robotics

I suggest that the notion of ‘safety’ in the Principles needs elaboration and specification. Of course, safety must be an essential consideration of the Principles, but to aspire to ‘safety’ is hardly inspirational. Good design which fulfils important human needs and aspirations should go beyond mere considerations of safety. Stated simply, safety is a minimum ethical goal. Moreover, if robots are products, then for the Principles to specify the goal of safety is redundant, since consumer protection legislation has long required that products are both safe and fit for purpose.

However, if a primary goal of the EPSRC Principles of Robotics is simply to give assurances that robots will not disrupt existing laws, and will not pose added dangers to the public, to have such modest goals may be considered adequate. Indeed, the Principles specifically state that they seek to fit with existing laws and fundamental rights, a point to which we will briefly return later on. However, even if the goal is no more than giving assurances of safety, there are reasons to consider that as they stand, the Principles may overlook ways in which serious safety issues may arise.

Furthermore, it is desirable to specify how safety is understood in the Principles. Does it refer simply to material damage to humans or property? By examining complex settings in which robots may be used, I will argue that safety needs to be understood broadly and in ways which encompass the danger of disruption to important norms relevant to the context of use of robots. That is, safety should include consideration of

how the use of robots might endanger moral values and other important social norms. This will be illustrated through briefly discussing the use of robots within a hospital setting.

3. Robots as products

These questions about the understanding of safety relate to how robots are understood as products in the Principles. Is a product something that a consumer simply takes out of the box and starts using? I argue that more attention needs to be made in the Principles to considering how important ethical issues could arise from the use of robots in an embedded context. Safety issues may arise further down the line when robots are employed within social settings, issues which may be hard to predict in advance. That is, concerns arising from the use of robots in specific contexts may need to be highlighted and distinguished from concerns arising from the design and manufacture of robots. This will then raise questions about lines of responsibility between robotics professionals and others, as shall be seen.

3.1 Safety and task identification

How issues of safety are identified in any particular case is relative to the task that is being undertaken. If safety is concerned with ensuring that important values are not disrupted, whether these are material, economic, social, psychological, aesthetic or moral values, then it's important to know what these values are, on any particular occasion, to identify dangers of disruption to these values. To illustrate: if you are simply told your task is to attend social event outdoors involving a picnic on an especially sunny day, you may be concerned with the dangers of sunburn, and attend wearing swimwear and covered in sunscreen, thinking you'd done a good job on safety. If you didn't realise that the event was also a crucial and highly sensitive diplomatic

event in the White House Rose Garden and televised live worldwide, you might spark a major diplomatic incident by being inadvisably clad.

How is this relevant to the EPSRC Principles of Robotics? It's relevant because of the complexity and intricacy of many important human tasks, especially within complex social systems (I could indeed say simply, 'within social systems', all of which are highly complex.) This complexity may on occasion mean that even those agents who are embedded in such systems don't fully understand what tasks are being accomplished. In some social settings, tasks may be clearly identified and formalised in various ways; in others, important functions may be less visible, and it may be difficult to determine with complete precision how a complex social system is functioning to achieve certain ends. Hence, where robots are deployed in place of humans, or supplementing human agency, identifying what safety issues are at issue may be challenging.

3.2 Robots as products employed in complex settings: downstream safety issues

If a robot is a product, which under current consumer protection law must be 'fit for purpose', then we need to understand very well what that purpose is. A robot might be entirely 'safe' in use, in that it does not 'go wrong' or harm people or damage property in its immediate operation, yet may have large ramifications for the overall safety of the system within which it is employed.

For example, a common and serious problem within hospital settings especially for elderly and vulnerable patients is dehydration. This can have serious health implications leading for instance to confusion. Sometimes dehydration is worsened by difficulty in reaching and managing drinks. Suppose a robotics system designed to assist such patients drink was put in place. This system may work with complete safety and reliability in terms of its immediate use, for

instance, never giving patients too much to drink, never spilling drinks, and never scalding patients.

However, such a system could have potentially large negative consequences in particular contexts. This might be especially the case where robots work in place of humans, and is worst when robots are particularly effective at their tasks. This is an important point to note, since one aim of robotics is to make certain human tasks more efficient. But efficiency in one part of a system may cause problems elsewhere. Increased hydration could cause increased rates of bedwetting in some patients. This could lead to staff fitting catheters, and could lead to hospital acquired incontinence. Patients with hospital acquired incontinence rarely achieve continence again. These patients may thus need specialist accommodation upon discharge, and then may cause ‘bed blocking’. Other patients may fall and receive injuries as they get out of bed to visit the toilet (Booth, Kumlien & Zang, 2009; Oliver, Healey and Haines, 2010).

3.3 Who is responsible for problems when using robots in complex settings?

Identifying such safety issues may be considered to be solely the responsibility of the users of a robot; whether responsibility belongs to the designers, the users, or is jointly held, may depend on the particular case and the particular safety issues involved.

However, robot designers and manufacturers may have a significant role to play in identifying and addressing such ‘further down the line’ problems in the implementation of robotics. The need for careful testing in situ and for dialogue between robot professionals and users over such issues of safety could usefully be addressed explicitly in the Principles. Let us go on to consider responsibility in more detail.

4. Responsibility in the EPSRC Principles of Robotics

Rule 2 of the Principles states that humans are responsible agents, but that robots are not. This is critical in considering the employment of robots, because it implies that

whenever robots are used to replace humans or part of human agency, then the responsibilities attributed formerly to the human agent or human actions will then either be displaced onto another human or humans within the wider system of interactions, or possibly overlooked. The displacement of responsibility produced by substituting a responsible agent (a human) with a non-responsible agent (a robot) may result in shifts in how responsibilities and accountabilities are understood. These shifts may be unexpected and complex.

4.1 Responsibility within complex social systems

Robots will be used within a system of human agents and behaviours. Such systems may be formalised with clearly expressed notions of responsibility and lines of accountability and communication, for example within a hospital setting, where there are not only legal requirements on behaviour, but ethical regulations and various codes of conduct and expectation. Even in such settings, there may be elements which are not fully understood or formalised with complete adequacy. And robots may be used in informal settings for instance, within a home setting of care for an elderly person. In fact, within such informal settings, social research shows that there may be strong local cultures and values regarding lines of responsibility and accountability (Arribas-Ayllon, Featherstone & Atkinson, 2011).

4.2 Responsibility when robots replace humans: the displacement of responsibilities

For an example of how responsibilities and accountabilities may be displaced, consider a robot taking over some of the roles of a health care assistant within the setting of a hospital ward. Responsibilities may be displaced in a variety of ways to different actors within the system of health care management.

These responsibilities may also change how tasks are understood. For instance, tasks which were previously seen as mundane may come to be seen as technical; or tasks previously seen as managerial may come to be seen as technical. Tasks which were previously considered skilled may come to be seen as routine if they can be managed by a robot. What might previously have been thought of as a culpable failure of competence or diligence on the part of a human might come to be seen as difficulty in understanding or operating machinery. There may be wide-ranging repercussions, including for instance for the relative status of jobs within the system (Daykin and Clarke, 2000).

Hence the use of robots within a complex workplace may be very disruptive to working practices. These considerations may go beyond the remit of any Principles of Robotics, but it is nonetheless worthy of consideration to ask where the responsibility of robotics professionals ends and the responsibility of robot end users starts when it comes to analysing, anticipating and addressing such complex and potentially very important issue. Robot professionals could surely have an important role in assisting with analysis of how robots placed within a complex system of human agencies may disrupt that system. I suggest the Principles could usefully indicate this issue.

As an aside, an observation could be made that performing such analyses may effectively bypass debates about whether robots can be developed to the point where responsibility can be attributed to them. For where a robot is part of a complex system of responsibilities and accountabilities, concerns of safety and audit often mean that these are duplicated and backed up in ways which mean that failings within the system can be identified and the load taken up elsewhere. This is certainly an ideal within healthcare systems (Donaldson, 2003).

If robots are not responsible, but humans are, can robots in fact replace humans in all cases? There may be a loss which must be acknowledged and addressed. For instance, within the NHS there are standards of care which aim to deliver person-centred care and to treat patients with dignity. Such statements are central to the provision of good health care (NICE, 2011). This is not simply a matter of delivering health care which has some value norms ‘sprinkled on top’ as an added extra, as it were; treating people well is a well-recognised aid to healing. Working out how the use of robots impacts upon such rich and contextualised values will be very important, and may well be a harder question than simply determining compliance with the law and with fundamental rights.

The EPSRC Principles of Robotics would be advised to add that robots should be designed to comply not just with existing law and fundamental rights, but also with the often nuanced and complex local norms and regulations within the specific systems in which robots are employed.

Furthermore, the development of robots which work well within such rich and nuanced practices as ‘person centred care’ will surely require careful interdisciplinary work and trial and testing in situ.

5. Safety, robot use, and multiplicity of task

We have seen how identifying safety issues requires identifying tasks. Where robots are replacing or extending human agency, this may be in a context where multiple tasks are undertaken within what might otherwise seem a simple transaction. There may be multiple human transactions of significance in a simple task, such as the communication of caring which occurs through routine use of language and body language. Having ‘mundane’ practical tasks taken over by robots might free up human workers to have more time to caring tasks for which humans may be suited; or we might find that robots

can perform some of these tasks as well, or possibly even better than humans. For instance, robot assistance with toileting may help to preserve human privacy and dignity, precisely because of qualities humans have which robots lack. But unless we can identify what these tasks are, disruption to the system may occur. In other words, the use of robots will then present safety issues, if we conceive of safety more broadly and richly simply as concerning injury to person or damage to property.

5.1 Identifying hidden tasks

But these layered tasks may not be transparent, even if they are ultimately of great importance. Working out the effects upon a system when robots enter that system may involve considerable analysis and research. For instance, in the setting of a hospital ward, seemingly ‘mundane’ caring tasks which may be nonetheless extremely important to the health and wellbeing of patients, are often carried out by staff working at lower grades. Hence there is a possibility that all aspects of this work may not be recognised or acknowledged by higher management or by formal audits of the system. Note, that in a healthcare setting, such issues may concern safety even if only conceptualised narrowly as encompassing simply physical harm to humans. With the best will in the world, highly skilled and careful observational work by social scientists may be needed to gain a good understanding of what interactions are actually taking place, and what their implications are, in such a complex setting as a hospital ward.

6. Implications for the responsibilities of robotics professionals

Rule 5 of the Principles of Robotics states that the person with legal responsibility for a robot must be identified. But there is more to the good design and responsible implementation of robots than simply identifying bare legal responsibilities. The complexity of the issues means that identifying just one person as having responsibility

may be unrealistic. The discussion so far indicates that determining issues of safety in use will require careful thought and collaboration between robotics professionals and those involved in the day to day use of robots especially within complex settings. If robots are seen as products, they certainly should not be understood as products in the sense that once you've opened the packet, then 'buyer beware' and it's up to you to use it well. The lines of responsibility between robot professional and robot user need to be considered and clarified.

6.1 Complying with existing laws, or developing laws?

In seeking such clarification, and in analysing exactly how to deal with situations where robot agency replaces human agency, developments in thinking and in response may be required. This may require the existing system of laws to be questioned in various ways. The aim of the Principles of Robotics to comply with existing systems of law and fundamental rights is laudable, but laws and understandings of rights and morality are subject to change and development in response to social and technological changes, as well as to developments in moral and political thought. However, any changes will require careful consideration. The implementation of robotics may be a push towards such changes, and these need not be negative. The very process of looking carefully at how robots and humans interact may teach us valuable lessons. The need to look carefully at lines of responsibility and accountability, and to look carefully at complex issues of safety in robotics, may inform any process of legal and ethical development.

7. Conclusions

Safety in robotics should be seen to encompass not only physical and material safety, but also the avoidance of disruption to psychological, social, moral and other important values.

Safety in robotics should encompass safety issues which may result from downstream effects of the implementation of robots in complex settings as well as from the immediate setting of their use.

If robots are seen as products, then issues arising from the use of these products needs to be considered carefully.

Identifying safety issues in robotics will involve careful analysis of the tasks of robots and humans in complex settings; some important tasks may be hard to identify.

Careful collaborative research between robotics professionals and others involved in the use of robots will be needed to identify issues of safety in use.

Where non-responsible robots replace or supplement responsible agents, attention must be given to analysing how responsibilities become redistributed within networks of agency.

Attention needs to be given to how responsibility for the safe and effective use of robots is shared and differentiated between robotics professionals and others.

The Principles of Robotics should pay attention not simply to existing laws and fundamental rights and freedoms, but also to local and institutional regulations, norms and principles, as relevant.

Existing laws and understandings of ethics are subject to development and improvement. If the Principles of Robotics seeks merely to comply with existing laws and fundamental rights, this may divert attention away from how careful attention to the employment of robotics may assist positively with such development.

Funding: This work was supported by the Future of Life Institute

References

- Arribas-Ayllon, M., Featherstone, K. and Atkinson, P. A. (2011). The practical ethics of genetic responsibility: non-disclosure and the autonomy of affect. *Social Theory & Health* 9(1), pp. 3-23. (10.1057/sth.2009.22)
- Boden M, Bryson J, Caldwell D, Dautenhahn K, Edwards L, Kember S, et al. (2011) Principles of Robotics. Swindon, UK: Engineering and Physical Sciences Research Council ESPRC.
- Booth, J., Kumlien, S. and Zang, Y. (2009). Promoting urinary continence with older people: key issues for nurses. *International journal of older people nursing*, 4(1), pp.63-69
- Daykin, N, Clarke, B. (2000) 'They'll still get the bodily care', *Sociology of Health and Illness*, 22, 3: 349 – 363
- Donaldson L. (2003) Making Amends. Department of Health. London: HMSO 2003.
- Glover J. (1971) Responsibility. Oxford: Blackwell.
- NICE. (2011). Service user experience in adult mental health: improving the experience of care for people using adult NHS mental health services. London: NICE.
- Oliver D, Healey F, Haines T. (2010) Preventing falls and fall-related injuries in hospitals. *Clinical Geriatric Medicine* 26:645–92
- Royal College of Psychiatrists. (2015) Dementia and People with Intellectual Disabilities Guidance on the assessment, diagnosis, interventions and support of people with intellectual disabilities who develop dementia. April 2015