

Ethics of healthcare robotics: Towards responsible research and innovation

1. Introduction

The past decade has seen a rapid growth of research in the area of ethics of robotics, also and particularly as applied to healthcare. This is unsurprising, since research and innovation in the area of healthcare robotics has seen a significant growth in recent years. Consider for instance research presented in this journal: in response to challenges related to ageing, care robots have been developed to support elderly people living at home (e.g. [\[1\]](#)), robotic nurses have been created to assist with care tasks (e.g. [\[2\]](#)), surgical robots have been designed and used in hospitals (e.g. [\[3\]](#)), and robots have been made more socially interactive (e.g. [\[4\]](#)), which also supports the development and use of robots in health care contexts.

Responding to what is taken to be the near future of health care, ethicists have been especially concerned with what is supposed to be the prospect of intelligent, autonomous, and often also humanoid robots that take care of the elderly. Questions addressed include: Will robots replace the nurses and other care givers, leaving the ill and elderly in the hands of machines? Could robots deliver the same quality of care? Can machines give the “warm”, “human” care we seem to expect from human care givers? Do robots used in care deceive vulnerable persons when they (the robots) “pretend” to be something else than they are, for example when they appear as pets (see Section [2](#))?

These reflections on the future of “machine” healthcare are helpful ways of exploring ethical sensitivities about healthcare, thinking through some of our ethical concerns, developing more refined arguments about what exactly we think might be problematic, and better understanding the current developments in the context of modern healthcare and its politics and the wider developments in robotics and our technological culture. However, they are somewhat limited when it comes to *changing* how things are done in healthcare research, innovation and practice. This is partly so since the context in which academic reflection and research in ethics takes place is largely divorced from the context of innovation and practice. How can this gap be bridged?

In this paper we argue that traditional ways of bridging this gap such as case studies and, more recently, desk-based “value sensitive design”-oriented work, are insufficient to transform ethics of healthcare robotics in a way that really engages with problems as they emerge in innovation, user, and stakeholder contexts. First we review the methods of philosophical reflection on ethical issues, using case studies, and thinking about values in design. Then we discuss what we take to be more dialogical, more democratic, and more effective ways of *doing* ethics: (1) collaborative, “embedded” ethics *in* healthcare robotics which directly and substantially involves ethicists in innovation and user processes and (2) various ways of *really* involving stakeholders in innovation and practice, thus rendering healthcare robotics more ethically and socially responsible. For this purpose we introduce the concept of responsible research and innovation (RRI) and show how a generic framework for RRI in ICT can be applied to healthcare robotics. We frame these options as representing forms of “internal” and “external” dialogue. We also reflect on what the current societal and organisational barriers are that prevent these methods from being widely adopted, and we critically discuss the problems and limitations of these methods. Finally, we reflect on what kind of policies may support these “closer” forms of ethical–technological innovation.

2. Traditional approaches to ethics of healthcare robotics: Philosophical reflection on ethics of healthcare robotics, case studies, and desk-based value sensitive design

There is a growing body of literature on the ethics of healthcare robotics and ICTs [5], [6],[7], [8], [9], [10], [11] and [12], sometimes also called machine (medical) ethics [13], [14],[8], [15], [16] and [17]. The literature gives a good overview of potential ethical issues in healthcare robotics and shows that philosophical reflection delivers valuable insights into what exactly might be problematic in this area and why.

2.1. Ethical concerns

Here are some ethical and social issues and philosophical discussions we identify as central. This subsection is not meant to be comprehensive; it is

meant as a pragmatic and heuristic tool to gain an overview, before we begin the development of the main arguments of this article.

First, there are critical evaluations of healthcare technology visions in terms of their implications for society and on healthcare, for example:

-

Replacement and its implications for labour: Are robots introduced to solve problems in healthcare and elderly care, or are they introduced to save money by replacing human care givers by robots, and to help robotics research and industry? For instance, in research concerning the development of robots for the elderly, robots are often presented as a response to demographic challenges (see again [\[1\]](#)). But are such technological solutions the main or only way we should tackle these challenges? And if there is truth in the suspicion that robots will replace humans, which problems exactly would they solve, and is robotics really a threat to employment? More generally, what are the consequences for healthcare work? For example, do robots and ICTs threaten “care craftsmanship” [\[7\]](#)?

-

Replacement and its implications for the quality of care:
de-humanisation and “cold” 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 empathic [\[39\]](#) or have emotions [\[18\]](#). Robots, it seems, are not capable of a “human” kind of attention and care, whereas healthcare seems to involve more than some “behaviours”; humans have various social and emotional needs, which are not necessarily met by giving them a robot. “Machine care” sounds cold and mechanical. There is the concern that elderly people are abandoned, handed over to robots [\[11\]](#) devoid of human contact [\[10\]](#). More generally, do machines in care “objectify” care receivers? Do they objectify *care givers* (see also the previous point)? What do we mean by good healthcare? Do we have good healthcare

today, without even considering robots? Is good care possible in the context of modernity [\[14\]](#)?

Second, there are issues that have less to do with the idea of replacement as such but are raised by human–robot interaction in healthcare and especially by the robot taking over tasks from humans, for instance:

- **Autonomy.** Not all health care robots are autonomous robots. For instance, surgical robots are remote controlled by the surgeon. Yet health care research often aims to give more autonomy to the robot. An important term in the field, for instance, is *autonomous systems* (see also the title of this journal). Autonomy means here that the robot is designed to carry out tasks without continuous human guidance and assistance, preferably in an unstructured environment. This development could lead to a future scenario in which robots would replace human care workers, for instance if care robots take over the work of the human nurse. As indicated before, this is ethically problematic. But even if robots in healthcare did not entirely replace human care workers, there is still the question how autonomous (in the sense of doing tasks on its own, unassisted by humans) the robot would be and should be in the context of the interaction and the care, and how autonomous it should be in the sense of operating without human supervision. For example, if robots are used in therapy for children, should the robot be supervised (and if so in what way) and what exactly and how much should it do without direct human intervention? (See for instance [\[19\]](#).)

- **Role and tasks.** Related to the previous point is the question regarding the role of the robot in the particular care process. Even if humans are still part of the care process, what exactly should the role of the robot be (and the role of the human)? What tasks can and should be delegated to robots? And in general: should they assist or take over human tasks? When and where should they do what?

-

Moral agency. Robots do not seem to have the capacity of moral reasoning or, more generally, of dealing with ethically problematic situations. Hence when a moral problem arises within the human–robot interaction and within the healthcare situation, there seems to be a problem: the robot is given (more) autonomy, in the sense of doing tasks by itself without human intervention, but does not seem to have the capacity of moral agency: it can do all kinds of things, but unlike humans does not have the capacity to reflect on the ethical quality of what it does. Some philosophers therefore propose to build-in a capacity for ethical reasoning, [\[13\]](#) and [\[20\]](#), whereas other philosophers deny that this is possible or think it is insufficient for dealing with complex ethical issues in healthcare. On the other hand, maybe the robot’s lack of moral agency is not a problem as long as humans are involved and included in the process. Again the issues of autonomy and role are raised.

•

Responsibility. This issue raises again the question regarding the autonomy and role of the robot and the human and, more generally, regarding human–technology relations. How does the introduction of robots (re-)shape ethical responsibilities? If the robot takes over human tasks, who is responsible for these tasks? What should be the new distribution of responsibility, when robots take over some tasks? Does it mean that humans remain responsible (assuming the robots cannot be morally responsible), and if so, how can they exercise this responsibility if they have not direct control over the robot (if they do not continuously intervene) or even do not supervise the robot?

•

Deception. If robots are used as “social” companions and are given other roles which encourage social–emotional involvement of the humans (e.g. elderly people or children), is this not a case of deception, and if so, is this deception justifiable? (See for example [\[6\]](#) and [\[11\]](#).)

•

Trust. In so far as the robot acts autonomously and human care givers withdraw from the care process (to some extent at least), can the robot be “trusted”, or is this term not applicable to robots [\[21\]](#)? Should we

only talk about reliability? Or do “social” robots raise the issue of trust? Shall we “trust” giving patients, elderly people, and children “in the hands of the robot” [\[11\]](#)?

Third, there are issues that are raised by many (research on and uses of) ICTs and by many technologies that involve human users, in particular:

-

Privacy and data protection: Robotics research and use of robots in healthcare 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.

-

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.

Note that generally researchers in the field of health care robotics are very well aware of the latter issues—privacy and data protection and safety—and usually take measures aimed at avoiding these ethical problems. For example, when developing their care robot Fischinger et al. write that their robot should not follow the user all the time (e.g. to the toilet) because of privacy reasons and say that it is their main goal to provide older adults with the feeling of safety [\[1\]](#). Whether or not measures taken are sufficient to ensure privacy and safety is (and should be) of course open to discussion.

There have also been various normative and theoretical approaches to ethics of healthcare robotics, ranging from Kantian ethics and utilitarianism to phenomenology, critical theory, and ethics of care. Each of these approaches has helped to highlight different but often converging moral sensitivities in this area, and sometimes they have also contributed to a better understanding of current healthcare practices, which is vital if we want to think about a better healthcare future. More generally, ethics of robotics and philosophy of robotics is a growing area of research and scholarship which attracts excellent people from various academic backgrounds, and is consolidating itself into a solid (sub)field of its own, even if it remains firmly connected to computer ethics, philosophy of technology, and related areas.

Researchers in robotics, clinicians, and (other) stakeholders may learn from this philosophical work and use it to guide their efforts to improve technologies and practices, and indeed to shape the future of healthcare. Even if some thought experiments or scenarios could be considered to belong to the realm of “science-fiction”, they may help us to think about our values and about what future of robotics and healthcare we want. For example, Coeckelbergh [22] has written a fictional scenario about a robotic dog in order to explore how ICTs and robots may re-shape elderly people’s capacities for social affiliation and engagement in relations with human and non-human others.

2.2. Limitations of the traditional approach

Nevertheless this “a priori” philosophical (and narrative) approach has its limitations when it comes to engaging more directly with technological research and innovation and its stakeholders. The robot ethics literature addresses mainly room for researchers within the robot ethics community and to what they think are important ethical issues. Of course when developing their work they might talk to robotics researchers and take into account public opinion. (In fact most of them do, and it is important to recognise this as a significant and important advance as opposed to an approach which *only* relies on conceptual–theoretical work and speculation.) But in their work they only involve these other parties *indirectly*. The voice of, say, robotics researchers and healthcare practitioners is not *directly* heard. Moreover, it is clear that each discipline has its own character and philosophers are *good* at conceptual work so this should at least be an important part of what they can and should contribute to thinking about ethical issues in healthcare robotics and other fields. But if robot ethics is to be more relevant to technological development and more socially responsive, it is important to also see the limitations of this approach and to explore more dialogical ways of working which may at least be added to the conceptual and speculative work. (See the next section.)

Furthermore, while some of these publications are more empirically oriented than others, and for instance use case studies and/or focus on programming or design, (e.g. [13] and [23]), the action is often still very much located “in the head” of the philosopher–designer based at her or his desk; there is little

explicit dialogue and engagement with (other) designers, users, and stakeholders. If and to the extent that value-sensitive design means in practice that philosophers imagine which values might be embedded in a particular design, without directly and explicitly involving what the designers themselves and (other) stakeholders think, there remains a distance between ethics and design (and more generally technological development). Even a case study which picks up on problems in the present use of healthcare technology is still a rather “remote” way of doing ethics in the sense that the people involved in or impacted by the research have no say in what issues are addressed, how they are addressed, and what normative solutions are provided. The interpretation and the discussion include dialogue with other academic literature but are still too *monological* when it comes to working together with robotics researchers and designers or involving stakeholders.

This is not to say that there is no dialogue or involvement in the research and design of robots. In fact, there is a long and distinguished history of participatory design in all types of information systems, going back to the 1950s and 1960s [\[24\]](#) and [\[25\]](#). This development towards participatory development of technologies was particularly strong in the Scandinavian countries [\[26\]](#) but influential across many of the technology research and design communities. A deeper discussion of this history of participation goes beyond the confines of this paper. Suffice it to say that certain types of participation are still very strongly represented, notably the engagement with intended users of a technology. In the case of healthcare robots this will often include patients and caregivers as well as healthcare professionals. What is less widely employed is participation with other stakeholders as well as participation at earlier stages of research agenda setting. We contend that such broader stakeholder engagement could make it easier to identify possible ethical issues which are currently beyond the scope of discussion.

We acknowledge that there are already many good practices and that there are already participatory approaches in use. Below we try to categorise movements in this direction and work towards contributing to the development of a more systematic framework that can guide and support ethics and responsible innovation in the area of health care robotics.

3. Collaborative and embedded ethics: Integration of ethics in research projects and the principles and practice of responsible research and innovation

Having argued that existing approaches to healthcare robotics are interesting and valuable but arguably not sufficient to make a material and substantial difference to the design, use and societal consequences of healthcare technologies such as care robots, we use this section to explore alternative ways of understanding, highlighting and implementing ethical aspects. There is continuous improvement in the fields of ethics and RRI, but we hope that our efforts at categorisation, systematisation, and critical discussion can contribute to taking ethics and RRI to a next level. We do this (1) by discussing the option to embed research ethics in technological projects and (2) by drawing on the literature on responsible research and innovation (RRI), in particular insofar as it pertains to ICT. These options represent what one could call “internal” and “external” forms of rendering technological development more dialogical, and should ideally be combined. In this paper we will pay extra attention to the second option.

3.1. Embedding ethics in research projects

First, instead of reflecting on ethical issues at a distance, it is possible to make ethics part of technological innovation projects and programmes. The idea is here to *collaborate* with developers of the technology rather than to write *about* what they do and what they should do. Ideally, such collaboration takes the form of an ongoing dialogue about ethics during the research project — with all researchers involved in it. This method recognises that evaluating the ethical and social consequences of technology is not a marginal and additional task which can be outsourced to philosophers and social scientists, but *is essential* to the quality of the project and is a collaborative task in which engineers and scientists play a key role. The importance of this direct interaction between science, research and society with a view to developing sensitivity towards ethical issues within projects has been recognised for some time. This immediate integration of broader concerns into research projects was prominently achieved during the Human Genome Project, where 5% of the overall research budget was dedicated to so-called ethical, legal and social issues (ELSI) (see [\[27\]](#)). This idea of integration of ELSI (sometimes also

called ELSA) work into project has been widely accepted and supported by research funders, such as the European Commission [28] and has led to a research stream in its own right (see e.g. [29]). To a large degree this has laid the foundations for the next step, namely RRI, as will be discussed in the following section.

The research project DREAM,¹ funded under the European Commission's 7th Framework Programme, may serve as a good illustration of how this may work in practice. The project aims to deliver the next generation robot-enhanced therapy by developing clinically relevant interactive capacities for social robots that can operate autonomously under supervision of a therapist. Interestingly, instead of seeing ethics as something that only needs to be discussed in a marginal section of the research proposal or that is dealt with if there is compliance with legal frameworks, the project has built in ethics into the project itself. Ethics of human–robot interaction is not a marginal issue in this project but instead one of the project's 'main challenges'.² Ethical issues raised by the human–robot interaction are explicitly addressed as part of the project, and are dealt with not only by ethics experts (including Coeckelbergh), but by means of collaboration within the entire research team including robotics researchers and clinicians. We believe this way of doing ethics provides a model of how to do healthcare ethics in a dialogical and collaborative way — at least when it comes to dialogue that is mainly “internal” to the project.

Of course this form of ethics presupposes that people have the opportunities and the capacity to engage in cross-disciplinary research and that funding agencies make possible and stimulate this kind of collaboration. Potential barriers to this form of doing ethics are therefore lack of cross-disciplinary education and funding programmes and structures that do not support or even discourage it. We will say more about implications for research and innovation policy in Section 4.

Second, responsible research and innovation provides pathways to “closer” ways of doing ethics not only by means of internal dialogue but also by means of dialogue with (other) stakeholders, for example in the healthcare sector and in industry. As indicated previously this will be our main focus here.

3.2. The concept of Responsible Research and Innovation (RRI)

RRI is a buzzword that takes much of its current relevance from the fact that the European Commission has integrated it into its current research funding programme Horizon 2020. Within Horizon 2020 it has been described a cross-cutting activity. As a consequence, all aspects of the programme which has an overall financial volume of approximately €70 billion between 2014 and 2020 have to adhere to principles of RRI. But what does this mean? The probably most widely cited definition describes RRI as 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)” [\[30, p. 9\]](#).

The discourse on RRI is growing rapidly, also in computer ethics and information systems, [\[31\]](#), and one can consequently find a large number of further definitions of RRI from both academic [\[32\]](#), [\[33\]](#) and [\[34\]](#) and policy sources [\[35\]](#), [\[36\]](#) and [\[37\]](#). One attempt to render the concept accessible to non-specialists was undertaken by the consortium of EU-funded project RRI-TOOLS which is tasked with identifying and disseminating principles and tools of RRI across Europe. For them RRI is characterised as follows:

- “Doing science and innovation with society and for society, including the involvement of relevant stakeholders groups ‘very upstream’ in the processes of research and innovation to align its outcomes with the values and expectations of society;
- A wide umbrella that brings together different aspects of the relationship between science and innovation and society: public engagement, open access, gender equality, science education, ethics and governance.
- A concept which has been adopted as a cross-cutting issue at Horizon 2020, the EU Framework Programme for Research and Innovation 2014–2020”.³

We believe that this latter definition highlights important aspects of RRI for the question of

healthcare robots. One is the importance of societal expectations that mirrors von Schomberg's focus on acceptability and desirability of innovation. The second important point that is highlighted in the second bullet point is that RRI is an umbrella term and that it encompasses a large number of activities and principles. This means that the traditional approaches to technology ethics such as case studies or value-sensitive design that we described above can legitimately be seen as being part of RRI. Adopting an RRI approach thus does not imply a break with the past and ongoing scholarship but an attempt to combine these with further positions and make them practically useful. This explains why we have referred to RRI in terms of a meta-responsibility [38]. The main argument behind this concept is that there are a number of existing responsibilities within the space of research and innovation in technology. RRI does not aim to replace them, but rather to shape, maintain, develop, coordinate and align existing and novel research and innovation-related processes, actors and responsibilities with a view to ensuring desirable and acceptable research outcomes [39].

These considerations demonstrate that RRI aims to be practical and relevant but leave open the question how this might be achieved. This question is discussed in the following subsection.

3.3. The framework for RRI in ICT

The idea behind the framework that we present here was to develop a practical tool that would help ICT researchers, funders or policymakers to find practical ways of integrating societal and ethical concerns in their work. Such a framework cannot be static as societal and ethical concerns change over time and the technologies contributing to such questions evolve rapidly. We therefore designed the framework as a series of questions that we see as a sort of scaffold that allows people grappling with them to find better responses than they had before. In order to provide practical help we compiled sets of possible answers to these questions, knowing well that these sets of answers and the good practice that informs them are also subject to further development.

Given that there is a large possible number of such questions the framework required more detailed structure. We adopted the EPSRC's framework for RRI⁴ that builds heavily on [40] views. It is based on the acronym AREA which stands for:

- **Anticipate** — describing and analysing the impacts that might arise.

- **Reflect** — reflecting on the purposes of, motivations for and potential implications of the research.

- **Engage** — opening up such visions, impacts and questioning to broader deliberation, dialogue, engagement.

- **Act** — using these processes to influence the direction and trajectory of the research and innovation process itself.

These are generic recommendations which can provide aspirations but will in many cases be difficult to implement. Part of our work undertaken to render RRI more applicable in ICT was therefore to consider how these general suggestions could be rendered more specific. We therefore discussed how these general points could be broken down with a view to rendering them more applicable. An initial indication was given by von Schomberg's widely used definition of RRI which distinguishes between the process and the product or outcome of research. The process of research can raise concerns which are often the subject to institutional ethics review. Such review, while widely adopted in biomedical research is much less widely used in ICT research. In addition it typically does not cover broader concerns regarding the consequences of widespread use of the technology, i.e. the outcome. It is therefore helpful to distinguish between process and product when implementing RRI.

However, ethical and societal concerns are not limited by process and product. They can refer to other aspects of research and innovation, notably its purpose. This was an insight gained by their debate concerning the UK SPICE project, which was a geo-engineering project where a public outcry led to the eventual withdrawal of the project, despite the fact that neither the process of this research nor its outcomes were particularly problematic [\[41\]](#). What the public objected to was the very idea of using geo-engineering to address the consequences of human-made climate change. Finally, it is important to underline that RRI aims to strengthen and support the human component of research, innovation and development. People are at the heart of RRI and the impact of research and innovation on them should be explicitly considered.

These considerations prompted us to extend the AREA framework and to develop what we called the 4P's:

- **Process:** covers all activities in preparing research, undertaking data collection and analysis, storage and presentation of data and interaction with respondents.
- **Product:** can refer to products or services. It includes the consequences of use as well as misuse of research products and the impact that research has on the natural and social environment.
- **Purpose:** covers the question why research is undertaken at all.
- **People:** are at the heart of RRI and need to be explicitly considered.

The idea was to use this combination of the AREA framework and the 4Ps to open up their space that would allow for a more fine-grained analysis of possible questions. At this stage, however, this combined framework was still rather generic and not specific to ICT. We therefore expanded our research to identify specific issues, problems or questions that are typical for ICT research and innovation.

The literature offers a discussion of several issues that set ICTs apart from other technologies and that can raise specific social and ethical questions. One specific feature is the ubiquity and pervasiveness of ICTs [42] and [43] which means that they are components of many socio-technical artefacts and have a strong but often underestimated impact on human lives. ICTs differ from other types of technologies in the speed of innovation and diffusion. A new app, for example, can go viral immediately and have large-scale consequences. This is one reason why existing governance mechanisms used in research can be problematic in ICT. A good example of this is informed consent. Informed consent is a fundamental principle in biomedical ethics and research ethics. However, in research involving large numbers of individuals, e.g. in social media research or big data applications using electronic health records informed consent can raise novel problems. ICT furthermore raises novel question due to the distributed and networked nature of the hardware and software artefacts involved. This has been

described as the problem of many hands, which is relevant to other areas such as climate change as well, [44], but which is pertinent in ICT where it is often difficult to trace consequences to individuals [45]. A final feature of ICT worth mentioning is its logical malleability [46], i.e. the broad range of often unforeseeable uses of ICTs that render prediction of consequences exceedingly difficult.

We believe that paying attention to these issues is likely to help researchers and other stakeholders in identifying possible areas of concern. We therefore tried to map the characteristics of ICT to the components of our framework by linking them to the 4Ps. Speed of innovation and diffusion has significant impacts on the way research is done and should therefore be considered when looking at the process. ICT products can raise particular issues and questions because of the ubiquity and pervasiveness which sets them apart from outcomes of research in other fields. This is to some degree caused by the difficulty of distinguishing basic and applied research in ICT. These two components can therefore be considered when looking at the product (or outcome) of research. The logical malleability of ICT raises particular issues with regard to the prediction of eventual uses. This means that the discussion of the purpose of the piece of research is rendered even more difficult, given that intended purpose and eventual use may diverge vastly. Finally, the problem of the many hands is one aspect of the human side of computing and should therefore be fed into the discussion when looking at the role of people.

These considerations concerning the AREA framework, the 4Ps and their specifics of ICT provided the basis for the development of the framework for RRI in ICT that is reproduced in [Table 1](#). The idea was to use the established AREA framework and develop it further to render it more useable for the purpose of ICT research. We need to stress that this is not the only way of conceptualising RRI in ICT. It should be understood as a heuristic device that allows stakeholders to broaden their perspective concerning the issues to be taken into consideration. A more detailed discussion of the relationship between the different components will need to be undertaken elsewhere. We concede that there is a certain amount of overlap between the rows and columns and that a number of issues and items could be located in more than one set of the matrix. We do not perceive this as a problem, however, because

the framework is meant to be a tool to be used by the stakeholders when reflecting on their work. Possible overlaps and redundancies can easily be identified during the process of application of the framework and eliminated at that point.

Table 1.

Scaffolding questions of the framework for RRI in ICT.

	Process –Speed of innovation and diffusion	Product –Ubiquity and pervasiveness –Applied and fundamental research	Purpose –Logical malleability	People –Problem of many hands
Anticipate	Is the planned research methodology acceptable?	Will the products be socially desirable?	Why should this research be undertaken?	Have we included the right stakeholders?
		How sustainable are the outcomes?		
Reflect	Which mechanisms are used to reflect on process?	How do you know what the consequences be?	Is the research controversial?	Who is affected?
	Alternatives:	What might be the potential use?	–Alternatives:	–Alternatives:
		What don't we know about?		
		How can we ensure societal desirability?		
Engage		–Alternatives:		
	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	What needs to be done to	How do we ensure that the	Who matters?

Process	Product	Purpose	People
–Speed of innovation and diffusion	–Ubiquity and pervasiveness	–Logical malleability	–Problem of many hands
	–Applied and fundamental research		
structure become flexible?	ensure social desirability?	implied future is desirable?	
What training is required?	What training is required?	What training is required?	What training is required?
What infrastructure is required?	What infrastructure is required?	What infrastructure is required?	What infrastructure is required?

[Table options](#)

The framework is work in progress that was developed and tested in different stakeholder communities. This initial testing was done via a case study approach where ICT researchers and other stakeholders were asked to consider ethical issues and then apply the framework to investigate whether it improved their awareness and solution strategies. The initial outcomes of this testing of the framework are promising in that they indicate that it has the potential to widen people’s understanding. They also indicate that further work is required to make it more user-friendly and unambiguous. The framework is subject to further debate and we expect it to be a dynamic resource that can guide stakeholders in ICT research. While we expect the shape and details of the framework to develop over time, we believe that its core content is sound and it is therefore suitable to reflect on the shape of RRI in healthcare robots.

The framework forms an integral part of the Observatory for RRI in ICT and the latest version is available on the website.⁵ For easier access we have made it available here in [Table 1](#).

This framework is aimed at ICT more broadly. In order to demonstrate how it can be used in particular application contexts, we now explore its relevance to healthcare robotics. It is important to note that the framework does not represent a static instrument that solves all problems of responsible innovation in ICT. Its purpose is to facilitate a more informed and better reflected discussion of such issues. It will need to develop with technologies and their

applications and requires continual input from the stakeholders involved. The following discussion serves as one way to reflect on the quality and content of the framework.

3.4. Implementation of RRI in healthcare robotics

Having introduced the framework for RRI in ICT, we are now in a position to discuss its application to healthcare robotics. Doing this in a journal article can only produce a somewhat abridged and exploratory version of the full use of the framework, as it is designed to be a communicative tool, a basis for a broader discussion with various stakeholders. The idea behind it is to use the scaffolding questions to explore various angles of the work. In this paper we will look specifically at the ethical issues highlighted above to see how the framework can deal with them. In a second step we then use the framework to discuss aspects that work on ethics and technology normally does not cover.

The way in which the framework is designed to work is to give potential users, be they researchers, funders, policymakers, and interested members of society or representatives from industry, a set of questions worth exploring. The live version of the framework that can be found on the Observatory website (www.responsible-innovation.org.uk) provides further links and suggestions for possible answers. Due to space restraints we can only pick up a limited number of questions and issues in this paper.

The first question that we need to answer is whether the framework is sensitive to the questions and problems highlighted in the debate on robotic healthcare outlined earlier. These were: replacement and its implications for labour, replacement and its implications for the quality of care: de-humanisation and “cold” care, autonomy, roles and tasks, moral agency, responsibility, trust, deception, privacy and safety.

The way to assess the sensitivity of the framework to these issues is to work through the individual questions that populate the matrix. In this paper we will focus on those questions that can be used to highlight the issues of healthcare robotics and explore how these relate to existing discourses.

The proposed RRI framework and the academic discourse summarised in the first part of the paper (see Section [2.1](#)) are closely related in the area of

'product', which is the place to explore outcomes and consequences of research and innovation activities. In particular questions in the cell defined by 'product' and 'reflect' are pertinent to the discussion. The following ethical concerns introduced earlier are good examples of such questions of the consequences of the "product" of the research: Replacement of labour, implications for care, privacy or safety. These questions are how future consequences can be known, what potential uses might be and how societal desirability can be ensured. Questions of anticipation of the product, namely whether products are likely to be socially desirable and the sustainability of outcomes are also of relevance. To some degree the existing literature also covers questions of purpose, namely when critically discussing the way in which robots in healthcare are portrayed as solutions to socio-economic problems. Questions of stakeholder inclusion are also sometimes touched upon.

Our interpretation is that the current discourse on robotics in healthcare constitutes a number of important answers to core questions of the framework for RRI in ICT. This indicates that the framework is well formed and addresses important aspects. At the same time it indicates that RRI can benefit from actively reaching out to other areas. The detailed knowledge represented in discussions of healthcare robots outlined earlier provide an example of the substantive knowledge and insights into ethical issues required for RRI to have the desired impact. We will return to this relationship between RRI and specific knowledge in the conclusion of this paper.

Another important question is which aspects of the framework are not addressed and whether conclusions can be drawn from such gaps. We have argued that the current discourse is strong in answering questions related to the product dimension, to some degree the purpose dimensions. It engages in anticipation and reflection. Areas that not as clearly covered are the process dimension, the people dimension and issues related to engagement and action.

The 'process' dimensions cover questions related to the process of research. It may well be that this is underdeveloped because scholars in healthcare robotics do not believe that such research raises interesting issues in terms of research process and methodology or that they believe that existing

governance mechanisms are sufficient to address such issues where they arise. Similarly, the people dimension may be deemed to be covered implicitly as much of the discourse covers the consequences for people, for example when discussing the loss of interpersonal relationships in care situations.

The 'engage' aspect of the framework is aimed at ensuring early and continuous engagement between researchers and broader stakeholder groups to ensure that civil society plays an appropriate role. This includes questions of methodology of identifying appropriate stakeholder groups and practical engagement activities. It furthermore requires questions on the acceptability of the research agenda and critical scrutiny of who benefits from research. It should be noted that this is an interpretation of engagement that goes beyond the relatively well established engagement with users and potential customers that forms part of many development methodologies. The idea here is to look into how societal concerns can be reflected at all stages of research and innovation processes.

Such questions may well play a role in the background of the healthcare robotics discourse, but they are not normally discussed in depth. This may be because the majority of the contributors of the discussion we cited in Section [2](#) are not themselves active in the technical research but take more of an observer position.

The last row of the framework, the one pointing to 'act' is probably the one that is most distant from the academic discourse. 'Act' looks at questions of flexibility of research process, mechanisms of agreeing desirable futures and their implementation through technical research. Moreover, it raises questions of training required to achieve the desired outcomes and infrastructure to be put in place to ensure that the aims of responsible innovation can be met.

Again, one can speculate that these questions are discussed in less detail because the scholars who take part in the discussion of ethics of healthcare robotics are often not directly involved in the research activities and these questions point to practical interventions in those areas where the technical work takes place. A further explanation for the different levels of attention paid to these different types of ethical questions may be that the more practical questions are seen as part of "research ethics" and therefore separated from ethics of robotics which is deemed to be about ethics of (future) use.

We hope that this short summary demonstrates that the traditional approaches outlined above are important and play a central role in being able to apply the proposed framework for RRI in ICT to the specific challenges raised by healthcare robotics. There is much good practice and awareness of ethical and social aspects within healthcare robotics research. We hope that the application of the framework can highlight good practice as well as point to aspects that are in need of further development. At the same time the framework highlights areas that are in need of further attention and points to broader questions of research policy that we discuss in the following section.

4. Further discussion and implications for policy

The current discourse on ethical and social issues in healthcare robotics shows a rich landscape of enquiry in the area that is important to understand and address possible issues. We have argued that this is an integral aspect of RRI and that the framework outlined above can be a way of integrating such questions. At the same time the analysis of the discourse using the framework has demonstrated that this discourse has limitations and that the application of the framework can highlight areas worthy of further development.

This raises the question what can be done in order to complement the current discourse with components that will allow for the coverage of currently underdeveloped aspects. On a very general level we conclude that the integration of ethical and societal reflection into research and innovation activities should be promoted. This implies initiating and enabling novel dialogues both within research projects and activities and between various research communities, both technical and reflective, and external stakeholders from user groups, research policy and civil society more general.

The following bullet points represent initial suggestions that we believe to be important for the broader uptake and implementation of RRI in healthcare robotics. These, of course, need to be understood in the social and political context in which the research is undertaken. This context varies greatly and our ideas would need to be tailored to specific environments. A more detailed analysis that could explain how these suggestions could be put into practice will need to be undertaken elsewhere. Numerous important questions, such as who would be responsible for implementing them, how would decisions be

made, which costs would be incurred and who should pay for them, etc. remain open for the moment. However, we hope that these initial suggestions can contribute to the wider debate on implementing RRI and rendering it practically relevant:

- - Reform education: Further integration is required of ethics in technological programmes. At present such integration is typically part of professional accreditation but it is rarely broad and comprehensive. Such educational reform should equip technical researchers to appreciate the relevance of these questions but they will not all be able to become experts in these fields as well. Education should therefore include encouraging humanities and social sciences to engage with technological research and development.
- - Reform existing funding programmes at national and supranational level: The EU is focusing on these questions in the H2020 funding programme but questions of implementations are still largely unclear. Many national funding organisations are far behind — we urgently need steps to redress this situation and ensure that RRI is integrated into all research and innovation funding activities.
- - Provide shared resources for healthcare robotics research stakeholders: We have seen that there is a rich discussion of ethical and social aspects of healthcare robotics. While this discourse is published and thus publicly accessible, it would be desirable to have a central point of contact that technical researchers are aware of that provides links to existing knowledge and good practice. An initial attempt to build such a resource was undertaken with the Observatory for RRI in ICT (www.responsible-innovation.org.uk) but such resources require community support and a sustained infrastructure, both in terms of ongoing ICT support and in terms of maintenance of networks.
- - Develop of standards and certifications: Individual interest and commitment to engage with these questions is important but in the

longer term a successful implementation of RRI will require integration into existing structures and processes. This refers to organisational and incentive structures as well as formal process of research and evaluation. There are numerous possible ways to achieve this. One particular one that we believe may be suitable and successful is the development of standards and certifications. There are already some initiatives that seem to go in this direction (see for example the work of the UK Robot Ethics platform <http://www.robotethics.org.uk/>). These mechanisms could be linked to all of the earlier points and provide objective and measurable ways of integrating RRI into healthcare robotics.

•
Include Civil Society in agenda setting: Where healthcare robotics are framed in terms of addressing a social problem, the definition of this social problem and the assessment of the success of addressing it should not be left to researchers alone. We believe that the ‘engage’ aspect of RRI, the inclusion of stakeholders and civil society in all aspects of research is key to identifying problems and possible solutions. One central component of this engagement is to allow the affected stakeholders to be engaged early and set the agenda for research and innovation.

While this paper has focused on the particular questions of healthcare robotics, it is clear that these cover just one area of research and innovation and that similar considerations are valid for most other scientific and technical research fields. We now spell out in the concluding section what this could mean for research and innovation more broadly.

5. Conclusion

In this paper we have outlined ethical issues related to healthcare robotics by summarising the main points of the debate. We have also argued that the present discussion has limitations when it comes to changing and engaging more directly with research and innovation (and healthcare) practices. We first offered a way of bridging the gap between ethics and research/innovation in this area by embedding ethics in research projects, constituting possibilities for “internal” dialogue. We then proposed RRI as a way to extend participatory

processes to dialogue with “external” parties and to deal with (other) issues not addressed by the former methods and approaches. We introduced the concept of RRI in ICT and outlined the framework of scaffolding questions that was developed to realise RRI. Using the framework we identified where the current discourse on the ethics of healthcare robotics contributes to RRI and which areas are in need of further development. We used this as a basis to suggest some policy advice to steer research policy in ways that would encourage RRI.

We realise that what we propose is not without problems. We are not suggesting that RRI will be simple, straightforward or linear. There are often different and sometimes contradictory interests at play, for example in cases where industry wants to sell robots but researchers have other aims, such as publishing their findings. RRI is not a panacea that will make conflicting interests disappear. However, we do believe that it can help surface and identify problems and provide the basis for a more intelligent discussion of options and possible solutions.

5.1. Further reflections on RRI and its political potential and ambitions

The radical potential of RRI lies in the fact that it promotes a different conceptualisation of the role of science and research in society. It constitutes a break with the traditional social contract of science [\[47\]](#) where science was given broad freedom in exchange for knowledge and trained scientists. Instead, it recognises the key role that society and its stakeholders can play not only as users of research and innovation but also in terms of setting agendas and exploring desirable futures to be achieved through research.

This view of RRI should be read in the broader context of research and science policy in modern Western societies. There are and have long been calls for more influence on research by affected stakeholders and civil society in general. This debate is closely linked to high-level public debates around particular scientific and technical developments, such as nuclear energy or genetically modified organisms. Science and science policy have reacted to this by a strengthened emphasis on public engagement and dialogue [\[48\],\[49\]](#) and [\[50\]](#). Very briefly, the question is whether and to what degree the public could and should have a say in research beyond its current role as users and consumers. This is a very large debate that this paper

cannot do justice to. However, it is important to see that this debate can lead to a different view of the role of research in society and the tasks that various stakeholders have in it.

This reconceptualisation of research points to its political and potentially contested nature. Not all researchers and research stakeholders will be comfortable with such a view of science but it has the great potential to open spaces for political decisions. The field of healthcare robotics clearly demonstrates the importance of such a political conception of research. The question whether robots should be used for the care of older people and how such care activities should be designed is not predominantly a technical one. It is strongly influenced by societal factors ranging from dominant family structures to the health insurance system. RRI requires rethinking what constitutes good research which may well entail a move away from the evaluation of research quality exclusively by researcher and the development of scientific excellence that takes into account broader societal aspects.

In this paper we explored what this might mean for the field of healthcare robotics. We propose more dialogical ways of doing ethics of healthcare robotics, involving both internal and external forms of dialogue. This does not mean that we should stop doing conceptual work, speculation, writing scenarios, etc. or that purely technical research has no place. In fact, such conceptual and technical work is a key component without which RRI cannot be effective. However, what it does mean is that other methods such as collaborative, embedded ethics should be added in order to render ethics of healthcare technologies, including ethics of healthcare robotics, more relevant to research and technological development. The vista of RRI is to render research and innovation activities more socially responsible, participatory and democratic. This is likely to require technical and scientific expertise to be mixed with the reflection on ethical and societal issues by philosophers and social scientists and the active integration of civil society.

This is the vista of RRI which is already implemented to varying degrees in current research and innovation projects in the area of healthcare robotics. But there is much room for development. At present there are still many open questions calling for better empirical insights into questions that range from incentive structures conducive to RRI to funding and management regimes

that support it. One outcome of RRI so far is already that traditional research and innovation systems are being questioned and rethought. This is most notable at the European level but also increasingly visible in nationally and privately funded research systems. We believe that this is an important initial step in ensuring that research and innovation in healthcare robotics and elsewhere fulfils their aim of providing acceptable and desirable outcomes.

Acknowledgements

This research benefited from the activities undertaken in the project “Framework for Responsible Research and Innovation in ICT”, EPSRC reference EP/J000019/1.

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007–2013) under Grant agreement Nos. [611391](#) (DREAM), [609817](#) (Responsible-Industry).