Implementing Responsible Research and Innovation for Care Robots through BS 8611

10

Bernd Carsten Stahl

Abstract

The concept of responsible research and innovation (RRI) has gained prominence in European research. It has been integrated into the EU's Horizon 2020 research framework as well as a number of individual Member States' research strategies. Elsewhere we have discussed how the idea of RRI can be applied to healthcare robots (Stahl and Coeckelbergh 2016) and we have speculated what such an implementation might look like in social reality (Stahl et al. 2014). In this paper I will explore how parallel developments reflect the reasoning in RRI. The focus of the paper will therefore be on the recently published standard on "Robots and robotic devices: Guide to the ethical design and application of robots and robotic systems" (BSI 2016). I will analyse the standard and discuss how it can be applied to care robots. The key question to be discussed is whether and to what degree this can be seen as an implementation of RRI in the area of care robotics.

10.1 Introduction

The concept of responsible research and innovation (RRI) has gained significant currency in recent years. It has been used to discuss how societies can influence their research and innovation activities in a way that will render these more likely to be socially acceptable, desirable and sustainable. The discourse around RRI has covered a broad number of fields of science and technology research. A current high profile area of

B. C. Stahl (✉)

Department of Informatics, Centre for Computing and Social Responsibility,
De Montfort University, Leicester, United Kingdom
E-Mail: bstahl@dmu.ac.uk

concern is the combination of artificial intelligence (AI), big data analytics and robotics which promises imminent disruptive changes. One particular application area in this field is the area of care robotics.

At present much research is undertaken to find ways of implementing and realising the principles of RRI in research and innovation practices. RRI raises some fundamental questions of research governance, such as the intractable problem of how the future can be forecast with sufficient accuracy to allow for meaningful interventions in the present. It also raises numerous practical questions concerning the integration of RRI thoughts into current research and innovation ecosystems. In addition RRI will have to be applied in a way that is sensitive to the local environment and research field, if it is to be successful.

This paper asks the question how RRI can be implemented in research and innovation activities concerning care robots. More specifically, it will investigate a recent development in the area of standardisation, the BS 8611 "Robots and robotic devices: Guide to the ethical design and application of robots and robotic systems" (BSI 2016). The question discussed in this paper is whether and to what degree this new standard can be understood as a way of implementing RRI in care robotics.

In order to answer this question I will first introduce RRI in more detail and discuss the limitations of the current RRI discourse. Following this I will introduce the characteristics and ethical concerns raised by care robots. The subsequent section will outline the BS 8611 standard. Having thus provided the full background, I will discuss how BS 8611 can be applied to care robots and review whether this can be seen as an implementation of RRI. The paper concludes with recommendations for both theory and practice that should help promote RRI in the field of care robots.

10.2 Responsible Research and Innovation

RRI has gained prominence since approximately 2010 as a key term that is used to describe open and participative approaches to research and innovation governance. Von Schomberg's (2011, p. 9) widely cited definition displays many of RRI's key characteristics:

> Responsible Research and Innovation is a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society).

This chapter does not offer the space for a comprehensive review of the quickly growing discourse on RRI. Suffice it to say that it is now the subject of edited books (Hankins 2012; Owen et al. 2013), conferences (Bogner et al. 2015; Hoven et al. 2014) and there is a dedicated journal, the journal of Responsible Innovation.

One of the reasons why RRI has been a highly successful term is that it has been adopted by research funders in attempts to steer research in certain directions. This started out

with the Dutch MVI programme and was taken up by other research funding organisations, notably the European Commission (EC). The EC has adopted RRI as a cross-cutting activity in its Horizon 2020 research framework programme. Given the significant financial value of H2020 of about € 80 billion and the leadership function that this programme has across Europe in inspiring other funders, this has been a key driver for the wider adoption of RRI. It is important to note that the EC has developed a conception of RRI that focuses on six keys: public engagement, gender equality, ethics, science education, open access and governance (European Commission 2013). This European conception of RRI highlights a number of important aspects of RRI but is arguably somewhat narrower than the academic RRI discourse calls for. It is therefore important to note that there are alternative conceptualisations of RRI. For the purposes of this paper, I will use the concept of RRI as put developed by Stilgoe et al. (2013) and subsequently adopted by the UK Engineering and Physical Sciences Research Council (Owen 2014). This concept represents RRI using the acronym AREA, which stands for anticipation, reflection, engagement and action. A piece of research or innovation activity, in order to count as having been undertaken responsibly, would need to incorporate anticipation about possible consequences, integrate mechanisms of reflection about the work, its aims and purposes, engage with relevant stakeholders and guid action of researchers accordingly.

Elsewhere (Jirotka et al. 2017) we have developed this AREA framework further by adding what we called the 4 Ps: Process, Product, Purpose and People. This represents an attempt to render the AREA framework more accessible and practical by guiding users to reflect on the process of undertaking the work, guiding their attention to the products or outcomes of the research, explicitly highlighting the importance of considering the purpose of the research and continuously focusing on the people who are involved and likely to be affected in the research and innovation. We have presented this AREA-4P framework as a matrix with each cell in the matrix containing some guiding questions that will help users to consider important parts and aspects of their work. This matrix looks as follows (Table 10.1):

I will use this view of RRI to explore whether and to what degree the BS 8611 standard represents an implementation of RRI in care robots. To do so, I now briefly discuss care robots and the ethical concerns that are typically associated with these.

10.3   Characteristics of Care Robots and Ethical Concerns

Robots as embodied information and communication technology (ICT) that can directly interact with their external environments have been a source of fascination and anxiety ever since they were first proposed. These mixed emotions are most strongly triggered by anthropomorphic robots, but they can be observed in the case of most other robots as well. Positions on robots range from the widely optimistic that see robots as integral part to the solution to most human problems (Brooks 2002) to dystopian visions of robots as

Table 10.1 AREA-4P view or RRI, adapted from (Jirotka et al. 2017)

	Process	Product	Purpose	People
Anticipate	Is the planned research methodology acceptable?	Will the products be socially desirable? How sustainable are the outcomes?	Why should this research be undertaken?	Have we included the right stakeholders?
Reflect	Which mechanisms are used to reflect on process? How could you do it differently?	How do you know what the consequences might be? What might be the potential use? What don't we know about? How can we ensure societal desirability? How could you do it differently?	Is the research controversial? How could you do it differently?	Who is affected? How could you do it differently?
Engage	How to engage a wide group of stakeholders?	What are viewpoints of a wide group of stakeholders?	Is the research agenda acceptable?	Who prioritises research? For whom is the research done?
Act	How can your research structure become flexible? What training is required? What infrastructure is required?	What needs to be done to ensure social desirability? What training is required? What infrastructure is required?	How do we ensure that the implied future is desirable? What training is required? What infrastructure is required?	Who matters? What training is required? What infrastructure is required?

AREA-4P matrix of RRI in ICT (www.orbit-rri.org/framework/)

the end of human civilisation or maybe even humanity, as represented in much popular culture and science fiction output.

In this paper the focus is on care robots, i. e. robotic devices that provide or support the human provision of care or aspects thereof. There is significant overlap with medical robots as well as social robots, such as artificial companions. The exact delineation of care robots is less important here than the concerns that one can find in the literature.

Elsewhere we have described these concerns in more detail (Stahl and Coeckelbergh 2016). In this paper I only briefly recapitulate what these concerns are, to provide the background of the ethical problems that RRI needs to address and that BS 8611 should be sensitive to. We distinguish between three types of ethical concerns, each with a set of individual problems.

The first set of issues arises from a critical evaluation of the vision that drives health-care technology and their implications for society. These include the replacement of human beings and the implications that such replacement has for labour. For instance, in research concerning the development of robots for the elderly, robots are often presented as a response to demographic challenges (Fischinger et al. 2015). But it is not clear that such technological solutions can or should be the solution to the problem. It is similarly not clear to what degree this really constitutes an economic problem and a threat to employment and what the consequences for human care work would be. A second example is the replacement of humans and its implications for the quality of care; the dehumanization of care. An important fear in discussions about robots in healthcare is that robots may replace human care givers, and that this may not only put these people out of job, but also remove the capacity for "warm", "human" care from the care process. It is highly doubtful, for instance, if robots could ever be truly empathic (Stahl et al. 2014) or have emotions (Coeckelbergh 2010). There is the concern that elderly people are abandoned, handed over to robots (Sparrow and Sparrow 2006) devoid of human contact (Sharkey and Sharkey 2010). Concerns arise both with regards to the potential objectification of care givers and care receivers.

A second set of issues has less to do with the idea of replacement and more with human-robot interaction in healthcare. A key issue discussed in this respect is autonomy. While autonomy comes in degrees and not all healthcare robots are autonomous, the concept of machine autonomy is often seen as problematic. In addition to the question of human replacement, it raises fundamental questions about the appropriateness of autonomous machines and the degree to which autonomy would be acceptable. In practical terms this raises questions about liability and responsibility. It is open to debate which roles and tasks should be undertaken by robots and to what degree it is legitimate to provide them with autonomy. On one extreme end of the spectrum of possible answers to this one can find fully autonomous robots that interact with care receivers without human input. In this case one could argue that robots should be endowed with a capacity to undertake ethical reasoning (Anderson and Anderson 2015; Wallach and Allen 2008). However, the very possibility of constructing such ethical reasoning in machines is contested. Increased use of robots in care and the possibility of robots acting increasingly

autonomously raise broad questions about responsibility and liability. The attribution of responsibility to designers, vendors and users of such robots and maybe even the robots themselves is not clear. If something goes wrong, who is to blame and who will be held legally liable? Further concerns include that using robots as social companions, in particular for vulnerable individuals may amount to deception, as the care receivers may not understand the way in which the technology works (Coeckelbergh 2012; Sparrow and Sparrow 2006). Finally, there are issues concerning trust that are raised because the transfer of human activities to robots may invite the transfer of trust originally vested in human carers. However, if is not clear that such a transfer of trust would be appropriate.

In addition to these specific issues related to robots and their role in healthcare and general human interaction, some concerns relate to ICT and its social impact in general. Many of these are applicable to robots in care situation as well. The most prominent among these are privacy and data protection. Robotics research and use of robots in healthcare can raise questions about which data are collected, how they are stored, who has access to them, who owns them, what happens to them, and so on. Furthermore there are questions of safety and avoidance of harm. Robots should not harm people and be safe to work with. This point is especially important in healthcare and related domains, since it often involves vulnerable people such as ill people, elderly people, and children.

This quick overview does not aim to be comprehensive but lists the key issues that can be taken as a first step to determine whether attempts to address ethical issues of care robots have been addressed. Previously we have suggested that RRI provides mechanisms to identify and address these issues (Stahl and Coeckelbergh 2016). In this paper the question is whether BS 8611 provides practical tools to realise this promise and thereby implement RRI for care robots.

10.4 BS 8611—Robots and Robotic Devices: Guide to the Ethical Design and Application of Robots and Robotic Systems

This section starts by giving a short overview of the ethical hazards described in BS 8611. This is followed by a discussion of how ethical risk assessment and management is to be undertaken according to BS 8611.

10.4.1 BS 8611 and Ethical Hazards

BS (British Standard) 8611 (BSI 2016) was published by BSI, the British Standards Institution in 2016. Its scope (section 1) clearly locates the document in the context of existing standardisation around robotics. The main aim is to identify "potential ethical harm" and provide "guidelines on safe design, protective measures and information for the design and application of robots" (p. 1). The overall document is framed in terms of risk identification

and risk management. Physical risks posed by robots are well documented and standards exist to help industry and developers to deal with these. BS 8611 therefore cites as "normative references" (section 2 of the standard) a set of existing standards in this area: BS EN ISO 12100:2010, Safety of machinery—General principles for design—Risk assessment and risk reduction (ISO 12100:2010); BS ISO 8373, Robots and robotic devices—Vocabulary; BS ISO 31000, Risk management—Principles and guidelines.

Unlike the existing body of standardisation, the focus of BS 8611 is on "ethical harm". This is defined in section 3 (terms and definition) of the standard as "anything likely to compromise psychological and/or societal and environmental well-being". An explanatory note elaborates that "Examples of ethical harm include stress, embarrassment, anxiety, addiction, discomfort, deception, humiliation, being disregarded. This might be experienced in relation to a person's gender, race, religion, age, disability, poverty or many other factors." Ethical hazards are defined as sources of ethical harms and ethical risks are said to be "probability of ethical harm occurring from the frequency and severity of exposure to a hazard". It is interesting to note that RRI is also defined in the document as the "process that seeks to promote creativity and opportunities for science and innovation that are socially desirable and undertaken in the public interest".

The subsequent substantive sections of BS 8611 then follow the order of "ethical risk assessment" (section 4), ethical guidelines and measures (section 5), ethics-related system design recommendations (section 6), verification and validation (section 7) and information for use (section 8).

The ethical risk assessment in section 4 starts with a table that lists ethical issues, ethical hazards and ethical risks. For each of these there is a suggested mitigation, space for comments and a validation mechanism. The ethical issues are the top level concerns. They are broken down into societal, application, commercial/financial and environmental. The largest group is that of societal issues. It includes the ethical issues of loss of trust, intentional or unintentional deception, anthropomorphisation, privacy and confidentiality, lack of respect for cultural diversity and pluralism, robot addiction, and employment. Application issues listed are misuse, unsuitable divergent use, dehumanisation of humans in the relationship with robots, inappropriate "trust" of a human by a robot and self-learning systems exceeding its remit. Under commercial/financial issues the BS document lists the appropriation of legal responsibility and authority, employment issues, equality of access, learning by robots that have some degree of behavioural autonomy, and informed consent. The environmental issues section, finally lists the hazards of environmental awareness (robots and appliances) and environmental awareness (operations and applications).

The ethical risk column spells out how these hazards translate into ethical risks. In each case this is followed by a suggested mitigation, comments and validation mechanisms.

10.4.2 Ethical Risk Assessment and Management in BS 8611

The practice and implementation of ethical risk management for robots is described in the text following overview table. Section 4.2 then spells out how ethical hazard identification should be put in practice. This starts with a reminder that the hazards need to be reviewed with the people and animals potentially affected by them. It is also pointed out that new developments in robotics may lead to new ethical hazards and risks.

The next steps of the risk management process are then discussed and related to existing risk management approaches that are described in other standards. It is suggested that ethical hazards can be treated in a similar way to ergonomic hazards. In line with BS EN ISO 14971 it is made clear that ethical hazards and risks for medical technologies will always have to be balanced with the benefits for users or beneficiaries.

Ethical risk assessment, as described in section 4.3 should be undertaken with regards to various human-robot interaction scenarios. This includes unauthorised use, reasonably foreseeable misuse, the uncertainty of situations to be dealt with, psychological effects of failure in the control system, possible reconfiguration of the system and ethical hazards associated with specific robot applications. As a rule of thumb, it is suggested that the risk of a robot performing an operation should not be higher than the risk of a human performing the same operation.

The final subsection of section 4 focuses on learning robots. It offers a categorisation of learning, depending on the degree of autonomy of the robot that covers three stages: environmental, performance enhancement and strategy. Learning robots, it is pointed out, raise additional ethical hazards because they can perform differently from otherwise identical robots.

Section 5 of the standard covers ethical guidelines and measures. Starting with general societal ethical guidelines, the document lists a number of norms that robots need to be taken into account when designing and building robots. These include that robots should not be designed solely or primarily to kill or harm humans, that humans and not robots are responsible agents, that they should be secure and not deceptive. The section refers to other norms, such as the precautionary principle and privacy by design. Roboticists are recommended to work responsibly by engaging with the public, addressing public concerns, demonstrate commitment to best practice, working with experts from other disciplines and the media and providing clear instructions. The document provides a list of groups that can help with engagement with various stakeholders. This is followed by a number of sections that spell out in more detail some of the high level ethical topics of concern, including privacy and confidentiality, respect for human dignity, human rights, cultural diversity and pluralism, dehumanisation of humans, legal issues, the balancing of risks and benefits, individual, organisational and social responsibility, informed consent, informed command, robot addiction and dependence on robots, anthropomorphization of robots and robots and employment. Section 5 concludes by pointing to a number of application areas and specific issues these may raise. The areas include rehabilitation, medical use, military use, commercial and financial guidelines and a reference to environmental and sustainability issues.

Section 6 covers ethics-related system design recommendations which promote the use of inherently ethical design or, where this is not possible, the use of safeguards and protective measures to protect robot users from harm. It also points to the protection against the perception of harm. Such perceptions, e.g. a close encounter or near collision with a self-driven vehicle, can itself count as harm and should be avoided.

Section 7 covers verification and validation. These are defined as follows: "'verification' checks that a system does what its specification requires it to do, whereas "validation" checks that a system does what its users expect. Precise specifications are needed in order to carry out verification, while user engagement is needed in order to carry out validation" (p. 14). Various methods for both verification and validation are then discussed.

The final section of BS 8611 covers information for use of robots. This covers general points such as the language in which usage information should be provided and which potential users need to be able to interact with robots and thus be capable of understanding use information. More detail is given in separate sections on markings or indications as well as the user manual and the service manual.

10.5   BS 8611 as Implementation of RRI in Care Robots

Having now introduced the concept of RRI, the ethical concerns about care robots and BS 8611 this section discusses how and to which degree BS 8611 can help realise or implement RRI.

It is probably not surprising that BS 8611 is closely aligned to RRI and offers a way of implementing or realising RRI in robotics. The standard quotes Rome Declaration on RRI (2014) as an inspiration. It subsequently touches on many of the aspects and components of RRI. From the perspective of RRI it is important that the standard signals to the robotics research community that issues of ethics and social responsibility are to be taken seriously. By having gone through the consultation process that is associated with standardisation and having the official seal of approval from the British Standards Institute, BS 8611 represents an important statement underlining the relevance of these issues.

In addition to the general political support that BS 8611 lends to the principle and idea of RRI, it provides important practical advice to roboticists. Most technicians and robotics researchers and developers are familiar with various ethical and social questions related to robots. The current debate surrounding artificial intelligence, big data analytics and robotics is difficult to ignore and researchers are generally happy to engage in it. There is a big difference, however, between generally engaging in a debate and practically changing one's work to address such issues. BS 8611 represents a tool that helps roboticists to do just that. Technology researchers and developers are used to working with standards and are very likely to be able to implement this standard in practice.

The list of ethical hazards that BS 8611 covers is significant and broadly covers the issues that the general discussion of robot ethics covers. The standard states that the list is not comprehensive, however, which leaves open the question of how the list will be

maintained and updated. And, like all such lists of issues, it carries the risk that researchers who have worked though all of the issues will assume that they have done all there is to do and subsequently overlook issues that are not covered. In light of the extensive coverage of the list this risk is probably low, but it should not be discounted alltogether.

A further positive aspect of BS 8611 is that it explicitly includes a focus on misuse and non-intended use. In our research on RRI in ICT (Jirotka et al. 2017) we have found that this is a topic that researchers and developers are not always comfortable with. A typical position is that research is a valuable good in itself because it contributes to knowledge, which is a value per se. Researchers are often reluctant to engage in the question of what the social consequences of their work will be, frequently and rightly citing the uncertainty of prediction of future use. They are even more reluctant to engage with the question of misuse, tending to see this as a social and policy problem that is beyond their remit. This position is not entirely unreasonable. However, in present research funding environments researchers are typically required to elaborate on the practical impact of their work. The higher the technology readiness level of the work, the closer to market it is, the more specific such views of the impact are expected to be. By the same token, it would seem to be appropriate to reflect on the non-intended practical consequences of research and development. These will not be possible to comprehensively predict, but they are not completely unknowable either. By highlighting the question of misuse, BS 8611 makes an important contribution in raising awareness of this important aspect of robot development.

A further strength of BS 8611 is its emphasis on validation and verification. Again, these are activities that are more prevalent at higher technology readiness levels and less obvious in more fundamental and basic research. Requiring reflection on validation and verification puts pressure on the individuals and organisations involved in work on care robots to develop useful metrics that can be applied to RRI.

While BS 8611 is thus a positive contribution to RRI and most likely applicable and relevant to work on care robots, it raises a number of further questions that will need debate. Firstly, the linked nature of the standard and its reliance on a set of other standards mean that it is not a stand-alone document but requires users to have access to various other standards and the ability to work through and implement those.

There are some questions about the provenance and justification of the content. Section 5 of the standard explicates a number of norms. All of these are perfectly reasonable and reflect the norms that modern democratic societies are trying to uphold. What is not clear, however, is how this list of norms was generated and what its status is meant to be. Presumably it is not a comprehensive list, which leads to the same problem observed earlier with regards to the list of ethical risks. It may lead to an assumption of comprehensive coverage where this is not given and maybe not even possible.

Another observation with regards to the norms in section 5 is that it does not discuss the question of conflicting norms. It is easy to imagine cases where different norms conflict in care robotics, where for example the requirement to protect private data may collide with the social responsibility of care for a patient. Quality of service delivery may

similarly collide with the norm of employment, where a human may be made redundant, if a robot can perform a particular task better than a human. Such value conflicts often arise and are subject of discussion in ethics and RRI (Fleischmann and Wallace 2010; Hedström et al. 2011). BS 8611 could be clearer on how individuals encountering such value conflicts are expected to deal with them.

One reason why value conflicts arise is that there are numerous different values held by members of society and they often do not agree on these or their priorities. This is one of the reasons why public engagement plays an increasing role in research and innovation governance and plays a central role in RRI. BS 8611 recognises this and makes explicit reference to such public engagement. Where the standard could go further is in the description of what such public engagement could look like or how stakeholders could be engaged. There is a rich literature on this topic which is not referenced (Arnstein 1969; Est 2011; European Commission 2014). Moreover, there are various traditions of public engagement. The traditional one sees public engagement as an activity that educates the public to ensure that technologies are met with acceptance. The other position sees public engagement as a mechanism of co-creation technologies in collaboration between researchers or experts and affected stakeholders (Jasanoff 2003). It is important to realise that RRI very much aims at the latter model, sometimes referred to as mode 2 of knowledge creation (Nowotny et al. 2003). While the standard cannot reasonably be expected to cover this complex topic in any depth, it could provide more pointers to existing debates.

Similar to the uncertainty that BS 8611 leaves with regards to public engagement, it also fails to give sufficient guidance on another aspect that is of crucial importance in developing robots responsibly, namely cost-benefit analysis. The standard rightly points out that in many cases different values and aims may need to be considered and decisions have to be made that lead to the overall best outcome. Risk-benefit analysis is thus presumably part of the answer to the question raised earlier, namely how value conflicts can be managed. However, the standard says very little about the practical challenges this raises and how these can be overcome. Similarly to the case of public engagement, at least some more references and pointers would have helped the practical applicability.

And, finally, there are questions to be asked concerning the overall positioning of the standard in the RRI discourse. On the one hand the standard cites the highly aspirational vision of RRI that describes RRI as an attempt to promote creativity and opportunities for science and innovation that are socially desirable and undertaken in the public interest. While this large, and difficult to achieve, aspiration informs the standard, the way it is developed focuses much more on immediate concerns in terms of risk assessment and mitigation. The potential for RRI to be a source of creativity is not developed in depth. Furthermore, the focus on risk assessment limits the type of issue to be discussed. In the case of care robots this means that some of the foundational questions, such as the question how we conceptualise care in the first instance, or which resources our societies are willing to invest in care of older people or people with particular diseases or conditions, cannot be addressed.

This shortcoming is not confined to BS 8611 but it is a general issue of RRI. On the one hand it tends to be advertised with great gestures of better societal embedding of science and research. On the other hand practical implementations tend to be more focused on the delivery of science and research in practice. As this happens in particular organisations, the concerns of these organisations take centre stage which includes a focus on specific risks. One can thus not fault the standard for using a language that its users are likely to understand and respond to. The consequence of this language and the focus it sets is nevertheless a possible impoverishment of RRI in practice.

10.6   Conclusion

This paper has discussed the question whether BS 8611 can help or even implement RRI with a particular emphasis on the application area of care robots. The answer is fundamentally positive and the analysis of the standard in light of the RRI discourse on care robots has demonstrated that BS 8611 can provide important support for researchers and developers working on care robots and wishing to take ethical and social considerations seriously.

The standard is of course not perfect and in the discussion I pointed to several areas that could benefit from additional insights, references or work. However, as successful standards are subject to regular review, one can hope that these changes can be incorporated, if they prove to be truly relevant to the user communities.

This paper was written exclusively on the basis of a textual analysis of BS 6811. The analysis and discussion is thus not informed by empirical insights into the way the standard can be used in practice. This will be the natural next step. Understanding how actual researchers and developers make use of the standard and how this influences their work and its outcomes will be an important next step. This will not only help the individuals working in the area and—ideally—the quality of the outcomes of their work, but it can make an important contribution to the development of RRI. BS 6811 is the first standard that has its roots at least partially in the RRI debate. Other standards are currently in preparation, such as the IEEE family of standards on ethics in ICT. If BS 6811 proves to be successful and can help individuals and organisations to do their work responsibly, then the next wave of standardisation may help promote RRI more broadly and with it its aim to ensure that science and technical developments contribute to a better world.

References

Anderson, S. L., Anderson, M. (2015). Towards a Principle-Based Healthcare Agent. In S. P. van Rysewyk, M. Pontier (Hrsg.), *Machine Medical Ethics* (S. 67–78). Cham: Springer.

Arnstein, S.R. (1969). A Ladder Of Citizen Participation. *Journal of the American Institute of Planners, 35,* 216–224. https://doi.org/10.1080/01944366908977225.

Bogner, A., Decker, M., Sotoudeh, M. (Hrsg.). (2015). *Responsible Innovation – Neue Impulse fuer die Technikfolgenabschaetzung, Gesellschaft – Technik – Umwelt.* Baden-Baden: Nomos.

Brooks, R.A. (2002). *Flesh and Machines: How Robots will change us.* New York: Pantheon Books.

BSI, 2016. BS 8611 – Robots and robotic devices: Guide to the ethical design and application of robots and robotic systems (No. BS 8611:2016). BSI Standards Publication.

Coeckelbergh, M. (2010). Health Care, Capabilities, and AI Assistive Technologies. *Ethical Theory and Moral Practice, 13,* 181–190.

Coeckelbergh, M. (2012). Are emotional robots deceptive? *IEEE Transactions on Affective Computing, 3,* 388–393. https://doi.org/10.1109/t-affc.2011.29.

Est, R. van, (2011). The broad challenge of public engagement in science. *Science and engineering ethics, 17,* 639–648. https://doi.org/10.1007/s11948-011-9296-9.

European Commission. (2014). Vademecum on Public Engagement in Horizon 2020 (Inofficial working document).

European Commission. (2013). Options for Strengthening Responsible Research and Innovation (Report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation). Luxembourg: Publications Office of the European Union.

Fischinger, D., Einramhof, P., Papoutsakis, K., Wohlkinger, W., Mayer, P., Panek, P., Hofmann, S., Koertner, T., Weiss, A., Argyros, A., Vincze, M. (2015). Hobbit, a care robot supporting independent living at home: First prototype and lessons learned. *Robotics and Autonomous Systems.* https://doi.org/10.1016/j.robot.2014.09.029.

Fleischmann, K., Wallace, W. (2010). Value conflicts in computational modeling. *Computer, 43,* 57–63.

Hankins, J. (2012). A handbook for responsible innovation, 1st ed. Fondazione Giannino Bassetti.

Hedström, K., Kolkowska, E., Karlsson, F., Allen, J.P. (2011). Value conflicts for information security management. *The Journal of Strategic Information Systems, 20,* 373–384. https://doi.org/10.1016/j.jsis.2011.06.001.

Jasanoff, S. (2003). Technologies of humility: Citizen participation in governing science. *Minerva, 41,* 223–244. https://doi.org/10.1023/a:1025557512320.

Jirotka, M., Grimpe, B., Stahl, B., Hartswood, M., Eden, G. (2017). Responsible research and innovation in the digital age. *Communications of the ACM, 60,* 62–68. https://doi.org/10.1145/3064940.

Nowotny, H., Scott, P., Gibbons, M. (2003). Introduction: 'Mode 2' Revisited: The New Production of Knowledge. *Minerva, 41,* 179–194. https://doi.org/10.1023/a:1025505528250.

Owen, R. (2014). The UK Engineering and Physical Sciences Research Council's commitment to a framework for responsible innovation. *Journal of Responsible Innovation, 1,* 113–117. https://doi.org/10.1080/23299460.2014.882065.

Owen, R., Heintz, M., Bessant, J. (Hrsg.). (2013). *Responsible innovation.* Hoboken: Wiley.

Rome Declaration. (2014). Rome declaration on responsible research and innovation in Europe.

Sharkey, A., Sharkey, N. (2010). Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and information technology.* https://doi.org/10.1007/s10676-010-9234-6.

Sparrow, R., Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines, 16,* 141–161.

Stahl, B.C., Coeckelbergh, M. (2016). Ethics of healthcare robotics: Towards responsible research and innovation. *Robotics and Autonomous Systems.* https://doi.org/10.1016/j.robot.2016.08.018.

Stahl, B.C., McBride, N., Wakunuma, K., Flick, C. (2014). The empathic care robot: A prototype of responsible research and innovation. *Technological Forecasting and Social Change, 84,* 74–85. https://doi.org/10.1016/j.techfore.2013.08.001.

Stilgoe, J., Owen, R., Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy, 42,* 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008.

Van den Hoven, J., Doorn, N., Swierstra, T. (Hrsg.). (2014). *Responsible innovation 1: Innovative solutions for global issues.* New York: Springer.

Von Schomberg, R. (Hrsg.). (2011). *Towards responsible research and innovation in the informa-tion and communication technologies and security technologies fields.* Luxembourg: Publica-tion Office of the European Union.

Wallach, W., Allen, C. (2008). *Moral machines: Teaching robots right from wrong.* New York: Oxford University Press.

Bernd Carsten Stahl is Professor of Critical Research in Technology and Director of the Centre for Computing and Social Responsibility at De Montfort University, Leicester, UK. His interests cover philosophical issues arising from the intersections of business, technology, and information. This includes ethical questions of current and emerging of ICTs, critical approaches to information systems and issues related to responsible research and innovation.