Bert-Jaap Koops · Ilse Oosterlaken
Henny Romijn · Tsjalling Swierstra
Jeroen van den Hoven *Editors*

Responsible Innovation 2

Concepts, Approaches, and Applications

Springer

Responsible Innovation 2

Bert-Jaap Koops · Ilse Oosterlaken
Henny Romijn · Tsjalling Swierstra
Jeroen van den Hoven
Editors

Responsible Innovation 2

Concepts, Approaches, and Applications

Springer

Editors
Bert-Jaap Koops
Tilburg Law School
Tilburg University
Tilburg, The Netherlands

Tsjalling Swierstra
Faculty of Arts and Social Sciences
University of Maastricht
Maastricht, The Netherlands

Ilse Oosterlaken
Philosophy Section, Faculty of TPM
Delft University of Technology
Delft, The Netherlands

Jeroen van den Hoven
Philosophy Section
Delft University of Technology
Delft, The Netherlands

Henny Romijn
Eindhoven University of Technology
Eindhoven, The Netherlands

ISBN 978-3-319-17307-8 ISBN 978-3-319-17308-5 (eBook)
DOI 10.1007/978-3-319-17308-5

Library of Congress Control Number: 2015937363

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

Contents

Editors and Contributors

About the Editors

Bert-Jaap Koops is professor of Regulation & Technology at the Tilburg Institute for Law, Technology, and Society (TILT), the Netherlands. His main research fields are cybercrime, cyber-investigation, forensics, privacy and data protection. He is also interested in topics such as identity, digital constitutional rights, ‘code as law’ and regulatory implications of human enhancement, genetics, robotics and neuroscience. Koops studied mathematics and general and comparative literature at Groningen University and received his Ph.D. in law at Tilburg University in 1999. He co-edited eight books in English on technology regulation, including *Emerging Electronic Highways* (1996), *Starting Points for ICT Regulatio*n (2006), *Constitutional Rights and New Technologies* (2008) and *Engineering the Human* (2013).

Ilse Oosterlaken has a background in engineering, social sciences and philosophy. In January 2013, she obtained her Ph.D. from Delft University of Technology, cum laude. Her dissertation provided a philosophical exploration of the application of the capability approach of Sen and Nussbaum to technology and its design. She co-edited the first book to have appeared on this topic, *The Capability Approach, Technology and Design* (Springer, 2012). In April 2015, Routledge will publish her monograph on the same topic. Her key research interests are in the area of ethics of technology, development ethics and value sensitive design (and their intersections).

Henny Romijn is an associate professor of Technology and Development Studies in the School of Innovation Sciences of Eindhoven University of Technology. Her research focuses on sustainable pro-poor innovation in the global South, using evolutionary economics and transition studies as analytical perspectives. Before turning into an academic, she worked for the International Labour Organisation on employment promotion projects in East Africa and South Asia. She is currently coordinating a programme funded by the Netherlands Organisation for Scientific Research’s programme on responsible innovation, which addresses the drivers and consequences of the emergence of contentious biofuel projects in Tanzania and India.

Tjalling Swierstra is full-time professor and head of the Philosophy Department of the University of Maastricht. He is a member of the Dutch Health Council and of the Program Committee of the 'Responsible Innovation' programme funded by the Dutch Research Council NWO. He has published widely on the ways in which ethical and political beliefs influence scientific and technological research and vice versa—how science and technology affect (moral) values, philosophies and political beliefs.

Jeroen van den Hoven is professor of moral philosophy at Delft University of Technology. Van den Hoven is Vice Dean of the Faculty of Technology, Policy and Management. He is former scientific director of the Centre for Ethics and Technology of the Three Technical Universities in the Netherlands (www.ethicsandtechnology.eu) and Editor in Chief of Ethics and Information Technology (Springer). Jeroen van den Hoven is chair of the responsible innovation ('Maatschappelijk Verantwoord Innoveren') programme of the Netherlands Organization of Scientific Research (NWO).

About the Contributors

Annelies Balkema is especially interested in innovations for a more just and sustainable development worldwide. In line with her background in environmental engineering and development studies, her research focus is on how economic and sociocultural factors influence the success of technological innovations and on the impact of technological innovations on societies and the environment. Her case studies include mostly renewable energy, drinking water and sanitation. She has worked for the Dutch Environmental Assessment Agency and the School of Innovation Sciences at Eindhoven University of Technology. The chapter she co-authored in this book is written for the NWO-MVI funded research project 'Biofuels; sustainable innovation or gold rush?'.

Vincent Blok is an assistant professor of Innovation and Entrepreneurship at the Social Sciences Group, Wageningen University (the Netherlands) and also research fellow in philosophy at the Faculty of Philosophy, Theology and Religious Studies, Radboud University Nijmegen (the Netherlands). In 2005, he received his Ph.D. degree in philosophy at Leiden University with a specialisation in philosophy of technology and philosophical method. He has published various articles on themes in the philosophy of technology and, responsible innovation, ecological innovation and innovation management.

Jacqueline Broerse is professor of Innovation and Communication in the Health and Life Sciences. She holds a master's degree in biomedical sciences and obtained her Ph.D. degree on the development of an interactive approach in research agenda setting processes. Her current research is focused on methodology development for facilitating a dialogue between science and society, and management of systemic change processes, to contribute to more equitable and inclusive innovation processes. Many of her research projects are in the health sector.

Christine Boshuijzen-van Burken holds a B.Sc. in Human Kinetic Engineering (The Hague University of Professional Education, 2002), a B.Sc. in Mechanical Engineering (Eindhoven, Fontys University of Professional Eduction, 2004) and a MA at the Amsterdam Vrije Universiteit, Faculty of Philosophy (2006). She was affiliated to Eindhoven University of Technology, Delft University of Technology, TNO Human Factors and the Netherlands Defence Academy as a NWO Ph.D. candidate in the field of ethics and military technology.

Aad Correljé is associate professor of Economics of Infrastructures at the Faculty Technology, Policy and Management, TU Delft, and research fellow with the Clingendael International Energy Programme (CIEP). He is a member of the Editorial Board of Energy Policy and an instructor at the Florence School of Regulation. His research involves an economic/institutional approach to the analysis of public policy, regulation and private strategy development in infrastructure-bound sectors, such as oil, gas, power, water supply and flooding protection. He is regularly commenting in the press, on radio and TV, and advising international organisations, governments, public and private enterprises and NGOs.

Eefje Cuppen is assistant professor at the Faculty of Technology, Policy and Management of Delft University of Technology. She has a background in innovation sciences and completed a dissertation on stakeholder participation in environmental policymaking. Her current research focuses on public engagement and stakeholder dialogue in the field of energy, innovation and sustainability transitions.

John Danaher is a lecturer in law at the National University of Ireland, Galway. He holds a BCL from University College Cork (2006); an LLM from Trinity College Dublin (2007); and a Ph.D. from University College Cork (2011). His research interests lie, broadly, in the areas of philosophy of law and emerging technologies and law. In the past, he has published articles on human enhancement, brain-based lie detection, the philosophy of punishment and artificial intelligence.

Rudi Darson studied Public Administration at Leiden University. He is an independent consultant in change management and development and also taught classes in economics, management, business information systems and knowledge management. He has been involved in developing projects in Suriname with attention to education and capacity building for sustainable development for more than 15 years.

Sarah R. Davies is a research fellow in the Department of Media, Cognition and Communication at the University of Copenhagen, where her work focuses on public engagement with science, deliberative democracy, and the governance of new and emerging technologies. She holds a Ph.D. in Science and Technology Studies (2007) from Imperial College's Science Communication Group. Her publications include the edited volumes *Science and Its Publics* (2008) and *Understanding Nanoscience and Emerging Technologies* (2010) and articles in journals such as *Futures*, *Science Communication*, *Science as Culture*, and *Public Understanding of Science.*

Marije de Jong holds master's degrees from Utrecht University (Biochemistry) and VU University Amsterdam (Management Policy-Analysis and Entrepreneurship in the Health and Life Sciences). Latter, an interest in the interplay between science, technology and society was ignited. Currently, she is pursuing a Ph.D. at the Athena Institute, Department of Science Communication (VU University Amsterdam). Her research project, funded by the NWO programme 'Responsible Innovation', focuses on the responsible embedding of neuroimaging technologies in justice and security. The context of justice and security has a particular appeal to her, considering her fascination with 'deviancy' and 'otherness'.

Filippo Santoni de Sio is postdoctoral researcher in philosophy at Delft University of Technology. He is a member of the project Enhancing Responsibility: the effects of cognitive enhancement on moral and legal responsibility. He studied concepts such as action, capacity, will, intention, accountability and blame, with an interdisciplinary approach (philosophy, law and neuroscience). His areas of specialisations are the philosophy of criminal law, moral responsibility, the theory of action, neuroethics and neurolaw. His personal Website is filipposantoni.net.

Marloes Dignum holds an engineering degree and a Ph.D. in the field of innovation sciences of Eindhoven University of Technology. Her expertise focuses on sustainability transitions in the energy domain. Her Ph.D. research comprised a longitudinal study of the influence of large technological visions by focusing on the vision of hydrogen energy. Marloes currently works as a postdoctoral researcher at the section economics of infrastructures of Delft University of Technology where she focuses on social responsible innovation in relation to shale gas developments.

Rosanne Edelenbosch has a background in medicine, medical anthropology and MPA (Management, Policy Analysis and Entrepreneurship in Health and Life Sciences). She is currently doing Ph.D. at the Department of Science Communication (Athena Institute) at the VU University Amsterdam. Her action-oriented research is concerned with the social and ethical aspects of the application of brain research in the field of learning and education, with a focus on neuroimaging. An important aim of this study was to further advance methods for bringing together stakeholders from science and society in creating a more socially robust research agenda.

Grace Eden is a Senior Research Associate at the University of Oxford Computer Science Department and leads ethnographic research for the EPSRC funded project 'Framework for Responsible Research and Innovation in ICT'. She engages in interdisciplinary research with a focus on understanding the ways in which technology effects communication, collaboration and knowledge exchange within scientific, work and home settings. Her research extends across three areas of social science research: the sociology of scientific knowledge (SSK), computer supported cooperative work (CSCW) and the broader impacts of science and innovation upon wider society.

Nadira Faber received her M.Sc. in Psychology (University of Munich, Germany) and Ph.D. in Experimental Social Psychology (University of Göttingen, Germany). Since 2012, she is a postdoctoral research fellow at the Department of Experimental Psychology, University of Oxford (United Kingdom). She is also a fellow of the Oxford Martin School and a Member of the Centre for Ethics and Technology at Delft University of Technology (the Netherlands). Nadira's current areas of research are social decision-making, moral decision-making and cognitive enhancement.

C. Flick is a lecturer in Computing and Social Responsibility at the Centre for Computing and Social Responsibility at De Montfort University, Leicester. Her Ph.D. was on the topic of Informed Consent in ICT, at the Centre for Applied Philosophy and Public Ethics, Charles Sturt University, Australia. Areas of research interest include Responsible Research and Innovation, video games and ethics, online child protection technologies, ethical governance of ICT and informed consent in ICT. Current and past funded project involvement include the following: Responsible-Industry (EU), FRRIICT (UK), Isis (UK), ETICA (EU) and EGAIS (EU).

Dirk Haen is a Ph.D. candidate at the Philosophy Department of Maastricht University, the Netherlands. He earned his master's degree at the University of Amsterdam (2008). His field of interest includes the ethics and politics of food. Focus of his current research is the discursive conditions under which citizen concerns about food and technological innovation gain access to the public agenda.

Maja Horst is Head of Department of Media, Cognition and Communication at University of Copenhagen. She holds a master's degree in communication and a Ph.D. in science and technology studies. Her research is focused on public understanding of science, science communication, research management and sociology of innovation. She has been directing a research project on the 'Scientific Social Responsibility' in 2009–2013 funded by the Danish Research Council for Social Science. Among other places, she has published in *Social Studies of Science*; *Science, Technology and Human Values*; *Public Understanding of Science*; *Science as Culture* and *Science and Public Policy*.

M. Jirotka is professor of Human Centred Computing in the Department of Computer Science at the University of Oxford, associate director of the Oxford e-Research Centre and associate researcher of the Oxford Internet Institute. Her research interests have long been concerned with bringing a richer comprehension of socially organised work practice into the process of engineering technological systems with a focus on supporting everyday work and interaction. Marina is a Chartered IT Professional of the BCS and sits on the ICT Ethics Specialist Group committee. She has published widely in international journals and conferences in e-Science, HCI and CSCW.

Bert-Jaap Koops is professor of Regulation & Technology at the Tilburg Institute for Law, Technology, and Society (TILT), the Netherlands. His main research fields are cybercrime, cyber-investigation, forensics, privacy and data protection. He is also interested in topics such as identity, digital constitutional rights, 'code as law' and regulatory implications of human enhancement, genetics, robotics and neuroscience. Koops studied mathematics and general and comparative literature at Groningen University and received his Ph.D. in law at Tilburg University in 1999. He co-edited eight books in English on technology regulation, including *Emerging*

Electronic Highways (1996), *Starting Points for ICT Regulatio*n (2006), *Constitutional Rights and New Technologies* (2008) and *Engineering the Human* (2013).

J. Otto Kroesen is assistant professor at Delft University of Technology, the Netherlands. He teaches ethics and intercultural communication and is an expert in development theory. They published a Ph.D. on the ethical philosophy on Levinas and he publishes the historical language philosophy of Rosenstock-Huessy and the nexus of values, society, technology and development.

Frank Kupper studied both biomedical sciences and philosophy of the Life Sciences at Utrecht University. He wrote his Ph.D. thesis at the Athena Institute, VU University, on the experimental development of deliberative instruments that facilitate a stepwise reflective learning process about the animal biotechnology controversy. Currently, he is assistant professor of Science Communication at the Athena Institute, focusing on playfulness, reflexivity and learning in science–society interactions. In 2007, he founded his own company 'Mens in de Maak', which is dedicated to the use of theatre and dialogue as instruments of playful reflection.

Pieter Lemmens studied biology and philosophy at the Radboud University in Nijmegen. He wrote his dissertation on the intimate relationship between the human and technology—entitled *Driven by technology. The human condition and the biotechnology revolution*—and got his Ph.D. in 2008 from the Radboud University in Nijmegen. Since 2008, he has conducted several postdoctoral research projects at both Wageningen University and Research Center and the Radboud University. He currently teaches philosophy and ethics at the Radboud University and sociology at Wageningen University and Research Center. He has published, among other things, on themes in the philosophy of technology and on the work of Martin Heidegger, Peter Sloterdijk, Bernard Stiegler and Michael Hardt and Antonio Negri.

Hannah Maslen is a postdoctoral research fellow in Ethics, working on the Oxford Martin Programme on Mind and Machine at the University of Oxford. She is also a junior research fellow at New College. Hannah's academic background is in philosophy, psychology and law: she received her BA in Philosophy, Psychology and Physiology (PPP) from Oxford in 2007, her M.Sc. in Criminology and Criminal Justice from Oxford in 2008, and her DPhil in Law from Oxford in 2011. Hannah's current research focuses on the ethical, legal and social implications of various brain intervention technologies.

David J. Ndegwah is senior lecturer in ethics and religious studies at Jaramogi Oginga Odinga University of Science and Technology, School of Humanities and Social Sciences, Bondo, Kenya. He published a Ph.D. on hermeneutics for the Pokot. He publishes on society of development and technology for the African context.

Quang A. Nguyen aka Bob is a master of multimedia and Internet computing. His research interests include database design and utilising CMS in ICT systems. Acting as the Technology Support Officer in the Centre for Computing and Social Responsibility at De Montfort University, Leicester, UK, he got involved with several projects such as CONSIDER, FRRIICT and RESPONSIBILITY.

Udo Pesch is assistant professor at Delft University of Technology. He is affiliated to the Department of Values and Technology of the department of Technology, Policy and Management. His disciplinary interests include science and technology studies, environmental politics, sustainable innovation, public administration and philosophy.

Auke Pols is a postdoctoral researcher at the School of Innovation Sciences, Eindhoven University of Technology. He got his Ph.D. at the Philosophy and Ethics section in Eindhoven, where he wrote a dissertation on how we should conceptualise what actions with artefacts are and how artefacts influence agents and agency. He currently works on the NWO-MVI project ‘Biofuels: sustainable innovation or gold rush?’ where he investigates how we should understand the notion of ‘sustainability’ in the context of biofuel production and use and what this implies for Dutch and European biofuel policy.

Zoë Robaey is a doctoral candidate in Ethics of Technology at Delft University of Technology. With a background in Public Policy and Science and Technology Studies, her work focuses on the intersection of ethics and public policy. She is currently investigating the link between ownership and moral responsibility for hazards of biotechnologies in the project ‘New Technologies as Social Experiments’ supported by the Netherlands Organization for Scientific Research (grant number 016.114.625).

Anneloes Roelofsen has a background in medical biology. She obtained her Ph.D. degree at the VU University Amsterdam on the development of a constructive technology assessment (CTA) approach that aims to facilitate broad societal reflection on emerging technologies and to guide research activities into societal desirable directions. In her current position at the Dutch Cancer Society, she focuses on patient participation in health research.

Andri D. Setiawan is a Ph.D. candidate at Eindhoven University of Technology. His Ph.D. research is fully funded by the Indonesia Endowment Fund for Education (LPDP), Ministry of Finance of Republic of Indonesia. His research interests lie primarily in innovation sciences and policy and now focus on developing a framework for responsible innovation in practice combining qualitative and quantitative approaches. Before coming to Eindhoven, he obtained his M.Sc. degree in Engineering and Policy Analysis from Delft University of Technology.

Arno Simons is a Ph.D. researcher at the Institute of Sociology of the Technische Universität Berlin. He currently completes his doctoral thesis on the discursive construction of emissions trading and how this process took shape as a real-world experiment mediated in an unfolding network of related documentary utterances. As a member of the Innovation in Governance Research Group, he also works on the development and spread of other environmental policy instruments. He has published on the dynamics of policy instrument constituencies, the political duality of scales in environmental markets and first experiences with a new policy foresight method developed by his group.

Rajbeer Singh is a research fellow at Delft University of Technology and Eindhoven University of Technology. Presently, his research focuses on responsible innovation, innovation systems, renewable energy and energy-saving technologies such as LEDs. He is also engaged in delivering lectures in courses run by his colleagues. He has an interest in innovation, futures studies (foresight approaches), energy, sustainability, STI policies, social policy, technology management, lighting system, climate issues, IT and e-governance, food and agriculture, higher education, discrimination and exclusion, and poverty. He has published on these issues in international conference proceedings and journals.

Petra Sneijder is a postdoctoral researcher at the Strategic Communication Group, Department of Communication, Philosophy and Technology of Wageningen University, the Netherlands. Petra finished her Ph.D. thesis on discursive identities, food choice and eating practices in 2006. Currently, she is mainly interested in the application of discursive psychology in the field of food innovation. She studied the interactional and rhetorical dimensions of expert and consumer talk on food innovation and explores how—and with which consequences for the public agenda—both parties discuss citizen concerns.

Tsjalling Swierstra is full-time professor and head of the Philosophy Department of the University of Maastricht. He is a member of the Dutch Health Council and of the Program Committee of the 'Responsible Innovation' programme funded by the Dutch Research Council NWO. He has published widely on the ways in which ethical and political beliefs influence scientific and technological research and vice versa—how science and technology affect (moral) values, philosophies and political beliefs.

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.

Behnam Taebi is assistant professor in ethics of technology at Delft University of Technology and a research fellow at Harvard Kennedy School's Belfer Center for Science and International Affairs. He has worked and published on the societal and ethical aspects of energy systems. He is the leading editor of a volume on 'The Ethics of Nuclear Energy' (under contract with Cambridge University Press) and a special issue of Journal of Risk Research on 'Socio-technical Challenges of Nuclear Power Production'. He is currently involved in projects on 'Multinational nuclear waste repositories' and 'The acceptability of shale gas', both funded by the Netherlands Organization for Scientific Research.

Job Timmermans M.Sc/MA is a Ph.D. candidate on Responsible Research and Innovation in ICT at the De Montfort University (DMU). He is working within the Framework for Responsible Research and Innovation in ICT (Frriict) project led by Marina Jirotka from the University of Oxford (PI) and Bernd Stahl from DMU (CI). Job lectured engineering ethics and information ethics at Delft University of Technology. He was a member of the FP7 ETICA project (Ethical Issues of Emerging ICT Applications) and has organised conferences and workshops such as the annual NWO Conference on responsible innovation and a workshop on the Internet of Things (IoT).

Hedwig te Molder is personal professor at the Strategic Communication Group, Department of Communication, Philosophy and Technology of Wageningen University, the Netherlands. She is also full professor of Science Communication at the University of Twente. She has a special focus on expert–citizen interaction and food talk. She uses discursive psychology and conversation analysis to examine people's everyday moralities regarding science and technology issues and to develop reflection and intervention methods. In 2007, she received the Distinguished Book Award from the ASA for conversation and cognition (with Jonathan Potter).

Bart van Bezooijen holds a master's degree and a Ph.D. in psychology from Tilburg University. He was postdoctoral researcher at Eindhoven University of Technology as part of a NWO project on responsible innovation. He currently works at Fontys University of Applied Sciences.

Contributors

Annelies Balkema Technology, Innovation and Society, School of Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands

Bart van Bezooijen Philosophy and Ethics, School of Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands

Vincent Blok Management Studies Group, School of Social Sciences, Wageningen University, Wageningen, The Netherlands

Jacqueline Broerse Athena Institute, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, The Netherlands

Christine Boshuijzen-van Burken Philosophy and Ethics, School of Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands

Aad Correljé Faculty of Technology Policy and Management, University of Technology Delft, Delft, The Netherlands

Eefje Cuppen Faculty of Technology Policy and Management, University of Technology Delft, Delft, The Netherlands

John Danaher Department of Law, National University of Ireland, Galway, Ireland

Rudi Darson Rotterdam, The Netherlands

Sarah R. Davies Department of Media, Cognition and Communication, University of Copenhagen, Copenhagen, Denmark

Marloes Dignum Faculty of Technology Policy and Management, University of Technology Delft, Delft, The Netherlands

Rosanne Edelenbosch Athena Institute, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, The Netherlands

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

Nadira Faber Department Values, Technology, and Innovation, Section of Ethics/Philosophy of Technology, Delft University of Technology, Delft, The Netherlands ; Department of Experimental Psychology, University of Oxford, Oxford, UK,

Catherine Flick Centre for Computing and Social Responsibility, De Montfort University, Leicester, UK

Dirk Haen Maastricht University, Maastricht, The Netherlands

Maja Horst Department of Media, Cognition and Communication, University of Copenhagen, Copenhagen, Denmark

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

Marije de Jong Athena Institute, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, The Netherlands

Bert-Jaap Koops Tilburg University, TILT, Tilburg, The Netherlands

J. Otto Kroesen Faculty of Technology, Policy and Management, Technical University Delft, Delft, The Netherlands

Frank Kupper Athena Institute, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, The Netherlands

Pieter Lemmens Department of Philosophy and Science Studies, Radboud University, Nijmegen, The Netherlands

Hannah Maslen Uehiro Centre for Practical Ethics, Oxford, UK

Hedwig te Molder Wageningen University and Research Centre, Wageningen, The Netherlands; University of Twente, Enschede, The Netherlands

David J. Ndegwah School of Humanities and Social Sciences, Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

Quang A. Nguyen Centre for Computing and Social Responsibility, De Montfort University, Leicester, UK

Udo Pesch Faculty of Technology Policy and Management, University of Technology Delft, Delft, The Netherlands

Auke Pols Philosophy and Ethics, School of Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands

Zoë Robaey Department of Values, Technology and Innovation, Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands

Anneloes Roelofsen KWF Kankerbestrijding, Amsterdam, The Netherlands

Andri D. Setiawan Department of Industrial Engineering & Innovation Sciences, Eindhoven Centre for Innovation Studies (ECIS), TU Eindhoven, Eindhoven, The Netherlands

Arno Simons Institute for Ecological Economy Research, Berlin, Germany

Rajbeer Singh Department of Values and Technology, Faculty of Technology, Policy and Management, TU Delft, Delft, The Netherlands

Filippo Santoni de Sio Department of Values, Technology, and Innovation, Section of Ethics/Philosophy of Technology, Delft University of Technology, Delft, The Netherlands

Petra Sneijder Wageningen University and Research Centre, Wageningen, The Netherlands

Bernd Carsten Stahl Centre for Computing and Social Responsibility, De Montfort University, Leicester, UK

Tsjalling Swierstra Maastricht University, Maastricht, The Netherlands

Behnam Taebi Faculty of Technology Policy and Management, University of Technology Delft, Delft, The Netherlands; Belfer Center for Science and International Affairs, Kennedy School of Government, Harvard University, Cambridge, USA

Job Timmermans Centre for Computing and Social Responsibility, De Montfort University, Leicester, UK

Chapter 1
The Concepts, Approaches, and Applications of Responsible Innovation

An Introduction

Bert-Jaap Koops

Abstract 'Responsible innovation' is an increasingly popular term, but it is by no means clear what exactly this term refers to, nor how responsible innovation can or should be approached. This chapter provides an introduction to the landscape of responsible innovation, drawing from the contributions to this volume and an emerging body of literature. First, the concept of responsible innovation is explored: what does 'responsible innovation' refer to? The concept can be seen as an ideal, of incorporating social and ethical values or aspects in the innovation process, and as a project, a joint enterprise of an increasingly large community of people who want to bring us closer to this ideal. Next, approaches to responsible innovation are discussed: how can we go about innovating responsibly? While all approaches seem to have in common a key role for stakeholder engagement, one can distinguish two broad types of approaches to make innovation in a certain context more responsible. There is a product approach, characterised by a focus on developing some kind of output—a method, a framework, or guidelines; and a process approach, focused on developing some kind of procedure, usually with an element of self-learning. Subsequently, the current landscape of responsible innovation is briefly sketched: who is doing what in which areas? The chapter ends with explaining the structure of this edited volume and a brief tour through the chapters, which together provide a rich body of work that anticipates, reflects, deliberates, and responds to the challenges of responsible innovation.

Keywords Responsible research and innovation (RRI) · Responsible innovation · Definition · Stakeholder engagement · Reflexive learning · Product approach · Process approach

---

B.-J. Koops (✉)
Tilburg University, TILT, PO Box 90153,
5000 LE Tilburg, The Netherlands
e-mail: E.J.Koops@uvt.nl

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_1

1.1 Introduction

This is the second volume in a series of proceedings on *Responsible Innovation*, corresponding with a conference series organized by the Netherlands Council for Scientific Research (NWO) in The Hague since 2011. The conferences present results of research projects funded under the NWO research program "Responsible Innovation" (*Maatschappelijk Verantwoord Innoveren*).[1] At the same time, the conferences provide a platform for the broad and rapidly growing international community of researchers—inside and outside of academia—who are interested and involved in research and innovation projects that investigate or implement responsible innovation. This volume contains work both from NWO-funded projects and from international researchers, which was presented at the second Conference on Responsible Innovation that took place on 13–14 December 2012 in The Hague.

Although 'responsible innovation' is a term that is increasingly used both in academic and in policy circles, it is by no means clear what exactly the term refers to, nor how responsible innovation, once we know what is meant by this, can or should be approached. With a wide variety of contributions that discuss conceptualisations, approaches, and applications of responsible innovation, this volume contributes to a better understanding of what responsible innovation means and what this implies for the theory and practice of innovation. The Dutch research program and conference series offer a space for anticipating, reflecting, deliberating, and responding to responsible innovation, and thus can be seen as an example of responsible innovation in research and policy itself (cf. Owen et al. 2013a, b, p. 46).

Although the contributions in this volume can only provide one step forward in what is as yet an early stage of a growing international effort at responsible innovation, slowly the contours of responsible innovation are taking shape. The aim of this volume is not to discuss certain particular aspects of responsible innovation, nor to take a position on how responsible innovation can best be undertaken; rather, we aim to offer a broad overview of what is currently happening in the field of responsible innovation, both in conceptual thinking and in actual research practice, in all its varieties and forms. Thus, the volume can contribute both to the further conceptualisation and theory-building of responsible innovation and to mutual learning of current approaches and practices, illustrated by more or less successful case studies and concrete challenges that responsible innovation projects are facing.

In this introductory chapter, I will sketch the contours of the broad landscape of responsible innovation, drawing from the contributions to this volume and the emerging body of literature (Von Schomberg 2011; Owen et al. 2013a). First, the concept of responsible innovation is explored: what does 'responsible innovation' refer to? Next, some of the major approaches to responsible innovation are discussed: how can we go about innovating responsibly? This sets the stage for a brief sketch of the current landscape of responsible innovation: who is doing what

[1] See http://www.nwo.nl/en/research-and-results/programmes/responsible+innovation for a description of the program and of projects funded under this program.

in which areas? The chapter ends with explaining the structure of the book and a brief tour through the chapters.

1.2 What Is Responsible Innovation?

The concept of responsible innovation—sometimes also referred to as responsible research and innovation to emphasise the relevance of 'responsibility' also in fundamental research—is gaining currency. A detailed definition is given by Von Schomberg (2011, p. 9):

> 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).

At a more abstract level, Stahl et al. (2013, p. 214) define responsible research and innovation as 'a social construct or ascription that defines entities and relationships between them in such a way that the outcomes of research and innovation processes lead to socially desirable outcomes.' Owen et al. (2013a, b, p. 36) define responsible innovation as 'a collective commitment of care for the future through responsive stewardship of science and innovation in the present.' Blok and Lemmens (Chap. 2) describe responsible innovation as 'a new approach towards innovation, in which social and ethical aspects are explicitly taken into account (…) and economic, socio-cultural and environmental aspects are balanced'. Similarly, Setiawan and Singh (Chap. 12) provide a contextualized working definition of responsible innovation as 'ensuring the accountability of innovation actors (the actors involved in the adoption of innovation) through the engagement of anticipation, reflexivity, responsiveness, deliberation and participation in the adoption of innovation while looking the impact of innovation on three aspects: environment, social, and economy'.

In these definitions, the emphasis is on articulating what 'responsible' means in 'responsible innovation'; the element of 'innovation' in 'responsible innovation' is not elaborated. Although this could imply that it is taken for granted that we know what 'innovation' means—with a very wide body of literature on innovation, perhaps there is no need to pin down what exactly 'innovation' means, as we all seem to know what it refers to—this need not be the case. Owen et al. (2013a) start their volume with a lengthy chapter on innovation, before embarking on a discussion of what is responsible in innovation, an approach that is followed in the discussion of definitions in Chap. 12. Blok and Lemmens (Chap. 2) argue that often a too limited understanding of innovation is applied: 'future research should broaden our conception of innovation, including non-technological innovations and non-market environments.' I think that the lack of elaboration of 'innovation' in definitions of the term 'responsible innovation' does not primarily indicate an unproblematic understanding of innovation; rather, it suggests that responsible innovation literature can be seen as a sub-field of innovation. It does not primarily aim at understanding or improving innovation *as such*, but rather at understanding how innovation, whatever it means in different contexts, can be made ‘responsible’.

What, then, does ‘responsible’ mean? In the definitions above, it refers to being (ethically) acceptable, sustainable, socially desirable (see also Von Schomberg 2013, p. 64), leading to socially desirable outcomes (Stahl, Eden, and Jirotka), care for the future (Owen et al.), and taking account of social and ethical aspects and balancing economic, socio-cultural and environmental aspects (Blok and Lemmens, Setiawan and Singh). The emphasis here is on incorporating social and ethical values or aspects in the innovation process. In this respect, responsible innovation is a close relative of corporate social responsibility, with which it shares a strong family resemblance. The Dutch term for responsible innovation, ‘maatschappelijk verantwoord innoveren’ (literally: socially responsible innovation) is modelled on the Dutch term for corporate social responsibility, ‘maatschappelijk verantwoord ondernemen’ (literally: socially responsible entrepreneurship). Insofar as industry innovates through research and development, there is considerable common ground between the ambitions of corporate social responsibility and responsible innovation. If we want to investigate the concept of responsible innovation (cf. Chaps. 3 and 4), we should therefore take care not to restrict literature searches to the sole term ‘responsible innovation’; one could easily miss out relevant developments in adjacent areas, such as corporate social responsibility or other ‘social responsibility’ family members.

In a similar vein, we should be careful in acting as if responsible innovation is a new enterprise, lest we overlook lessons of the past. Responsible innovation builds on various areas that have known a very active and productive scholarship over the past decades. Most prominently (cf. Chap. 4), responsible innovation’s roots lie in various strands of Science, Technology, and Society Studies (STS), such as technology assessment, particularly the later generations of Constructive Technology Assessment and Participatory or Public Technology Assessment (for an overview, see Van Eindhoven 1997), and Value-Sensitive Design (Friedman 1998). It also has roots in applied ethics, often combined with STS approaches into a reflection on how new technologies, society, and values co-evolve (e.g., Bijker and Law 1992). There are also close parallels between the emergence of responsible innovation and of sustainable development, sharing a focus on value-sensitive and process-oriented guidance of processes of change (cf. Chaps. 11 and 15).

Less prominent but equally important are the roots of responsible innovation in legal theory, governance and regulation studies, in which the past decades have witnessed a ‘wider governance turn [that] is expressed through a move away from top-down, command and control regulation and towards a broader distribution of “soft law” activities’ (Chap. 3). The development of ‘smart regulation’ or ‘responsive regulation’ (Ayres and Braithwaite 1995) and ‘participatory governance’ (Fung and Wright 2003) shares many characteristics with developments in STS and applied ethics, such as a focus on an on-going and reflexive process of learning, and a ‘participatory turn’ of stakeholder involvement. Thus, regulatory innovation (Black et al. 2005) is a close relative of responsible (research and) innovation. Similarly, the rise of the study of 'techno-regulation', i.e., the use of technology as a regulatory tool (Lessig 1999; Brownsword and Yeung 2008), parallels the development of value-sensitive design, in an enterprise of embedding, in a responsible way, values and norms in the design of technology. Considerable experience with reflexive learning and stakeholder engagement has also been built up within the field of risk governance, within which risk assessment bears considerable similarities with Technology Assessment, and risk management has developed approaches fitting the paradigm of responsive regulation. The increasing emphasis on stakeholder engagement and public participation in risk governance (e.g., Thayyil 2014) likewise parallels the focus on stakeholder engagement in responsible innovation.

Perhaps the most interesting aspect of responsible innovation is that it builds on these various traditions, and tries to bring together the insights from different research traditions and communities, ranging from innovation studies, STS, ethics, and governance and regulation studies. Moreover, responsible innovation research is not conducted in ivory towers, but in labs and work spaces where innovation happens in practice; it brings together scientists, social scientists, and humanities scholars to jointly explore how research and innovation can be 'responsibly' shaped. And this need not, and should not, be restricted to technological innovation stricto sensu: since technology interacts with society and norms in a complex process of mutual shaping, responsible innovation is broadly relevant wherever society innovates, be it in technologies, institutions, social practices, or regulation. It is the combination of all these that should, ideally, be incorporated into responsible innovation.

Which leaves me with a final question: how can we best characterise the concept of responsible innovation that we are exploring? Is it just a 'concept' (Chap. 4), a 'notion' or a 'discourse' (Chap. 2), an 'approach' (Chap. 2) or a 'strategy' (Von Schomberg 2013, p. 51), or an 'ideal' or 'aspiration' (Owen et al. 2013a, b, p. 27)? Is it a new field of study or an emerging discipline, or a trend in scholarship and policy, or perhaps even a hype (Chap. 4)? Perhaps it is all of the above, but if we must choose, I think responsible innovation can best be characterised as a combination of two things. It is, first, an ideal: something we strive for even though we realise it can never be fully attained. Second, it is also a project, a joint enterprise of an increasingly large community of people who want to bring us closer to this ideal.

1.3 Approaches to Responsible Innovation

There are many ways to approach responsible innovation in practice. With roots in several different traditions, responsible innovation can draw from a variety of methods, practices, and lessons, as the chapters in part II (on approaches) and part III (on applications) attest. A single element stands out as a common factor within the variety of approaches: the engagement with stakeholders (Chaps. 4, 6–12, and 15). In line with the 'participatory turn' in the social sciences, responsible innovation researchers emphasise the importance of listening to stakeholders in innovation processes. They explore stakeholder opinions and views both substantively, as part of an effort of responsible innovation in certain contexts, such as cognitive enhancement (Chap. 7) or biofuels in Tanzania (Chap. 15), and at a meta-level, reflecting on what are productive methods to investigate and incorporate stakeholder views, such as focus groups (Chap. 8) or dialogue workshops (Chap. 9). The risk of seeing stakeholder engagement as a silver bullet in responsible innovation is pointed out by Blok and Lemmens, who emphasise that power asymmetries between stakeholder groups affect the framing of societal problems and the responsiveness and the 'response-*ability*' of actors in the innovation process, leading them to conclude that the 'practical applicability of the concept of responsible innovation is highly questionable' (Chap. 2). This contrasts sharply with the tone and findings of Correljé et al., whose framework sees public engagement as a symmetrical process giving equal attention to the project stakeholders and the local community ('the public'); according to them, 'symmetry is considered crucial for value sensitive design of energy projects' (Chap. 10).

Apart from stakeholder engagement, there are no substantial elements that seem to be common to all approaches to responsible innovation. Within the wide range of ways to deal with responsible innovation, two broad types of approaches can be distinguished: a product approach and a process approach. This distinction is similar but not identical to the two dimensions that Von Schomberg (2013, pp. 65–66) discerns. Von Schomberg's product dimension describes responsible innovation for products, which 'should be evaluated and designed with a view to (…) normative anchor points' (Von Schomberg 2013, p. 65). The process dimension describes responsible innovation as an attempt 'to arrive at a more responsive, adaptive, and integrated management of the innovation process' (Von Schomberg 2013, p. 65). These dimensions thus focus on the *object* of responsible innovation: either a (usually technological) product, or the process of innovation.

The product and process approaches that I discern relate more to the *approach* to responsible innovation: the enterprise of responsible innovation can be seen as a product (something that is developed and then used) or a process (something that is on-going and recursive). A product approach can be applied both to product innovation (e.g., a value-sensitive memory detection test, Chap. 13) and to process innovation (e.g., developing a normative framework to assess shifts in responsibilities associated with cognitive enhancement, Chap. 7). A process approach can be applied to product innovation (e.g., organising 'responsiveness' in food technology, Chap. 9) and to process innovation (e.g., developing a learning platform for innovation in ICT research and development, Chap. 6). There is no clear-cut border between these approaches. Approaches to responsible innovation lie on a spectrum with the product approach at one end and the process approach at the other end, and most responsible innovation projects lie somewhere in between. For analytical purposes, it may nevertheless be useful to articulate which end of the spectrum tends to get more attention.

The *product approach* can be characterised by a focus on developing a method, a framework, or guidelines that can be used to make innovation in a certain context more responsible. Often, it involves the development of a normative framework (consisting of ethical and legal values and norms) that is subsequently applied to a technology (concrete applications or a more abstract class of technology), and this often is accompanied by an argument that the normative framework should be applied from the start of the technology development process. Responsibility in innovation processes has to move 'upstream', and many projects aim at developing tools that actors at the source of the stream can use to take account of ethical and social values. Risk assessment methods and the precautionary principle (discussed in Chap. 5) are examples of such tools.

Although the value set of stakeholders is an important element of the normative framework, the product approach seems often also to incorporate values from external sources in the normative framework. For example, Robaey and Simons draw upon the classic principles in biomedical ethics to develop a set of higher values that is used to take a critical look at three current approaches to responsible innovation, and they argue that the value of justice is underrepresented in these approaches (Chap. 5). Danaher uses legal-systematic principles to develop a legitimacy-enhancing test that can be applied forensic applications such as lie detection (Chap. 13). Maslen, Santoni de Sio, and Faulmüller combine analysis of stakeholder values with legal and philosophical analysis in their application of the method of reflective equilibrium to contribute to the responsible development and use of cognitive enhancers such as psychopharmaceuticals (Chap. 7).

Moving towards the other end of the spectrum, the *process approach* can be characterised as a focus on developing self-learning procedures that can be used to make innovation in a certain context more responsible. In contrast to the product approach, the aim is less to develop substantively responsible frameworks or methods, but rather procedures or practices that are procedurally responsible. It is often associated with procedural values such as legitimacy, inclusiveness, and accountability, while the substantive values that guide a certain technology or system transition are generated internally to the context, through stakeholder involvement. This procedural, self-learning focus is most visible in Owen et al.'s approach to responsible innovation: 'to innovate responsibly entails a continuous commitment to be *anticipatory, reflective, inclusive, deliberative*, and *responsive*' (Owen et al. 2013a, b, p. 29, emphasis in original). Procedures developed along these lines can be found, for instance, in adaptive management (discussed in Chap. 5) and the Observatory for Responsible Research and Innovation in ICT, which fosters self-learning in the UK ICT research community with respect to responsible innovation practices (Chap. 6). The process approach is also well illustrated in Correljé et al.'s framework for responsible innovation in the energy sector, which reflects

> 'the dynamic nature of value specification through stakeholder interaction. This is an important notion for value sensitive design. The emergent and dynamic nature of value specification demonstrates that a full *ex ante* assessment of relevant and conflicting values is not possible. Designing for values requires a continuous and flexible participatory approach.' (Chap. 10)

Perhaps it is ironic that many projects that subscribe to a process vision of responsible innovation, with a focus on developing reflective, inclusive, deliberative, and responsive procedures, nevertheless have a tendency to adopt somewhat of a product approach to the procedures or methods they develop, in the sense that they hope to develop, ultimately, a good (reflective, responsive, etc.) procedure that works to achieve responsible innovation. Although reflection and self-learning are evident in the critical approach that researchers adopt towards what they are doing (see, e.g., Chap. 9), responsible innovation researchers do not always critically and recursively reflect on what they are doing themselves. Researchers operating in the value-sensitive design frame should, for example, carefully consider the critique by Blok and Lemmens that '[i]nnovation implies pain, annihilation and destruction, and it is questionable whether this "Faustian aspect" of innovation can ever be overcome by integrating social and ethical aspects in the design process' (Chap. 2).

It is important to realise that 'as an innovation itself, responsible innovation must abide by its own framework in this regard, and be anticipatory, reflective, deliberative, and responsive in its constitution and implementation' (Owen et al. 2013a, b, p. 46). Possibly, responsible innovation research and implementation projects should make a more consistent effort to internalise the very process approach that many researchers advocate as the best approach to responsible innovation. This implies that these research and implementation projects should also build in reflection and deliberation—with peers in the responsible innovation research community—in the design of their projects. Innovation projects, if they are to be responsible, should incorporate sensitivity to the dynamic character of the systems that are being innovated, by building in reflexivity and recursiveness in their tools, methods, procedures, and approaches.

1.4 The Landscape of Responsible Innovation

While the previous sections articulate the growing importance of and attention for responsible innovation, it remains to be seen to what extent the ideal of responsible innovation is brought any closer in practice by the collective project of responsible innovation. As Davies and Horst observe, '[w]e also found it easier to identify discussion of the principles of RI than examples of it being carried out in practice' (Chap. 3), and scientists seem as yet little involved in discussing responsible innovation in their own practice (Chap. 4). At this point in time, responsible innovation is, perhaps, largely an international, macro-level discourse (Chap. 3). But such a view would ignore the many efforts that *are* presently being undertaken to give meaning and substance to responsible innovation. Parts II and III of this volume attest to a significant number of applications in responsible innovation practices and projects. Gradually, the contours of the landscape of responsible innovation are taking shape. It is impossible to provide a detailed picture of this landscape here, both because of space constraints and because the contours are as yet fuzzy, but a charcoal sketch might look as follows.

The landscape of responsible innovation is very diverse, first and foremost because responsible innovation requires a context-sensitive approach. Any attempt at responsible innovation should therefore be carefully positioned along the relevant dimensions that constitute its context—and perhaps responsible innovation exists not only in our four-dimensional space/time world but also in the metaphorical ten-dimensional space that I have elsewhere sketched of technology regulation research (Koops 2010). The context depends not only on place and time, but also on the type and innovativeness of a technology, on the type and framing of the problem, the disciplines that are or can be used to approach the problem, the available knowledge, and the normative outlooks involved. Ideally, a context-specific approach would explicitly position responsible innovation along each of these dimensions.

In terms of technologies, a significant number of projects described in this volume are situated in the fields of energy, environment-related technologies, and neuroscience. This may well be a coincidence as the volume is just a snapshot of one meeting point (the December 2012 conference). It demonstrates how topical energy, food, and neuroscience are in the current innovation landscape, but it should be pointed out that responsible innovation in information and communication technologies (ICT) is being discussed extensively in other platforms, such as a Privacy by Design approach to ICT systems (Cavoukian 2010), while ICT, finance, and nanotechnologies have dedicated chapters in Owen et al.'s (2013a) overview. Thus, the picture in this volume of technology areas is not representative of the entire landscape of responsible innovation, which stretches from ICT to synthetic biology, from nano-medicine to biorobotics, from renewable energy to technology transfer, and beyond.

The same applies to the places where responsible innovation takes place. Many chapters in this volume describe projects based in the Netherlands, which is not surprising since the conference and associated research program have Dutch roots. But then, the Netherlands have a strong international position in STS and applied ethics, and it is not without reason that an innovative research program and conference series arise in this context. Notable is the connection that Dutch researchers have with developing countries in efforts of locally-sensitive responsible innovation projects (Chaps. 11, 12 and 15). Other countries featuring in this volume are the United Kingdom (Chaps. 3 and 6) and Denmark (Chap. 3), evidencing a European focal point in responsible innovation, which may partly be credited to the presence of policy entrepreneurs such as René von Schomberg. The United States, however, is also visible in the landscape (Chaps. 3, 13 and 14).

As to the actors populating the landscape, Davies and Horst provide a good overview:

> 'a wide range of persons, organisations, groups and categories appear as implicated in RI [responsible innovation] in some way. These actors include, for instance: "Europe", "the public", "governments, companies and research funders", "civil society", "business", "stakeholders", "NGOs", "research councils", "investors", "citizens", and "researchers".' (Chap. 3)

In line with the emphasis on stakeholder engagement, the main characters featuring in many of the chapters are those most directly involved in the innovation process: developers (scientists, researchers) and users (consumers, professional users, 'the public'). Less visible, however, are important stakeholders who are not involved hands-on in the process or product itself that is being innovated, but who affect the innovation trajectory more indirectly though not less forcibly: research funders (both public and private), corporate leaders (corporate innovation seems underrepresented in current research, according to Chap. 2), and regulators (politicians, law-making officials, supervisory authorities).

The relative absence of regulators in the landscape might imply that the governance of responsible innovation is as yet underdeveloped, and perhaps that processes of responsible innovation are to a significant extent self-governed. But absence of regulators does not necessarily imply absence of governance structures besides self-regulation: most innovation takes place within existing regulatory frameworks, for example norms and institutions for health and safety, intellectual property, liability, and standardisation. Nevertheless, existing governance structures and regulatory frameworks can sometimes be poorly equipped to accommodate certain innovations or system transitions. For example, regulation of the energy market based on the tradition energy market has to be transformed to deal with the transition to smart grids (cf. Bellantuono 2012); regulation of chemicals based on quantity thresholds is not well-suited to govern nano-materials that due to different properties may also be toxic in small quantities (Fleurke and Somsen 2011: 372); and the incentive structure embedded in liability law might impede certain innovations in robotics due to the unpredictability, and possible lack of insurability, of autonomic robotic applications (cf. Bertolini 2013). This underlines the importance of integrating the governance perspective within the whole cycle of responsible innovation processes, so that technologies, practices, and governance can fruitfully co-evolve.

One possible explanation for the relative lack of visible interaction of regulators with responsible innovation processes can be found in Collingridge's dilemma: controlling a technology is difficult in its early stages because not enough is known of its possible or probable effects, and it is also difficult once the technology is well-developed because by then intervention is expensive and drastic (Collingridge 1980). Possibly, regulators are verging towards regulating technologies more downstream, when at least they have better knowledge of the technology and its effects, instead of exploring regulatory interventions upstream where responsible innovation is situated. Seeing, however, that regulatory innovation does take place, for example in Privacy by Design requirements in data protection law (European Commission 2012: Art. 23), model codes of conduct for responsible nanotechnologies (European Commission 2008), and a reflexive and cyclical approach in European chemicals regulation (Fleurke and Somsen 2011), there may also be another explanation for the lack of visibility of governance in the responsible innovation landscape. Possibly, responsible innovation and regulatory innovation, although close cousins, are as yet shaped within different communities, forming two cultures that need to get better acquainted. Hopefully, in the coming generation of responsible innovation research and practice, regulators—and other indirect but powerful stakeholders—will also be involved more visibly in the enterprise of responsible innovation.

The landscape is coloured by the values and normative outlooks that constitute the 'responsible' element in responsible innovation contexts. Which values and outlooks are relevant is highly context-dependent, not only depending on the socio-cultural location but also on the technology and type of problem that are being studied. As observed above, higher-level values that colour the landscape more strongly are often procedural values, perhaps because the substantive values are too context-specific to show in depth, or because innovation is radically unpredictable (Chap. 2) and hence can only be approached in terms of procedural fairness and not substantive fairness. It is interesting to note that where substantive values do show through in this volume, they are often centred on autonomy and a fair (re)distribution of responsibilities. This not only repeats a long-standing concern in discussions on new technologies that they might (typically negatively) affect the autonomy of individuals, but it might also reflect a possible step-change in the role of technology in society: with fast-growing developments in robotics, we are now witnessing the emergence of technologies that can, and perhaps will, function really independently from human agency (cf. the discussion of drones in Chap. 14). The concern over autonomous technologies has of course a respectable tradition in science fiction, philosophy, and future studies, but so far the prospect of technologies going 'their own way' has been essentially speculative. With emerging robotics applications, we may have to negotiate our relationship with technology in new ways, moving from craftsman-tool and subject-object relationships toward relationships between autonomous subjects (cf. Matthias 2007).

1.5 Overview of This Volume

We have organised this volume in three parts. Part I contains chapters that primarily attempt to conceptualise responsible innovation. These chapters discuss, at a meta-level, the concept(s) that are used in policy and academia discourse and practice. In Chap. 2, Vincent Blok and Pieter Lemmens provide a counterpoint to the dominant narrative of responsible innovation, arguing that this narrative misunderstands the nature of innovation, raising significant doubts on the practical feasibility of 'responsible innovation'. A more critical inquiry into the concept is needed, with a less naïve understanding of the process of innovation, a wider understanding of innovation that also covers institutions and social practices, and a study of alternative strategies of innovation in contexts that may be more amenable to 'responsibilisation' than mainstream, often power-imbalanced, innovation contexts. In Chap. 3, Sarah R. Davies and Maja Horst sketch the policy landscape of responsible innovation in the United States, the United Kingdom, and Denmark, asking how the concept is being constituted in policy and governance discourse. A descriptive map of the landscape is followed by a critical reflection, which finds responsible innovation currently to be a general and macro-level process, associated with de-individualisation of responsible actors, and the authors call for more research and initiatives at the micro-level where responsible innovation should be operationalised. In Chap. 4, Marije de Jong and colleagues investigate whether and to what extent researchers in neuroimaging are familiar with the notion of responsible innovation. As the results show a low level of awareness with hands-on researchers, the authors argue that, responsible innovation being a co-constructive process, extra efforts should be made to actively engage scientists in this enterprise.

Parts II and III contain chapters that discuss approaches to and applications of responsible innovation. Most chapters contain a combination of an approach—a method, tool, or framework to practice responsible innovation—and an application context, such as food technology or solar energy. The focus in some chapters tends to be more on the approach, while other chapters seem to focus more on the application context. We have therefore arranged the chapters in a part on approaches to responsible innovation (Part II), where the discussion tends to be particularly concerned with learning about the process of responsible innovation itself, and a part on applications of responsible innovation (Part III), where the discussion tends to have more the character of case studies of how responsible innovation is, can be, or should be approached in concrete application contexts. The distinction is an analytic one that may help to guide the reader through the landscape of responsible innovation, going roughly downhill from more abstract to more concrete levels of discussion, but the reader should bear in mind that all chapters in these parts contain elements of approaches as well as applications.

Part II starts with Chap. 5, in which Zoë Robaey and Arno Simons discuss responsible innovation conceptualised as a process of social experimentation. They critically discuss three types of policy approaches featuring in the literature to deal with introducing new technologies in societies: the precautionary principle, participatory technology assessment, and adaptive management. An ethical analysis of these demonstrates that key ethical values are present to a greater or lesser extent in these approaches, and the authors plea for policy-makers to take more account of the value of justice. Chapter 6 presents results by Bernd Carsten Stahl and colleagues from a UK-funded project to develop a Framework for Responsible Research and Innovation in ICT. By developing an Observatory that serves as a platform for ICT scholars who face responsibility-related questions, the project contributes to sharing of experiences and good practices and serves as a community-based resource for responsible ICT innovation. A workshop reflecting on the idea of an observatory suggested that such a platform is most likely to be successful if it connects to issues that ICT researchers identify with, and thus is best built bottom-up. Hannah Maslen, Filippo Santoni de Sio, and Nadira Faulmüller, in Chap. 7, take up the theme of responsibilities in their discussion of psychopharmaceuticals used as cognitive enhancers. The authors outline a theoretical framework and methodology for investigating the claims made in the current enhancement debate that some professionals (such as surgeons or aircraft pilots) have a responsibility to enhance, and the claim that this might lead these professionals to acquire more responsibilities once enhanced. By examining normative hypotheses, psychological data from valorisation groups and lay people, and legal analysis of liability, the authors show how research into the implications of psychopharmaceuticals can be made sensitive to the evolving landscape of public attitudes and professional duties, and at the same time detail the current state of affairs required for immediate policy-making.

Continuing the theme of neuroscience developments, in Chap. 8, Rosanne Edelenbosch, Frank Kupper, and Jacqueline Broerse discuss how focus groups can be used to elicit stakeholder views on the opportunities and concerns of emerging technologies, in this case neuroimaging applications in the context of education and personalised learning. The focus groups show not only that value conflicts emerge, but also that arguments about opportunities and concerns are influenced by the frames in which situations or issues are presented. Frame analysis thus provides a useful addition to the method of focus groups in stakeholder engagement exercises. In Chap. 9, Dirk Haen and colleagues describe how they developed a 'Discursive Awareness & Techno-Ethical Imagination' tool which can be used to organise stakeholder dialogues that are hospitable to moral, political, and cultural concerns around technological innovations. The authors critically reflect on the possibilities and limitations of this tool, based on their experience in two stakeholder dialogues on emerging food technologies.

Chapter 10 presents a framework developed by Aad Correljé and colleagues that can be used for responsible innovation in energy projects with a local site impact. The framework is based on Value-Sensitive Design, but applied not (only) to energy technologies but particularly (also) to the institutional design and the set-up of stakeholder interactions. The framework helps to elicit and accommodate the variety of values that is involved in the design of energy projects, and thus may prevent conflicts from arising or escalating between stakeholder groups due to controversies over their value sets. In Chap. 11, Otto J. Kroesen, Rudi Darson, and David J. Ndegwah argue that the development debate and research should pay more attention to the moral aspects of capacity building as part of responsible innovation. Illustrated by two case studies—the introduction of tropical greenhouses in Kenya and the renovation of a vocational school in Surinam—the authors show that different capacities are required in different stages of change trajectories. A responsible innovation approach therefore needs to be flexible, learning, and responsive as a project evolves, so that different values and cultural traits, and the capacities required to observe these, can be accommodated according to time, situation, and need.

Part III offers a set of experiences in concrete applications of responsible innovation, evidencing the growing range of attempts to bring responsible innovation to fruition in everyday practices. The first practice is the adoption of solar photovoltaic technologies in telecommunications towers in Indonesia, described by Andri D. Setiawan and Rajbeer Singh in Chap. 12. The case illustrates the importance and complexities of distributing responsibilities in technological innovation processes. Through stakeholder and impact analysis and applying the process-oriented dimensions of responsible innovation, the authors show how a proportional distribution of accountability and responsibility among stakeholders contributes to a socially acceptable and sustainable innovation. In Chap. 13, John Danaher presents a framework for responsible innovation in neurotechnological applications that are being developed for memory detection (or lie detection) in a forensic context. The framework consists in a legitimacy enhancing test, which results in finding an application acceptable if it would (probably) enhance the legitimacy of a courtroom procedure. The author claims that the test may be usable more widely, to foster responsible innovation in any social epistemic system.

In Chap. 14, Christine Boshuijzen-van Burken and Bart van Bezooijen discuss the case of Unmanned Aerial Vehicles in an attempt to understand how soldiers deal with innovations on the battlefield. Within the conceptual framework of normative practices, decision-making by soldiers is investigated through structural analysis and using findings from psychological research that reveal potential changes in behavioural patterns of people exposed to morally laden situations and technology-pervaded situations. These insights may inform designers to help them develop technologies that support responsible decision-making on the battlefield, taking into account both psychological mechanisms and the specific normative context in which decisions are taken. Finally, Annelies Balkema and Auke Pols describe in Chap. 15 how several of the themes discussed in earlier chapters crop up in their case study of jatropha (a plant used for biofuel) cultivation in Tanzania and its negative socio-economic and environmental impacts. By engaging with the most vulnerable group of stakeholders—small farmers in the South—the authors show that biofuel innovation has so far not taken sufficient account of societal and ethical values. They develop a normative framework grounded in sustainability and moral responsibility that can be used to reflect on innovations in biofuels.

Where the conceptual and more reflective chapters on approaches tend to emphasise the challenges of responsible innovation and suggest that the theory—although also under construction and not unproblematic in itself—is easier than the practice, the chapters on applications in responsible innovation show a vibrant practice of responsible innovation, however difficult it may be. Taken together, the body of work in this volume anticipates, reflects, deliberates, and responds to the challenges of responsible innovation. The challenges may be formidable, but the context-sensitive and case-based approaches that pervade this volume, and the many examples of more or less successful attempts at responsible innovation discussed, hold promise to take up those challenges and ultimately to make a difference in how innovation is shaped. As a Chinese saying holds: many small people doing many small things in many small places, can change the face of the world.

References

Ayres, I., and J. Braithwaite. 1995. *Responsive regulation: transcending the deregulation debate*. Oxford: Oxford University Press.

Bellantuono, G. 2012. Comparing Regulatory innovations for climate change: smart grids policies in the USA and the EU. SSRN eJournal, http://ssrn.com/abstract=2153881.

Bertolini, A. 2013. Robots as products: the case for a realistic analysis of robotic applications and liability rules. *Law, Innovation and Technology* 5: 214–247.

Bijker, W.E., and J. Law (eds.). 1992. *Shaping technology/building society. Studies in sociotechnical change*. Cambridge (MA): MIT Press.

Black, J., M. Lodge, and M. Thatcher, eds. 2005. *Regulatory innovation*. Cheltenham: Edward Elgar.

Brownsword, R., and K. Yeung (eds.). 2008. *Regulating technologies*. Oxford: Hart Publishing.

Cavoukian, A. 2010. Privacy by design: the definitive workshop. A foreword. *Identity in the Information Society* 3: 247–251.

Collingridge, D. 1980. *The social control of technology*. London: Frances Pinter.

European Commission. 2008. *Commission Recommendation of 07/02/2008 on a code of conduct for responsible nanosciences and nanotechnologies research*. Brussels, 7 February 2008, C(2008) 424final.

European Commission. 2012. *Proposal for a general data protection regulation*. Brussels, 25.1.2012, COM(2012) 11 final.

Fleurke, F.M., and Somsen, H. 2011. Precautionary regulation of chemical risk: How REACH confronts the regulatory challenges of scale, uncertainty, complexity and innovation. *Common Market Law Review* 48(2): 357–393.

Friedman, B., ed. 1998. *Human values and the design of computer technology*. University of Chicago Press.

Fung, A., and E.O. Wright (eds.). 2003. *Deepening democracy. Institutional innovations in empowered participatory governance*. New York: Verso.

Koops, B.J. 2010. Ten dimensions of technology regulation. Finding your bearings in the research space of an emerging discipline. In *Dimensions of technology regulation*, eds M.E.A. Goodwin, B.J. Koops and R.E. Leenes, 309–324. Nijmegen: Wolf Legal Publishers.

Lessig, L. 1999. *Code and other laws of cyberspace*. New York: Basic Books.

Matthias, A. 2007. *Automaten als Träger von Rechten. Plädoyer für eine Gesetzänderung*. PhD thesis, Humboldt University, Berlin.

Owen, R., J. Bessant, and M. Heintz (eds.). 2013a. *Responsible innovation*. Chichester: Wiley.

Owen, R., J. Stilgoe, Ph Machaghten, M. Gorman, E. Fisher, and D. Guston. 2013b. A framework for responsible innovation. In *Responsible innovation*, ed. R. Owen, J. Bessant, and M. Heintz, 27–50. Chichester: Wiley.

Stahl, B.C., G. Eden, and M. Jirotka. 2013. Responsible research and innovation in information and communication technology: identifying and engaging with the ethical implications of ICTs. A framework for responsible innovation. In *Responsible innovation*, eds. R. Owen, J. Bessant, and M. Heintz, 199–218. Chichester: Wiley.

Thayyil, N. 2014. *Biotechnology regulation and GMOs. Law, technology and public contestations in Europe*. Cheltenham: Edward Elgar.

Van Eindhoven, J.C.M. 1997. Technology assessment: product or process? *Technological Forecasting and Social Change* 54: 269–286.

Von Schomberg, R. 2011. *Towards responsible research and innovation in the information and communication technologies and security technologies fields*. Brussels: European Commission.

Von Schomberg, R. 2013. A vision of responsible research and innovation. In *Responsible innovation*, ed. R. Owen, J. Bessant, and M. Heintz, 51–74. Chichester: Wiley.

Part I
Conceptualisation—Understanding Responsible Innovation

Chapter 2
The Emerging Concept of Responsible Innovation. Three Reasons Why It Is Questionable and Calls for a Radical Transformation of the Concept of Innovation

Vincent Blok and Pieter Lemmens

Abstract In this chapter, we challenge the presupposed concept of innovation in the responsible innovation literature. As a first step, we raise several questions with regard to the possibility of 'responsible' innovation and point at several difficulties which undermine the supposedly responsible character of innovation processes, based on an analysis of the input, throughput and output of innovation processes. It becomes clear that the practical applicability of the concept of responsible innovation is highly problematic and that a more thorough inquiry of the concept is required. As a second step, we analyze the concept of innovation which is self-evidently presupposed in current literature on responsible innovation. It becomes clear that innovation is self-evidently seen as (1) *technological* innovation, (2) is primarily perceived from an *economic* perspective, (3) is *inherently good* and (4) presupposes a symmetry between moral agents and moral addressees. By challenging this narrow and uncritical concept of innovation, we contribute to a second round of theorizing about the concept and provide a research agenda for future research in order to enhance a less naïve concept of responsible innovation.

Keywords Responsible innovation · Responsible research and innovation · Innovation management · Science and technology studies · Engineering ethics

V. Blok
Management Studies Group, School of Social Sciences, Wageningen University, Hollandseweg 1, 6707 KN, Wageningen, The Netherlands
e-mail: vincent.blok@wur.nl

P. Lemmens (✉)
Department of Philosophy and Science Studies, Radboud University, Nijmegen, The Netherlands
e-mail: p.lemmen@science.ru.nl

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_2

2.1 Introduction

The concept of 'responsible innovation' is relatively new. The use of the term suggests that over the past decades, innovation wasn't all that responsible; the negative impact of innovations on individuals, societies and eco-systems was largely neglected in favour of economic growth and creating shareholder value. The emergence of responsible innovation has to be understood, then, as a new approach towards innovation, in which social and ethical aspects are explicitly taken into account (cf. European Commission 2011) and economic, socio-cultural and environmental aspects are balanced.

Because the primary responsibility for economic, socio-cultural and environmental aspects is allocated to different players in society—the profit sector on the one hand and governmental organizations, NGO's and civil society on the other—it is argued that the balancing of *People, Planet* and *Profit* in sustainable business development presupposes the active involvement of and partnership with various elements of society (Hens and Nath 2003; Eweje 2007). These partnerships are also crucial from an innovation perspective per se. Nowadays, it is widely acknowledged that only a few firms have all resources and networks available to innovate in isolation (Ireland et al. 2002). Innovation seems to flourish in an open innovation environment, in which the interaction with various stakeholders is seen as a resource of competitive advantage (Chesbrough 2003). From this perspective, it can be expected that cross-sector partnerships with multiple stakeholders will lead to innovative *and* responsible solutions. Stakeholder engagement seems to be a key characteristic of responsible innovation (cf. Chap. 1, Koops 2015) and is understood then as "a strategy of stakeholders to become mutually responsive to each other and anticipate research and innovation outcomes underpinning the 'grand challenges' of our time for which they share responsibility" (von Schomberg 2013); the ongoing involvement of society in innovation processes will help to achieve social and environmental benefits (cf. Matter 2011).

The guiding assumption here is that "[r]ight from the start, research, development and design [can] incorporate relevant ethical and societal aspects" so that "technological and scientific advances become properly embedded in society" (NWO 2012). Responsibility is seen here as an *add-on* or extension to the concept of innovation; responsible innovation = regular innovation + stakeholder involvement with regard to ethical and societal aspects. With the help of this extension, innovation processes will be better enabled to balance economic (profit), socio-cultural (people) and environmental (planet) interests.

Although the concept of responsible innovation is relatively new and still evolving in different directions, we think the time is ripe to challenge the presupposed concept of innovation in the responsible innovation literature. Till now, most research is done from a policy or socio-ethical perspective and focusing on academic R&D environments, while most innovations take place in commercial or industrial settings (cf. Flipse 2012). It is precisely corporate innovation, which is underrepresented in current research (cf. Penders et al. 2009). An additional problem is that empirical evidence how to put responsible innovation into practice is still scarce (Chap. 3, Davies and Horst 2015; Blok et al. 2015).

In this chapter, we primarily attempt to contribute to the conceptualization of responsible innovation. Based on a review of the existing literature on responsible innovation, we raise questions concerning the possibility of responsible innovation and point at difficulties with regard to the input, throughput and output of innovation processes. It will become clear that these difficulties can even undermine the responsible character of innovation processes (Sect. 2.2). In Sect. 2.3, we ask what concept of innovation is presupposed in responsible innovation. It will become clear that the presupposed concept of innovation is uncritical and cannot be *upgraded* to a more responsible concept of innovation; responsible innovation calls for a radical transformation of the concept of innovation itself. By challenging the presupposed concept of innovation in this article, we contribute to a second round of theorizing about the concept in order to enhance a less naïve concept of responsible innovation (Sect. 2.4).

2.2 Challenging the Concept of Responsible Innovation

2.2.1 The Input of Responsible Innovation Processes: Fundamental Differences and Conflicts of Interest Among Stakeholders in Case of 'Grand Challenges'

According to much of the responsible innovation literature, the *input* of responsible innovation processes is not to be found in clear-cut and isolated problems which have to be solved, but in the so called 'grand challenges' of our time; climate change, resource depletion, poverty alleviation, ageing societies, etc. (von Schomberg 2013). This shift towards grand challenges as the main driver for innovation is also reflected in the most important framework program for research and innovation in the EU; Horizon 2020 prioritizes research and innovation based on these grand challenges (cf. European Commission 2011).

Grand challenges like global warming or sustainable development are also called 'wicked problems'. According to Rittel and Webber, who described the concept of wicked problems as opposed to tamed problems for the first time in 1973, the former are difficult to pin down, highly complex and not amenable for definitive solutions. Wicked problems concern complex systems in which cause and effect relations are uncertain or unknown. Rittel and Webber have specified ten characteristics of wicked problems. Examples are that there is no definitive formulation of a wicked problem, that solutions to wicked problems are not true-or-false but better or worse and, that they have no stopping rule, i.e. the problem-solver does not know when an acceptable solution of the problem is found etc. (Ritter and Webber 1973; Batie 2008).

The complexity of wicked problems is partly related to the multiple stakeholders involved in solving these problems. Many stakeholders have different ideas about what the 'real' problem is (Kreuter et al. 2004). Also, the solution to the problem is based on judgments of multiple stakeholders, which can differ widely and are not (always) based on shared values (Batie 2008; Ritter and Webber 1973). With regard to responsible innovation, various stakeholders have different ideas about the problem and its solution in general and about the societal and ethical aspects which have to be taken into account during the innovation process in particular (cf. Chap. 11, Kroesen et al. 2015). These differences among stakeholders are due to differences with regard to the *content* of the grand challenges, but also due to different agendas and divergent motives of profit and non-profit organizations for instance (Yaziji and Doh 2009). While non-profit organizations are mainly motivated by altruistic motives for instance (Milne et al. 1996), profit organizations are mainly self-interested (Iyer 2003). Furthermore, profit and non-profit organizations have divergent approaches to value creation; companies will naturally focus on economic value creation by producing and selling products and services, while NGOs for instance will focus on social value creation by advocating social norms and values (Yaziji and Doh 2009; cf. Bos et al. 2013).

Because of these differences between various stakeholders, actual efforts to involve stakeholders in innovation processes are liable to failure. The fundamental differences among stakeholders with regard to their vision, goal, sector and motive, can be seen as bottlenecks in responsible innovation. These bottlenecks at least indicate that it is not so easy to 'incorporate relevant ethical and societal aspects' so that 'technological and scientific advances become properly embedded in society', as is sometimes suggested in the literature (cf. Chap. 10, Correljé et al. 2015).

According to the collaboration and partnership literature, an important condition for stakeholder involvement is the initial agreement among stakeholders on the problem definition and the goals of collaboration (Selsky et al. 2005; Bryson et al. 2006). Such a common ground may be found in a process of public dialogue about the priorities, directions, implications and consequences of innovations (Jackson et al. 2005). Hardy et al. (2005) pointed at political processes involved in defining the problem and the objective of collaboration. Political processes are important, because the specific formulation of the problem definition already determines what potential innovative solutions are sought for and who are legitimate partners in realizing these innovations. If we conceive a grand challenge like sustainable development for instance in such a way that it affords a systems change, a wholly different set of innovations is at stake than if it is defined at a product level and only involves innovations in order to substitute depletable resources. This example shows that stakeholders have an interest in the specific way the problem is defined, because it has consequences for the shared objective of the innovation process and with this, for the investment of resources to solve the problem.

It is clear that some actors are more powerful than others in defining the problem and the objectives of the innovation process, i.e. that the involvement of societal and ethical aspects is not without any 'push and pull'. It is presumable that power imbalances are especially at stake in the case of grand challenges, exactly because of the different problem definitions and different value frames of the stakeholders involved. These power imbalances can be seen as a prime source of conflicts among stakeholders (Bryson et al. 2006).

The first reason to question the possibility of responsible innovation is therefore, that there is no consensus about the scope of the grand challenges and the goal of the innovation process among stakeholders. It is not only questionable whether responsible innovation is possible in case of grand challenges or wicked problems. Fundamental power imbalances among stakeholders can even undermine the incorporation of societal and ethical aspects in the innovation processes. When the *input* of responsible innovation processes is found in grand challenges, we may conclude, the presupposed responsiveness towards stakeholders is highly questionable.

2.2.2 The Throughput of Innovation Processes: Transparency and Mutuality Among Stakeholders Is Limited and Does not Make Innovation Processes More Manageable

With regard to the *throughput* of responsible innovation processes, it is acknowledged that social and ethical aspects are usually *not* included in the innovation process (cf. Flipse 2012). In order to develop a more responsible model of innovation, various technology assessment approaches have been developed (see Flipse 2012 for an overview). In these approaches, it is assumed that stakeholders should be involved 'right from the start' in order to incorporate relevant ethical and societal aspects in the innovation process (NWO 2012; Delgado et al. 2010; Owen and Goldberg 2010). Von Schomberg for instance defines responsible innovation as a "transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to 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)" (von Schomberg 2013, p. 19). The throughput of the innovation process should be characterized by transparency, interaction and mutual responsiveness. Also in the collaboration and partnership literature, the necessity to reduce so called information asymmetries is acknowledged; by "linking and sharing of information, resources, activities, and capabilities", it is expected that partnerships between companies and their stakeholders "jointly achieve an outcome that could not be achieved by organizations in one sector alone" (Bryson et al. 2006, p. 44).

Apart from the fact that the fundamental deviations in power, vision, goal, sector and motive will limit the possibility of the mutual responsiveness among stakeholders, it is questionable whether finding a common ground is desirable (Roelofsen et al. 2011). If too much emphasis is placed on finding such a common ground, certain stakeholders will become hesitant to participate in the collaboration (Hagendijk and Irwin 2006, cited in Roelofsen et al. 2011). Moreover, the call for transparency of innovation processes is highly naive. From a business perspective, innovation is the main source of competitive advantage in the current economy. This advantage is exactly based on information asymmetries, i.e. additional knowledge which enables companies to identify business opportunities in the market. This additional or 'prior' knowledge (McMullen and Shepard 2006) may consist in the ability to "see where a good can be sold at a price higher than that for which it can be bought" (Kirzner 1973). In this case, information asymmetries are due to imperfect knowledge of market participants with regard to *existing* information and new business opportunities "arise out of the entrepreneur's alertness to [these] information asymmetries existing in the economy" (Dutta and Crossan 2005). Information asymmetries may also be *created* by the development of new information or new knowledge. This information provides opportunities for new or alternative solutions for existing or anticipated problems, and in case of responsible innovation, for existing and anticipated grand challenges. From the perspective of a company or a consortium of collaborating companies,[1] therefore, information asymmetries have to be seen as a source of competitive advantage. Although transparency towards stakeholders is a necessary condition of open innovation processes, the call for a mutual responsiveness among stakeholders—i.e. the reduction of information asymmetries—in the responsible innovation literature is highly naive. For this reason, collaborations with stakeholders are sometimes explicitly restricted, especially in case of intellectual property (IP) and secrecy (Flipse 2012).

The full transparency to stakeholders—especially with regard to the core technology or the innovation itself—is not only highly riskful from a business perspective. Innovative companies even have an incentive to increase information asymmetries because it enables them to claim capabilities or features of their new products or services which are not (yet) justified, in order to receive economic (i.e. investment) or societal (i.e. societal or governmental) support (Millar et al. 2012).

The call for *mutual* responsiveness and *collective* responsibility is also unrealistic. Innovations are risky and involve high amounts of investment. The societal and ethical acceptability of an innovation can be seen as an important investment criterion and in this respect, stakeholder involvement is indeed an important indicator of the societal embeddedness of the innovation process. Nevertheless, the investor alone is responsible for the risk-reward assessment and therefore, for the investment decision as such. The mutuality of the responsiveness is also limited by stakeholders like NGO's. Is it reasonable to expect that stakeholders cooperate constructively and live up to the commitments they make in the innovation process, when faced with an innovation process which is highly uncertain and with final impacts which are unpredictable? In the end, NGO's will never take

[1]On the one hand, it is assumed that the reduction of information asymmetries among partners will increase performance; by leveraging their resources, knowledge and capabilities—saving resources, elimination or reduction of waste, improving productivity etc.—inter-organizational partnerships may contribute both to the cost efficiency and the competitive advantage of the allied partners over other firms (Gulati 2007). On the other hand, this competitive advantage of the allied partners over other firms is based on increased information asymmetries.

responsibility for an innovation whose outcomes are so much uncertain. A further problem that occurs in the case of mutual responsiveness and collective responsibility is the blurring of tasks and responsibilities. This can lead to the loss of legitimacy; if an NGO collaborates with a company in responsible innovation and the innovation turns out not to be that responsible, they are nevertheless responsible as well. Instead of putting companies under pressure to innovate in a sustainable way, they could be accused of co-operation with their "traditional enemy" (Hemmati 2002; Van Huijstee et al. 2007).

We can even question whether transparency and mutual responsiveness during the innovation process will increase the desirability of the outcomes. Any purportted attempt to steer the development of technologies in directions that will maximize their social benefits and minimize their negative social and environmental impacts sees itself confronted with the so-called "dilemma of control" or "Collingridge dilemma". David Collingridge, chemist and analyst of technology policy, gave this dilemma its classic formulation: "The social consequences of a technology cannot be predicted early in the life of the technology. By the time undesirable consequences are discovered, however, the technology is often so much part of the whole economic and social fabric that its control is extremely difficult. This is the *dilemma of control*. When change is easy, the need for it cannot be foreseen; when the need for change is apparent, change has become expensive, difficult and time consuming" (Collingridge 1981, p. 11). In other words: In the early stages, when the technology is still malleable and thus amenable to social intervention, its effects are not yet sufficiently known to warrant such an intervention; when later on the effects become apparent, however, it is no longer easy to control or adapt the technology because it has meanwhile become 'hardened' and socially 'entrenched' (cf. Flipse 2012). Thus the dilemma results from the combination of an *information* problem at the earlier stages and a *power* problem at the later stages of development.

The Collingridge dilemma is widely known in technology policy circles and presents a clear challenge to anyone who aims to feed social concerns about possible negative consequences of a new technology back into the design and innovation process itself.

The second reason to question the possibility of responsible innovation is that the 'transparency' and 'mutuality' among stakeholders is limited. The Collinridge dilemma showed that transparency and interaction with multiple stakeholders doesn't make innovation processes more manageable perse. It is not only questionable whether responsible innovation is possible in the case of grand challenges or wicked problems. Information asymmetries among stakeholders can even undermine the responsible innovation; it enables firms to deploy these asymmetries in support of misleading claims about the features or capabilities of responsible innovations under construction, in order to attract societal and ethical legitimacy (Millar et al. 2012). In conclusion: when the throughput of responsible innovation processes is characterized by information asymmetries, the presupposed mutual responsiveness towards stakeholders becomes questionable.

2.2.3 *The Output of Responsible Innovation Processes: The Foresight of Responsible Innovation Is Limited Because of Our Epistemic Insufficiency with Regard to the Grand Challenges*

With regard to the *output* of responsible innovation processes, it is assumed that the incorporation of societal and ethical aspects in the innovation process decreases unintended societal impact of technological developments and will therefore lead to more responsible innovations (cf. Rogers-Hayden and Pidgeon 2007). Although stakeholder involvement may result in more desirable outcomes, the involvement of stakeholders doesn't necessarily guarantee a more responsible output of the innovation process. For example, many stakeholders were involved during the development of biofuels and this innovation seemed to be promising according to various stakeholders; biofuel is inherently renewable, locally produced, less polluting, etc. An unexpected outcome of the increased demand for biofuels however, was that food prices increased because farmers started to grow more and more crops for biofuel production. Although multiple stakeholders were involved, it turned out that especially people in developing countries were negatively affected by the increased food prices.

Unexpected outcomes can be seen as a main characteristic of innovation processes. Recent literature tends to see innovation as a cumulative, stepwise process of collaboration between multiple actors with often unexpected outcomes (Rammert 1997). If innovation is a highly uncertain process, though, which is characterized by interdependency, serendipity etc., it is questionable whether this uncertainty can ever be steered in desired directions through relatively simple means. Well-known puzzles like the Jevons Paradox—innovations that increases energy efficiency for instance tends to increase the consumption of energy as well—show that the outcome of innovations may be contrary to what was intended (cf. Owen 2012).

Also, most of the responsible innovation literature starts from the plausible premise that innovations are radically uncertain and that their societal and environmental consequences are virtually unpredictable (for an example, see Ozdemir et al. 2011). The question then is what responsibility could mean if this premise were indeed true. Are responsible innovations those forms of science and technology development whose protagonists are willing and able to take accountability for, or 'stand up for', the societal and environmental consequences of their endeavors? If there is practically no way to predict or foresee such consequences, however, all talk about responsibility in this context would seem groundless and misleading.

This point can be illustrated with the requirement of 'accountability' that has found its way in the recommended Code of Conduct for Responsible N&N (Nano sciences and Nanotechnologies) Research for the European Union: "Researchers and research organizations should remain accountable for the social, environmental and human health impacts that their research may impose on present and future generations" (European Commission 2008).[2] This extremely far-reaching commitment exposes Nano researchers, as two analysts rightly point out, to "unbounded hazards of moral luck" (Grinbaum and Groves). In the future, researchers could be held liable for any untoward effects of their research that are as yet impossible to predict but that may manifest themselves only after decades. Their fate could be even worse than that of the Italian seismologists who were sued and convicted for failing to accurately predict an impending earthquake. Or is being held 'accountable' deemed to have no consequences for the researchers? Then the stipulation in the Code of Conduct is just an empty play of words.[3]

We have to admit that the EU Code of Conduct for Responsible N&N Research is not representative for the essential meaning of responsible innovation. Nevertheless, while the radical uncertainty of innovations and the unpredictability of their societal and environmental effects is fully acknowledged, it is suggested that this uncertainty can somehow be overcome by making the whole process more inclusive and more reflexive from the very outset (cf. Flipse 2012). Owen and Goldberg for instance argue: "Embedding iterative risk (and benefit) analysis with technology assessment and public/stakeholder engagement approaches within innovation research proposals was seen as offering a mechanism that considers technical risk issues and associated uncertainties, but that could also provide opportunities for identifying as yet unforeseen effects (economic, societal, and ethical) as these emerge. It may also facilitate upstream engagement with stakeholders and the public as to how these emerging impacts are received" (Owen and Goldberg 2010, p. 1705). Regardless of the inherent unpredictability of the future, 'foresight' can still be exercised, it is claimed, when stakeholders are involved to deliberate on various scenarios for possible futures; societal values and concerns can still be taken on-board 'midstream' and then 'modulate' on-going technological trajectories. René von Schomberg expresses the underlying view as follows in an interview in LEV: "You need a smart innovation process in which you do not place your bets too much on one particular technology, because technology development is unpredictable. Bring all involved societal parties together and let them jointly determine what they expect from the research trajectory. Then you can formulate research agendas, which can be mutually adjusted in order to arrive at the desired end result". In another publication he wrote that societal actors become "co-responsible for the innovation process by a constructive input in terms of defining societal desirable products" (Schomberg forthcoming). While he acknowledged that "an ethics focused on the intentions and/or consequences of actions of individuals is not appropriate for innovation", he nevertheless held that there is "collective responsibility both for the right impacts and negative

[2]In the final Code of Conduct for Responsible N&N Research, the formulation has been slightly nuanced: "Researchers and research organizations should remain accountable for the social, environmental and human health impacts of their work". However, this reformulation of the Code doesn't solve the underlying issue.

[3]For an effort to develop a concept of ethical oaths which imply actual ethical behaviour, see Blok (2013).

consequences" (von Schomberg 2013). Arguably this still presumes that these impacts and consequences, although perhaps not attributable to the actions of individuals, can somehow be foreseen by the various societal actors that are involved in the innovation process, a presumption that had initially been denied.

The third reason to question the possibility of responsible innovation is, that the 'responsibility' or 'accountability' is principally limited because of the radical uncertainty of innovation processes. This uncertainty even increases in case of wicked problems like sustainable development, because no simple solutions exist; all our proposed solutions will have unintended consequences and remain finite and provisional compared to the complexity and depth of the sustainability problem itself. Because our *knowledge* of the solution of wicked problems is principally limited, we can call this an *epistemic insufficiency* (cf. DeMartino 2013). If the output of responsible innovation processes is characterized by a fundamental uncertainty, which means that our *knowledge* of the impact of our innovations is not only limited but principally *insufficient*, the presupposed 'foresight' of responsible innovation becomes questionable. In other words, our knowledge is principally insufficient to assess the impact of innovation processes and there will always be unintended consequences of our innovations which can be harmful.

The analysis of the input, throughput and output of innovation processes raised several questions with regard to the possibility of responsible innovation and pointed at several difficulties which undermine the responsible character of innovation processes. It became clear that the practical applicability of the concept of responsible innovation is highly problematic and that a more thorough inquiry of the concept is required.

In the next section, we therefore ask what concept of innovation is presupposed in the responsible innovation literature.

2.3 The Presupposed Concept of Innovation in the Responsible Innovation Literature

What concept of innovation is self-evidently presupposed in the responsible innovation literature? In a recent paper, Benoît Godin sketched the history of the innovation concept. Although innovation always existed throughout history, it became only very recently "the emblem of the modern society" (Godin 2009, p. 5; cf. Nowotny 2008). While the concept originally concerned *novelties* in the broadest sense of the word—including imitation, invention, creative imagination, change—it became only recently restricted to *technological* innovation and *commercialized* innovation. Nowadays, it is almost self-evident that innovation does not only concern the *exploration* of new technologies, but also the commercial *exploitation* of these new technologies.

According to Godin (2009, p. 21), the restriction of innovation to technological innovation is rooted in what he calls, the 'culture of things' or material culture: "The origins of this culture go back to the Renaissance: due to commercial exchanges, exploration and travel, natural and artificial objects have been what is valued in arts, science, and real life". The focus on technological innovations is further enhanced by the introduction of patent laws from the fourteenth century onward, because only these kinds of innovations can be patented (cf. Macleod 1988). With the emergence of economic thought in the seventeenth century, the *utilitarian value* of innovations became most important (Godin 2009; Schumpeter 1943). As a consequence of this economic paradigm of technological innovation, alternative forms of innovation like systems innovations (for instance agro-ecological innovations) or attitudinal innovations (for instance prevention or life style interventions) receive less attention because it is difficult to develop a business model on the basis of these kinds of innovations. A good example is the case of agricultural science and technology (S&T). Because of the current technological regime of agricultural S&T, technological innovations like genetic engineering are locked in and systems innovations like agro-ecological innovations are locked out (Vanloqueren and Baret 2009; cf. Mortensen et al. 2012).

The same economic paradigm of technological innovation is presupposed in case of responsible innovation. The Dutch research council for instance started recently with a new responsible innovation program which is intended for "short-term research projects into the ethical and societal aspects of new technology". Besides the scientific quality of the research projects on responsible innovation, the "added value, societal relevance and knowledge utilisation" are the most important criteria for funding (our emphasis, NWO 2012; cf. Technology Strategy Board 2012). Also from the EU perspective, responsible innovation is, among others, characterized by "Assessing and effectively prioritising social, ethical and environmental impacts, risks and opportunities, both now and in the future, alongside the technical and commercial" (Matter 2011). The economic paradigm of technological innovation is also reflected in the fact that investments in responsible research and innovation (RRI) is primarily legitimized by the macro-economic arguments, i.e. that RRI will lead to the creation of jobs and economic growth (von Schomberg 2013). In the Lund declaration for instance, it is claimed that "meeting the grand challenges will be a prerequisite for continued economic growth and for improved changes to tackle key issues" (cited in von Schomberg 2013, p. 12).

The applicability of the economic paradigm of technological innovations on the concept of responsible innovation may be questioned, because exactly the imperative of economic growth and profit is at odds with the other P's People and Planet, to say the least. Some philosophers even argue that the growth imperative is above everything else 'responsible' for the rapid environmental destruction, resource depletion and impoverishment of populations all over the globe. Huesemann and Huesemann (2011, p. 256) for instance argue, that "our economic system's inherent dependence on growth for survival (i.e. more of everything: more markets, more consumers, more raw materials, more energy, more cheap labor, etc.) is the root cause of many environmental problems and is in direct conflict with sustainability". For our discussion of the applicability of the economic paradigm of technological innovations on the concept of responsible innovation, it is sufficient to raise the question whether the fundamental tensions between the growth imperative and the demand for sustainability can ever be overcome, as long as we embrace the economic paradigm in our concept of responsible innovation. It is striking, in this respect, that corporate innovation is underrepresented in current research on the concept of responsible innovation.

Another assumption in responsible innovation is that innovations are inherently good, as they produce prosperity and jobs and meet societal challenges at the same time (von Schomberg 2013). Innovation is seen as a panacea for all problems (Godin 2009) and responsible innovation in particular serves the 'public good' (Matter 2011). Some researchers even argue for cognitive enhancement in order to increase the innovative abilities of the species (Greely et al. 2008, cited in Godin 2009).

With this focus on the inherently good of innovation, the 'Faustian' aspect of innovation processes is neglected. According to the godfather of innovation studies, Joseph Schumpeter (1883–1950), innovation is the product of *creative destruction.* To Schumpeter, "Capitalism […] is by nature a form or method of economic change and not only never is but never can be stationary. […] The fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers' goods, the new methods of production or transportation, the new markets, the new forms of industrial organization that capitalist enterprise creates. […] The opening up of new markets, foreign or domestic, and the organizational development from the craft shop and factory to such concerns as U.S. Steel illustrate the same process of industrial mutation […] that incessantly revolutionizes the economic structure *from within*, incessantly destroying the old one, incessantly creating a new one. This process of Creative Destruction is the essential fact about capitalism. It is what capitalism consists in and what every capitalist concern has got to live in" (Schumpeter 1943, p. 82–83). According to Schumpeter, it is exactly innovation which is responsible for creative destruction. The *construction* of new and innovative solutions is accompanied by the annihilation or *destruction* of the old rules and the old order (including the 'writing off' of 'obsolescent' skills, technologies and capital stocks), i.e. the positive impact of responsible innovation and contribution to the public good, is therefore accompanied by negative impacts elsewhere. This Faustian aspect of innovation processes is largely ignored by the responsible innovation literature. Furthermore, in line with the idea that innovation is a cumulative process with often unexpected outcomes (Sect. 2.2), we cannot claim that the outcome of responsible innovation will automatically contribute to the public good. Well-known puzzles like the Jevons Paradox—innovations that increases energy efficiency for instance tends to increase the consumption of energy as well—show that the outcome of innovations may be contrary to what was intended (cf. Owen 2012). Innovation implies pain, annihilation and destruction, and it is questionable whether this 'Faustian aspect' of innovation can ever be overcome by integrating social and ethical aspects in the design process.

The effort to integrate social and ethical aspects in the design process shows another assumption of responsible innovation. Mutual responsiveness means that multiple stakeholders participate or are able to participate in the formation of societal and ethical norms. These norms are valid if the participants agree to it as participants of practical discourses (Habermas 1990, 1993). In order to form societal and ethical norms together, stakeholders should be able to listen to others, to be empathetic with regard to the interests of others, to take the perspective or standpoint of the other etc. (Kaptein and van Tulder 2003). Mutual responsiveness therefore presupposes symmetry between moral agents and moral addressees (Mackin 2011). Only because stakeholders can hear the voice of the other and can take the perspective of the other, they can become mutual responsive. Given the existing information asymmetries and investment imbalances, let alone the epistemic insufficiency with regard to future generations for instance (cf. Sect. 2.2), it is highly questionable whether the presupposed symmetry between moral agents and moral addressees is legitimized.

The analysis of the concept of innovation which is presupposed in current literature on responsible innovation, shows that innovation is self-evidently seen as (1) technological innovation, (2) which is primarily perceived from an economic perspective, (3) is inherently good and (4) presupposes a symmetry between moral agents and moral addressees.

2.4 Conclusions

In Sect. 2.2, we raised several questions with regard to the possibility and applicability of responsible innovation processes. We pointed at several difficulties with regard to responsible innovation, i.e. with attaining a more socially just and environmentally sound trajectory of innovation that will appeal to all the stakeholders involved. Especially with respect to the 'grand challenges' or wicked problems, it turned out to be very difficult if not impossible to satisfy the ideal of responsible innovation. But also more generally, it became clear that the wish for (more) responsibility clashed with the realities of existing innovation processes. As we have pointed out with respect to the input, throughput and output of innovation processes, the demand for responsibility runs into serious problems. The main difficulty of responsible innovation revolves around the response-*ability* of actors in the innovation process, due to 'epistemic' factors like the inherent complexity, uncertainty and unpredictability of technological innovation on the one hand, and 'moral' and 'political' factors like conflicting worldviews, interests and value systems among stakeholders and power imbalances on the other. We concluded that the practical applicability of the concept of responsible innovation is highly questionable.

In Sect. 2.3, we analyzed the presupposed concept of innovation in responsible innovation. With regard to the first two characteristics—innovation is technological and primarily seen from an economic perspective—we have to conclude that the prevailing concept of innovation is uncritically taken over in the concept of responsible innovation. With regard to the third characteristic of innovation—innovation is inherently good—we showed that the current concept of responsible innovation is highly naïve. Finally, our discussion of information asymmetries and our epistemic insufficiency to assess the future impact of our innovations (Sect. 2.2) raised fundamental questions with regard to the fourth assumption of responsible innovation—the symmetry of moral agents and moral addressees.

Because the possibility and applicability of responsible innovation is highly questionable and the presupposed concept of innovation is uncritical, narrow and naïve, we have to conclude that the conventional concept of innovation—innovation is technological and seen from an economic perspective—cannot be *upgraded* to a more responsible concept of innovation; responsible innovation calls for a radical transformation of the concept of innovation itself. Based on the findings in this article, we propose the following research agenda for future research on responsible innovation:

- With regard to the input of innovation processes, future research should focus on the question how to deal with power-, vision-, goal-, sector- and motive-deviations among stakeholders involved in responsible innovation processes, especially with regard to social and ethical aspects.[4]
- With regard to the throughput of innovation processes, future research should focus on the question how stakeholders become mutual responsive to each other, given the remaining investment- and risk imbalances and given the remaining information asymmetries among stakeholders involved in responsible innovation processes.
- With regard to the output of innovation processes, future research should focus on the question how to assess the future impact of responsible innovation processes, given the fundamental uncertainty of innovation processes and given our epistemic insufficiency with regard to the future impact of responsible innovation processes.

We can also raise more fundamental question with regard to the concept of responsible innovation itself. With regard to the concept of innovation which is presupposed in the majority of the responsible innovation literature, future research should broaden our conception of innovation, including non-technological innovations and non-market environments. Furthermore, future research should consider the concept of response-*ability* in the context of alternative strategies of innovation, in which the relationships between stakeholders substantially differ from those involved in conventional innovation processes. A particularly promising direction for future research might be the examination of user-based or user-centered innovations, free and open source, and commons-based peer-to-peer (p2p) innovation strategies. These alternative strategies have steadily proliferated over the last two decades in various domains of the economy, may possess characteristics that are less susceptible to the flaws that we have diagnosed in this article, and may therefore provide directions for more responsible innovation processes in the future.

With regard to the concept of responsibility which is presupposed in the majority of the responsible innovation literature, future research should consider the question what consequences our epistemic insufficiency will have for our concept of responsibility in general and responsible innovation practices in particular.

---

[4]For an attempt to deal with fundamental differences among multiple stakeholders during stakeholder dialogue, see Blok (2014).

Acknowledgments This article owes much to the inspiring discussions about RRI that the authors were fortunate to have with Henk van den Belt.

References

Batie, S.S. 2008. Wicked problems and applied economics. *American Journal of Agricultural Economics* 5: 1176–1191.

Blok, V. 2013. The power of speech acts: reflections on a performative concept of ethical oaths in economics and business. *Review of Social Economy* 71(2): 187–208.

Blok, V. 2014. Look who's talking: Responsible innovation, the paradox of dialogue and the voice of the other in communication and negotiation processes. *Journal of Responsible Innovation* 1(2): 171–190. doi:10.1080/23299460.2014.924239.

Blok, V., Hoffmans, L., and Wubben, E. 2015. Stakeholder engagement for responsible innovation in the private sector: Critical issues and management practices in the dutch food industry. *Journal of Chain and Network Sciences*.

Bos, J., V. Blok, and R. van Tulder. 2013. From confrontation to partnership. The role of a Dutch non-governmental organisation in co-creating a market to address the issue of animal welfare. *International Food and Agribusiness Management Review* 16(A): 69–75.

Bryson, J.M., B.C. Crosby, and M. Middleton Stone. 2006. The design and implementation of cross-sector collaborations: Propositions from the literature. *Public Management Review* 66: 44–55.

Chesbrough, H.W. 2003. Open innovation: The new imperative for creating and profiting from technology. Boston, MA: Harvard Business School Press.

Collingridge, D. 1981. *The social control of technology*. Palgrave: Macmillan.

Correljé, A., E. Cuppen, M. Dignum, U. Pesch, and B. Taebi. 2015. Responsible innovation in energy projects: Values in the design of technologies, institutions and stakeholder interactions. In *Responsible innovation, volume 2: Concepts, approaches, and applications,* ed. B.J. Koops, I. Oosterlaken, J. van den Hoven, H.A. Romijn, and T.E. Swierstra. Dordrecht: Springer.

Davies, S.R., M. Horst. 2015. Responsible innovation in the US, UK and Denmark: governance landscapes. In *Responsible innovation, volume 2: Concepts, approaches, and applications,* ed. B.J. Koops, I. Oosterlaken, J. van den Hoven, H.A. Romijn, and T.E. Swierstra. Dordrecht: Springer.

Delgado, A., K.L. Kjølberg, and F. Wickson. 2010. Public engagement coming of age: From theory to practice in STS encounters with nanotechnology. *Public Understanding of Science* 20(6): 826–845.

DeMartino, G.F. 2013. Professional ethics, codes and oaths: What's appropriate for economics? *Review of Social Economy* 71(2): 166–186.

Dutta, D.K., and M.M. Crossan. 2005. The nature of entrepreneurial opportunities: Understanding the process using the 4I organizational learning framework. *Entrepreneurship Theory and Practice* 29: 425–449.

European Commission. 2008. Recommendation on 'A code of conduct for responsible nanosciences and nanotechnologies research. Brussels.

European Commission. 2011. Horizon 2020—The framework programme for research and innovation. Brussels.

Eweje, G. 2007. Strategic partnerships between MNEs and civil society: The post Wssd perspectives. *Sustainable Development* 15(1): 15–27.

Flipse, S.M. 2012. Enhancing socially responsible innovation in industry. Practical use for considerations of social and ethical aspects in industrial life science & technology. Delft.

Godin, B. 2009. Innovation: The history of a category. Working paper.

Gulati, R. 2007. *Managing network resources: Alliances, affiliations, and other relational assets*. Oxford: Oxford UP.

Habermas, J. 1990. *Moral consciousness and communicative action*. Cambridge: MIT press.
Habermas, J. 1993. *Justification and application remarks on discourse ethics*. Cambridge: MIT press.
Hardy, C., T.B. Lawrence, and D. Grant. 2005. Discourse and collaboration: The role of conversations and collective identity. *Academy of Management Review* 30(1): 58–77.
Hemmati, M. 2002. *Multi-stakeholder processes for governance and sustainability. Beyond deadlock and conflict*. Earthscan: London.
Hens, L., and B. Nath. 2003. *Managing to collaborate: The theory and practice of collaborative advantage*. New York: Routledge.
Huesemann, M., and J. Hueseman. 2011. *TechNoFix. Why technology won't save us or the environment*. Gabriola Island: New Society Publishers.
Ireland, R.D., M.A. Hitt, and D. Vadyanath. 2002. Alliance management as a source of competitive advantage. *Journal of Management* 28(3): 413–446.
Iyer, E. 2003. Theory of alliances: Partnership and partner characteristics. *Journal of Nonprofit & Public Sector Marketing* 11(1): 41–57.
Jackson, R., F. Barbagallo, and H. Haste. 2005. Strengths of public dialogue on science-related issues. *Critical Review of International Social & Political Philosophy* 8(3): 349–358.
Kaptein, M., and R. van Tulder. 2003. Toward effective stakeholder dialogue. *Business and Society Review* 108(2): 203–224.
Kirzner, I. 1973. *Competition and entrepreneurship*. Chicago: University of Chicago Press.
Koops, B.J. 2015. The concepts, approaches, and applications of responsible innovation; an introduction. In *Responsible innovation, volume 2: Concepts, approaches, and applications,* ed. B.J. Koops, I Oosterlaken, J van den Hoven, H.A. Romijn, and T.E. Swierstra. Dordrecht: Springer.
Kreuter, M.W., C. de Rosa, E.H. Howze, and G.T. Baldwin. 2004. Understanding wicked problems: A key to advancing environmental health promotion. *Health, Education, and Behavior* 31: 441–454.
Kroesen, J.O., R. Darson, D.J. Ndegwah. 2015. Capacities, development and responsible innovation. In *Responsible innovation, volume 2: concepts, approaches, and applications,* ed. B.J. Koops, I Oosterlaken, J van den Hoven, H.A. Romijn and T.E. Swierstra. Dordrecht: Springer.
Mackin, G. 2011. The *aporia* of practical reason: Reflections on what it means to pay due respect to others. *Contemporary Political Theory* 10(1): 58–77.
Macleod, C. 1988. *Inventing the industrial revolution: The english patent system, 1660–1800*. Cambridge: Cambridge UP.
Matter. 2011. A report on responsible research & innovation. Available at: http://ec.europa.eu/research/science-society/document_library/pdf_06/rri-report-hilary-sutcliffe_en.pdf.
McMullen, J.S., and D.A. Shepard. 2006. Entrepreneurial action and the role of uncertainty in the theory of the entrepreneur. *The Academy of Management Review* 31(1): 132–152.
Millar, C., Y. Udalov, and H. Millar. 2012. The ethical dilemma of information asymmetry in innovation: Reputation, investors and noise in the innovation channel. *Creativity and Innovation Management* 21(2): 224–237.
Milne, G.R., E.S. Iyer, and S. Gooding-Williams. 1996. Environmental organization alliance relationships within and across nonprofit, business, and government sectors. *Journal of Public Policy and Marketing* 15(2): 203–215.
Mortensen, D.A., J.F. Egan, B.D. Maxwell, M.R. Ryan, and R.G. Smith. 2012. Navigating a critical juncture for sustainable weed management. *BioScience* 62(1): 75–84.
Netherlands Organisation for Scientific Research (NWO). 2012. Responsible innovation research program. Available at: http://www.responsible-innovation.nl/conference/conf11/index.php?option=com_content&view=article&id=24&Itemid=34.
Nowotny, H. 2008. *Insatiable curiosity: Innovation in a fragile future*. Cambridge: MIT Press.
Owen, D. 2012. *The conundrum*. New York: Riverhead Books.
Owen, R., and N. Goldberg. 2010. Responsible innovation: A pilot study with the U.K. engineering and physical sciences research council. *Risk Analysis* 30(11): 1699–1707.

Ozdemir, V., S.A. Faraj, and B.M. Knoppers. 2011. Steering vaccinomics innovations with anticipatory governance and participatory foresight. *OMICS: A Journal of Integrative Biology* 15(9): 637–646.

Penders, B., J.M.A. Verbakel, and A. Neis. 2009. The social study of corporate science: A research manifesto. *Bulletin of Science, Technology & Society* 29(6): 439–446.

Rammert, W. 1997. Innovation im Netz. Neue Zeiten für Innovation: heterogen verteilt und interaktiv vernetzt. *Soziale Welt* 48(4): 394–416.

Rittel, H.W.J., and M.M. Webber. 1973. Dilemmas in a general theory of planning. *Policy Sciences* 4: 155–169.

Roelofsen, A., W.P.C. Boon, R.R. Kloet, and J.E.W. Broerse. 2011. Stakeholder interaction within research consortia on emerging technologies: Learning how and what? *Research Policy* 40(3): 341–354.

Rogers-Hayden, T., and N. Pidgeon. 2007. Moving engagement upstream? Nanotechnologies and the royal society and royal academy of engineering's inquiry. *Public Understanding of Science* 16(3): 345–364.

von Schomberg, R. 2013. A vision of responsible research and innovation. In *Responsible innovation*, ed. R. Owen, M. Heintz, and J. Bessant. London: Wiley.

Schumpeter, J.A. 1943. *Capitalism, socialism & democracy*. London/New York: Routledge.

Selsky, J.W., and B. Parker. 2005. Cross-sector partnerships to address social issues: Challenges to theory and practice. *Journal of Management* 31(6): 849–873.

Technology Strategy Board. 2012. Responsible innovation framework for commercialization of research findings. http://webarchive.nationalarchives.gov.uk/20130221185318/www.innovateuk.org/_assets/responsible_innovation.pdf.

Van Huijstee, M.M., M. Francken, and P. Leroy. 2007. Partnerships for sustainable development: A review of current literature. *Environmental Sciences* 4(2): 75–89.

Vanloqueren, G., and P.V. Baret. 2009. How agricultural research systems shape a technological regime that develops genetic engineering but locks out agroecological innovations. *Research Policy* 38(6): 971–983.

Yaziji, M., and J. Doh. 2009. NGOs and corporations: Conflict and collaboration. Cambridge: Cambridge UP.

Chapter 3
Responsible Innovation in the US, UK and Denmark: Governance Landscapes

Sarah R. Davies and Maja Horst

Abstract This chapter explores the notion of responsible innovation (RI) as it is currently being imagined in policy and governance practice. It does this in the context of three different countries: the UK, US and Denmark. We ask how RI is being constituted within policy discussion. What is it understood as being? What kinds of actors are implicated in it? And what is its scope, or field of action? In exploring these questions we argue that responsible innovation is currently a largely international discourse, and that it remains unclear, from current policy discussion, how it should be put into practice. Though it is tied to a linear model of science and technology, in which both the process and outputs of scientific research are, through RI, imbued with responsibility, the actors involved and the fields in which they are assumed to operate are exceedingly general. As such, RI appears to be a fundamentally de-individualised process.

Keywords Responsible innovation · Civic epistemology · Discourse analysis · De-individualisation

3.1 Introduction

This chapter discusses a number of issues. It will explore the notion of responsible innovation as it is currently being developed in policy and governance practice, and it will do this in the context of three different countries: the UK, US and Denmark. In doing so our discussion will chart some of the contours of this thing called responsibility. What is it understood as being? What kinds of actors are implicated in it? And what is its scope, or field of action? As Bert-Jaap Koops

S.R. Davies (✉) · M. Horst
Department of Media, Cognition and Communication,
University of Copenhagen, Copenhagen, Denmark
e-mail: dxq327@hum.ku.dk

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_3

noted in the Introduction to this volume, responsible research and innovation is a concept that is both increasingly used and highly ambiguous. Drawing on a number of very different theoretical and practical traditions, and focusing on at least two different aspects of innovation, its *processes* and its *products* (Chap. 1), responsible innovation remains a practice in the making. Our contribution to this discussion is to explore some of the ways in which, at least with regard to written discourse, it is being framed and made. In exploring these questions we argue that, despite differences between contemporary imaginations of the relationship between science and society in different national contexts, responsible innovation is currently a largely international discourse. For the most part it remains unclear, from current policy discussion, how it is—and should be—put into practice. As such this analysis functions to clarify the ways in which an important concept within contemporary management of the science-society relationship, that of 'responsibility', is being used.

This analysis, which focuses on the policy landscape of responsible innovation, draws on work carried out in the 'Scientific Social Responsibility' (SSR) research project.[1] The research, which is based in Denmark but which takes a comparative approach in order to look at the US and UK as well as the Danish context, explores the discourse and practice of responsibility in science at a number of levels: that of national and international policy, the meso-level of research and laboratory managers, and that of bench scientists and everyday practice. In addition, it focuses on new and emerging science, and in particular on nanotechnology and synthetic biology. As a whole the research thus explores the operationalisation of and meanings attributed to discourses of responsibility in emerging science and technology (see Davies et al. 2014; Glerup and Horst 2014). Here we primarily report on analysis of macro-level, policy-oriented discussion of responsibility—the national and international backdrop to the ways in which scientists working at different levels understand, articulate, and put into practice calls for responsibility.

The SSR research is qualitative and discourse-oriented, focusing on explicating the meanings that circulate around particular concepts and on understanding their construction and uses (Law 1994). We view calls for responsible innovation as part of a wider governance turn within the management of science and technology (and indeed within politics more generally). This turn is expressed through a move away from top-down, command and control regulation and towards a broader distribution of 'soft law' activities such as voluntary reporting, codes of conduct, and the outsourcing of technological management (Benz and Papadopoulos 2007; Kearnes and Rip 2009; Shamir 2008). Shamir, in particular, has written about the importance of notions of responsibility within these moves. The dominance of corporate social responsibility as a frame for business, for instance, reflects a tendency for "corporations to assume socio-moral duties that were heretofore assigned to organizations, governmental entities and state agencies … the very

[1]See project website: http://mef.ku.dk/forskning/fokusomraaderogprojekter/forskningenssamfundsmaessigeansvarlighed/. The SSR project is funded by The Danish Council for Independent Research: Social Sciences.

notion of moral duty [becomes grounded] within the rationality of the market" (2008, 4). Our discussion will reflect upon the ways in which this moralisation of the corporate, and concomitant dissolution of distinctions between market and society, are mirrored (or not) in moves around the responsible conduct and development of science. As we outline the ways in which notions of responsible innovation are being articulated, we also want to speak to discussions around civic epistemology by tying the different national contexts we describe to particular historical trajectories around scientific citizenship (cf. Mejlgaard 2009).

We start, then, by exploring the history of three different national contexts, outlining, from extant literature, some characteristics of the 'civic epistemologies' of the US, UK and Denmark. We then describe a more focused review of web-based content relating to responsible innovation, retrieved through a series of searches in Google for "responsible innovation".[2] These searches were deliberately open: we were interested not just in academic literature but in references within grey (policy) literature and in the 'real world' of research and innovation (cf. Shelley-Egan 2009). We therefore identified results in academic literature, policy discussions and reports, and from more general websites. As Table 3.1 indicates, this search pulled up results from a wide range of sources, including academic articles (for instance, written by scientists calling for responsible research; Drenth 2006), policy discussion (for instance European Commission reports; DG Research 2011), and a wide range of more 'meta' references, such as the titles of funding programmes (see, for instance, Chap. 1), research projects, or think tanks. Some fifty references with 'responsible innovation' in their title were collated in the reference manager Zotero. This content was then coded by hand by the lead author, and discussed by the project team, with reference to themes of the nature of responsible innovation, the actors implicated in it, and its scope or frame of reference. In vivo coding was also used to identify emergent themes.

3.2 National Contexts: The US, UK and Denmark

In exploring contemporary policy discussion of responsible innovation we take three different national contexts as our case studies: Denmark, the UK, and US. In focusing on these we do not only assume that, for instance, regulation might focus on different issues in these different locations, or that it will take somewhat different forms (though of course these things are true; see Falkner and Jaspers 2012; Morris et al. 2010). Rather we are drawing on traditions which explore the links between particular national imaginations and identities and epistemic cultures (see Horst and Irwin 2010; Jasanoff 2005; Jasanoff and Kim 2009; Macnaghten and Guivant 2011). Such work locates questions of scientific citizenship, policy and governance within broader cultural dynamics involving, for example, shared assumptions about public knowledge or 'sound science'. Miller writes that:

[2] The web search therefore focused on English language discussion of responsible innovation.

Table 3.1 Key references to the scope and actants of 'responsible innovation' in grey literature and webpages

Document	Actants referenced	Scope of their actions
Policy documents		
Von Schomberg (2011). Also a number of similar book chapters and online discussions	"the consumer" "the European citizen" "society" "producers, users, stakeholders" "industry, civil society and research" "policy makers"	The application phase of new technologies "the innovation process" "European legislation"
Sutcliffe (2011)	"citizens" "the public" "civil society stakeholders" "the EC, governments, business and NGOs" "Europe" "governments, companies and research funders" "customers" "professional stakeholders" "experts" "the EC" "participants in dialogues" "the environment" "individual scientists"	Research and innovation "genuine innovation" The innovation process Corporate responsibility "oversight" "adaptive governance"
Technology Strategy Board (2012). www.innovateuk.org/_assets/responsible_innovation.pdf	The Technology Strategy Board "companies"	Regulation The creation of a sustainable economy
DG Research (2011)	"civil society" "Europe"	
Webpages		
KARIM Network: http://www.karimnetwork.eu/Pages/Responsible-Innovation.aspx	"academics, businesses and supply chain organisations" "business" "SMEs"	Competitiveness in global markets

(continued)

Table 3.1 (continued)

Document	Actants referenced	Scope of their actions
Unilever: http://www.unilever-esa.com/innovation/innovationinunilever/responsibleinnovation/	"consumers" "our business"	Managing and growing business Product safety
Oslo Research Group on RI: http://www.afi-wri.no/modules/module_123/proxy.asp?I=5553&C=416&D=2&mids=	"key users, stakeholders, affected parties and the society at large" "a wide range of disciplines and perspectives, including the natural sciences, the social sciences, the humanities and lay people"	Modern society
Nordic RI conference: http://www.csreurope.org/events.php?action=show_event&event_id=782	"Nordic and international investors, companies and stakeholders" The center for corporate diversity	
University of Leeds, Medical Technologies Support for Innovation: http://www.medical-technologies.co.uk/support-for-innovation/responsible-innovation/	"Researchers" "funders and regulators" "stakeholders" ("users, providers, regulators") "patients, consumers, families, activists, producers, funders and regulators"	The environment society
Franco-British Workshop on RI: http://www.ambafrance-uk.org/Franco-British-workshop-on,18791	"governments" "stakeholders" (academia, industry, government)	Particular nation states
Research Councils UK: http://www.rcuk.ac.uk/research/xrcprogrammes/Digital/Impact/Pages/ResponsibleInnovation.aspx	"researchers" "responsible innovation advisory panel"	A particular area of research ("the field of digital economy")
3TU Center for Ethics and Technology: http://www.ethicsandtechnology.eu/research/socially_responsible_innovation/	"companies" "customers, stockholders, the governments, investors and NGOs" "society" "investors" "Netherlands Organisation for Scientific Research"	
Observatory for Responsible Innovation: http://www.debatinginnovation.org/		"finance and financial services" "renewable energy in architecture"
Framework for Responsible Research and Innovation in IT: http://responsible-innovation.org.uk/frriict	"researchers" "the ICT community" "EPSRC" (Engineering and Physical Science Research Council)	ICT

> …the concept of civic epistemology seeks to capture the public knowledge ways or ways of knowing that operate within and across [different social and institutional spaces] … research on civic epistemologies inquires into how knowledge is dynamically constructed and applied in the search for meaning and design and implementation of policy in modern societies. (Miller 2008, 1897–8)

Here, then, we want to outline, very briefly, some of the hallmarks of the civic epistemologies of the US, US and Denmark as they have been discussed in the literature. What kinds of assumptions frame processes of knowledge production in each of these contexts?

We start with the UK—largely for the reason that the interactions between science and society have been so well characterised here that there is now a well-defined narrative around the move from 'deficit to dialogue' (see, for instance, Bauer et al. 2007; Burchell et al. 2009; Gregory and Lock 2008; Wilsdon and Willis 2004). This narrative runs as follows: interest in enhancing public understanding of science (PUS), which grew out of a 1985 Royal Society report (Royal Society 1985), was criticised as implicitly holding a deficit model of the public, in which laypeople were viewed as passive recipients of scientific knowledge (Miller 2001). This 'critical PUS' research, in combination with a series of very public scientific controversies such as that around British beef and vCJD (Jones 2004), resulted in a new emphasis on dialogue and public engagement with science (PEST). Science and scientists would no longer instruct publics; instead, they would draw on lay knowledges in open and transparent discussion (House of Lords 2000; Jackson et al. 2005), within formats such as consensus conferences or the science communication-oriented 'dialogue event'. Most recently, there have been calls to move this engagement even further upstream (Wilsdon and Willis 2004) and to incorporate attention to impact throughout the academy (HEFCE 2010).

This (obviously simplistic) narrative can and has been complicated and problematised (Gregory and Lock 2008; Irwin et al. 2012; Wynne 2006)—for instance in terms of the degree to which the move to dialogue has remained part of a fundamentally neoliberal agenda (Thorpe and Gregory 2010). But while what the move to PEST means in practice is still very much open to debate, it remains fair to say that UK science policy retains an unusual openness to public debate, deliberation, and the integration of social and ethical thought into scientific development (Irwin 2006). Most recently, as we will discuss, this has been expressed through the use of the notion of 'responsible innovation' by a number of key actors; more generally, the UK government, research councils, and scientific societies remain committed to issues around 'science and society' (Department for Innovation Universities and Skills 2012).

The UK's move towards dialogue on science was strongly influenced by Danish deliberative activities such as the consensus conference (Horst and Irwin 2010). It is ironic, then, that the configuration of science and society in Denmark has moved in the opposite direction to that of the UK—away from institutionalised participation, and towards an emphasis on deficit model-style, unidirectional communication (Horst 2012; Mejlgaard 2009). The policy context has now moved away from the use of the highly deliberative and egalitarian structures—exemplified most of all in the internationally-recognised Danish Board of Technology, or DBT—that were developed through the 1980s and '90s, and towards a new emphasis on innovation and knowledge dissemination. Since 1999, there have been a large number of reforms in Danish research and innovation policy, all designed to make Danish universities and research policy a vehicle for innovation, commercialisation, and economic growth, and these reforms have tended to erode traditions of participation. For instance, a 2003 Action Plan from the Science Ministry (*Fra tanke til faktura*, or *From thought to invoice*) explicitly introduced knowledge dissemination as a third mission for universities, but as primarily targeted at innovation and commercialisation rather than the public. Similarly, the *Forsk og Fortæl* (*Research and Tell*) report of 2004, also released by the Science Ministry, was developed not in collaboration with the key actors in Denmark's long history of deliberation and participation around science (such as DBT, the Council of Ethics, or the scholars who had studied these) but with, primarily, journalists with little background in debates about science and society—and perhaps as a result tended towards a PUS-style emphasis on the need for public scientific literacy.

To note these developments is not to criticise: in fact, this emphasis on PUS and the need for communication can be understood in terms of the development of new forms of scientific accountability, and a shared sense of the necessity of access to scientific knowledge for lay citizens to operate within increasingly technological societies (Mejlgaard 2009). But it does indicate a very different kind of context for dealing with questions of responsibility in science from the UK—one which does, certainly, retain the hallmarks of a peculiarly Danish culture around equality of status (and thereby the problematisation of experts and expertise), but which is increasingly turning that drive to equity against 'experts of community' such as the DBT—or, indeed, social scientists (Horst 2012).

The US presents a very different context again—one which, despite the presence of some of the language of 'public engagement' (see, for instance, Besley 2010; McCallie et al. 2009), has largely been isolated from European dynamics around scientific controversy, issues of public trust, and participatory democracy. At least part of the reason for this is the US's distinctive geography (a nation the size of a continent, and incorporating dense, cosmopolitan urban centres alongside vast tracts of rural land or wilderness), which has influenced a largely celebratory imagination of technoscience and its governance (Nye 2004), and its political culture—a culture that was formed at least in part through contrasts with the 'old world' (Anderson 2006) and which incorporates a set of characteristic legislative, democratic, and political procedures (Jasanoff 2005). Rather than single, dominant narratives such as those that we find around 'PUS to PEST' in the UK or 'PEST to PUS' in Denmark, the US presents uncoordinated and sporadic activity around the management of the relationship between science and society (cf. Kleinman 1995). Within different academic disciplines, geographic locations, and substantive policy areas there has variously been interest in, for example, public scientific literacy and how to improve this (Miller 1998), the use of dialogue events in informal science education (Bell 2008), the application of traditions of deliberative democracy (exemplified, in the US, by the town hall meeting) to science and technology (Brown 2009; Fischer 2009), and the use of 'real time technology assessment' (Guston and Sarewitz 2002) and other forms of public assessment of technoscience (Guston 1999). While there is, increasingly, a build-up of interest in 'anticipatory governance' of emerging technologies—the integration of public debate, social and ethical reflection, and policy and scientific engagement at an early stage of a technology's development (Guston 2010)—including the use of this language within government coordination of nanotechnology (Roco 2001; Roco et al. 2011)—the overall sense is of a pluralistic approach to scientific citizenship and of a context of "a well-established culture of technological optimism, [set] against a robust tradition of scientific activism" (Jasanoff 2005, 45).

These accounts are sketches, with no claim to being comprehensive. They are not designed to give final analyses of civic epistemologies. What they have done is to hint at the kinds of policy logics that have been reported in the literature as present in each location, and thus to suggest what kinds of arguments, discourses, and regulatory rationales may be applied to questions of responsibility in science. In the next section we explore to what extent this is the case by discussing the constitution of responsible innovation within national and international literatures.

3.3 Constituting Responsible Innovation

In this section we will outline what we found when we started looking more closely at the ways in which responsibility in science, and specifically the notion of responsible innovation (RI), is being used and discussed in practice. In doing so we are drawing on the web search around "responsible innovation" described in the Introduction. We studied what we found with an eye to location—geographic and sociocultural—as well as to our interests in the implied meanings, actors, and scope of RI. In what follows we describe the topline findings from this analysis, providing illustrative references to the body of literature identified and deconstructing in more detail some of the notions implied by the language used in this literature (Fairclough 2003).

3.3.1 An International, Rather Than National, Agenda

Strikingly, what emerged first was the international nature of this discussion. Of course, this is in some ways an artefact of our search method, which focused on the English language articulation 'responsible innovation' rather than the Danish 'social og videnskabelig ansvarlighed'. However, we expected that we might find different articulations (in translation or otherwise) of RI inflected by the different political environments of, for instance, the US and Denmark. Instead, much of the discourse was tied to the international level, rather than to specific national contexts—and in fact to a limited number of key actors. There were, certainly, clusters within which the term tends to occur. In more academic literature, it is primarily used by scholars in Europe (and in particular in the UK; see, for instance, Owen et al. 2009) and, to a lesser extent, in the US (where it co-locates with discussion either of the responsible conduct of research or of nanotechnology; Frankel 2000; Roco et al. 2011). But in terms of policy discourse, its use is currently fairly limited. A number of European Commission reports discuss the term (albeit in conjunction with concerns, in at least one case, about the relatively limited degree of impact such discussions might have; DG Research 2011; Sutcliffe 2011). The UK was the only context in which we found more extensive usage, with reference to principles of RI from policy actors such as the Technology Strategy Board (a government funded body which aims to "accelerate economic growth by stimulating and supporting business-led innovation"[3]) and Research Councils UK, as well as instances of more developed use of RI principles by the Engineering and Physical Sciences Research Council (Owen and Goldberg 2010; also Macnaghten and Owen 2011).

Given the UK's history, as outlined above, and current status as enthusiastic user of at least the language of public engagement (Davies 2013), it is perhaps not surprising that it also appears to be a frontrunner in developing and operationalising thinking on responsible innovation (Owen et al. 2009). Indeed, responsible innovation is being framed by some as a 'next step' for public engagement with science—one which turns attention back onto the innovation process rather than (what are often framed as problematic) publics. Thus Stilgoe suggests that RI has the potential to move policy debate beyond the increasingly stale use of public engagement, writing that:

> Public engagement has been too disconnected from policy. Rather than focussing on The Public, reinforcing the sense that they are somehow the problem, Responsible Innovation turns our attention towards innovation – how it is governed, who is responsible and what the alternatives might be. (Stilgoe 2012, 22)

In the UK, then, RI is developing against a backdrop of interest in public participation in science, and a strong collective memory of processes such as *GMNation?* and influential dialogues on nanotechnology (Jones 2008). In contrast, neither the US or Denmark have such a historical backdrop. In both cases, attention to responsible innovation seems to be developing in a rather piecemeal manner—in the US, through the involvement of the social science-oriented Centers for Nanotechnology in Society in the continuing articulation of the National Nanotechnology Initiative (see, for instance, Roco et al. 2011), and in Denmark, through publications such as the Royal Academy's *Samfundets bevågenhed—Universiteternes ansvarlighed* ('Societal Attention—Universities' Responsibility'; Harder et al. 2010) which, however, emphasise the importance of researcher autonomy (twinned with accountability) rather than public involvement (cf. Chaps. 1 and 4). Features of these

[3]See http://www.innovateuk.org.

differing national contexts seem, as yet, to be influencing the unfolding of RI only to a very limited extent.[4]

3.3.2 A Rather Rarefied Discussion

We also found it easier to identify discussion of the principles of RI than examples of it being carried out in practice (with the exceptions of the British trials noted above; Owen and Goldberg 2010).[5] The majority of the references that we found were at a fairly rarefied level: only a limited number emerged from business, for instance, and the majority were writing from within the academy. At this point it seems that the concept remains in a state of experimentation and flux, and has had limited take-up by practitioners within policy and industry. It is not clear, for instance, to what extent it could or should overlap with discussion of CSR, which is now a key narrative within both commercial and government practice around emerging technology, innovation, and economic development (see Groves et al. 2011; Vallentin and Murillo 2012). While CSR is increasingly integrated into industry notions of 'good business' and used by governments as a tool for soft regulation (Midttun et al. 2012; Vallentin and Murillo 2012), scientific social responsibility seems to not yet be coupled to these moves.[6]

3.3.3 The 'What' of RI: An Emphasis on the Outcomes and Outputs of Science

Given these trends—and in particular the sense that RI is currently being developed by a limited number of actors working on an international level rather than in

---

[4]Of course, both the UK and Denmark are part of the European Union, and therefore part of Commission discussion of and policy on responsible innovation. Anecdotal evidence from early empirical work suggests that this may in fact be influencing the context of discussions of responsibility through assumptions of a division of labour. In such assumptions, the EU is seen as a key player in the articulation of policy on emerging technology (and indeed is both a key funder of such technological research—Shapira and Wang (2010)—and, as we have outlined, the primary source of the language of responsible innovation), such that there is little sense in replicating these discussions. In national scientific and policy imaginations of responsible research and innovation, then, some responsibility for responsibility may itself be being delegated to supranational institutional arrangements.

[5]The present volume will, of course, offer many more practical instances of RI than we could readily find at the time of the research (2012); see Part III of this volume in particular.

[6]CSR is now an enormous field, both in terms of academic research and as an industry in its own right, which we cannot do justice to here. The key point is that there is an extant tradition around responsibility, which emerged from industry but which has now been enthusiastically taken up by policy actors such as the European Commission (Midttun et al. 2012) as a means of governance, which so far seems not to have been tapped by work on responsible research and innovation. The degree to which it might be a helpful model for practices of RI, and the possibilities for successfully connecting to existing CSR activity and EC/government policy, deserves an essay in itself.

specific national contexts—what kinds of imaginations of RI can we identify? We want to discuss this 'what' question with reference to some of the definitions that various actors have offered up. Though, as one EU-supported report notes, the RI notion is new and "definitions are evolving" (Sutcliffe 2011, 3), it remains helpful to focus in on some of the ways in which it has been explicitly framed within different arenas.

We start with René von Schomberg's definition, which has been particularly influential: versions of this definition are found in a number of places—including von Schomberg's personal webpage, a 2011 book chapter on technology assessment, and in European Commission (EC) documents. We cite it here from a 2011 EC edited volume on RI in Information and Communication Technologies (ICT).

> 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). (von Schomberg 2011, 9; emphases ours)

We can take a number of points from this. First, RI is in this framing a process, albeit one with a clearly defined endpoint (the "proper embedding of scientific and technological advances in our society"). In this respect the definition builds on critiques of the UK's turn to public engagement, and the tendency of practice to focus on one-off events rather than long term engagement between science and publics (Stilgoe 2012). Second, this process has a number of characteristics: it is transparent, interactive, involves 'mutual responsiveness' between "societal actors and innovators", and focuses on issues of ethics, sustainability and desirability. Third, its focus is the "innovation process and its marketable products". RI is therefore explicitly framed as being about the outputs of science—its products and their markets and the way in which these are embedded into society. And, finally, there is little sense of contestation within this definition, which anticipates 'mutual responsiveness' and, apparently, the development of homogeneous 'views' on issues such as what will ensure that particular technologies are ethically acceptable. This is striking given that several decades of work on deliberation and public engagement suggest that consensus on issues such as these, when under discussion by varied stakeholders, is at best difficult and at worst impossible (Horst 2007). This (admittedly concise) account of RI, at least, thus gives little information on how to deal with tensions and (perhaps irreconcilable) differences in opinion.

The UK's Technology Strategy Board (a government instrument for supporting UK innovation) outline of RI—found in a call for studies in synthetic biology—is not dissimilar, though shorter and less developed:

> Responsible innovation requires careful consideration of *ethical, societal and regulatory issues* and appropriate response throughout the process, including (i) during *the process of carrying out the R&D*, and (ii) for *commercial use of the findings*. (Technology Strategy Board 2012, 1; emphases ours)

Again, there is reference to "ethical, societal and regulatory issues" and to the final output of the research—its "commercial use"; in addition, however, there is a more explicit mention of reflection during the process of the research. A later note on this (ibid, 1–2) explains that this refers to "good scientific practices" and to the clearance of research through gatekeepers such as university ethics review boards.

Both of these definitions emerge, broadly, from the policy domain. We want to cite two further references to RI which come from more applied areas, the first being taken from Unilever's website. Under the heading 'Responsible Innovation', they write:

> We believe our *products* should make a real contribution to an *individual's wellbeing* and that of their *community*, while having the *least possible adverse impact* on the *environment* at every stage in the *product lifecycle*.[7] (Emphases ours)

This is a public-facing webpage and, perhaps because of this, depicts a rather more general imagination of what RI can be. The emphasis, however, is again on *products*—the outcomes of innovation—and the way in which these should impact the "wellbeing" of individuals and communities. These products should also have "the least possible" negative effect on the environment. The European-funded KARIM network—a 'Knowledge Acceleration and Responsible Innovation Meta Network' oriented towards enhancing knowledge transfer across industry, universities and policy—contains a similar emphasis on safety, in the form of "environmental and economic risks":

> There is increasing *pressure* to explore ways of *developing new technologies* whilst also taking into account *environmental and economic risks* as well as broader *societal issues*. Moreover there is a pressure to take more responsible approach to innovation at *every stage of the innovation process* whether it is in *laboratories*, *industry*, *manufacturing* or *policy making*. … Responsible Innovation is a new and exciting concept. There is no globally agreed definition at present but we believe that *new values* can create better opportunities for *individuals*, *societies* and the *natural environment*. Over the life of KARIM we will provide access to the right knowledge at the right time during the *innovation process* in order to facilitate *more responsible outcomes*.[8] (Emphases ours)

Here, again, we find a dual emphasis on outcomes and the process of innovation (in "laboratories, industry, manufacturing or policy making"). Framed as a response to external drivers (the "increasing pressure" that is mentioned), RI is a "new and exciting concept" that rolls together risk, societal issues, "new values" and the "right knowledge at the right time" in order to ensure "responsible outcomes". While it is not entirely apparent how this will come about (what kinds of "new values" are at play, and what kinds of "opportunities" can they present? And what precisely are "more responsible outcomes"?), the extract shares, along with those above, a clear emphasis on the products and outcomes of scientific research. The central framework is that of "developing new technologies", and what is being imagined is a broadly linear progression between R&D and new products—both of which, as the process and outcomes of innovation, should be made more responsible.

[7] See http://www.unilever-esa.com/innovation/innovationinunilever/responsibleinnovation/.

[8] See http://www.karimnetwork.eu/Pages/Responsible-Innovation.aspx.

What links the discussions of RI we have examined, then, is this central, linear model of innovation they are based around, and the assumption that the purpose of (at least some) scientific research is to enable the development of beneficial products which will, as one RI conference website puts it, "enhance … economic growth and ensure prosperity".[9] This is not, of course, surprising (and see also Blok and Lemmens, Chap. 2 of this volume, for a more extended discussion of the model of innovation mobilised within RI). Academic literature has already made explicit the links between discourses of responsibility and the rise of new agendas around the entrepreneurial university, innovation and growth, and corporate social responsibility (Eztkowitz et al. 2000; Hellström 2003; Mahlouji and Anaraki 2009).[10] But if the 'what' of RI is that it is something that will make an already existing process—research that leads to profitable products—better, more 'responsible', in some way, it raises questions concerning the degree to which responsibility is thus delimited and constrained. If RI is framed as supporting the development of scientific outcomes, where does this leave, in Ferrari and Nordmann's terms (2010), the possibility of and capacity for saying 'no' to particular lines of technological research? We return to this theme in the conclusion.

3.3.4 The Actors of RI

Our second area of interest in this analysis was the 'who' of responsibility. What kinds of actors are being implicated in it? Again, we draw on our search through the literature of RI to discuss this.

The most striking aspect of this literature—policy, web-based and academic—is the sheer range of actors referenced. As Table 3.1 (which represents key language used around the implied actants of RI, and the scope of activity attributed to them) shows, a wide range of persons, organisations, groups and categories appear as implicated in RI in some way. These actors include, for instance: "Europe", "the public", "governments, companies and research funders", "civil society", "business", "stakeholders", "NGOs", "research councils", "investors", "citizens", and "researchers". Systematic assessment of the roles assigned to these different actants is thus difficult given the very broad categories that tend to be used. As noted earlier, discussion of RI tends to remain on a fairly rarefied level. It is unclear, for example, what kinds of specific, concrete actions or activities might be allocated to 'Europe'.

[9]Quoted from the website for the Franco-British Workshop on Responsible Innovation, http://www.ambafrance-uk.org/Franco-British-workshop-on,18791.

[10]Similarly, critiques of public engagement—which is, in the UK, at least, a key forerunner of RI (Stilgoe 2012)—have pointed out the intermeshing of participation and neoliberal regimes (Thorpe and Gregory 2010).

Many different actors are understood as being involved in RI, then, but they tend to take the form of rather general categories. Take, for instance, this extract from a report on RI supported by the European Commission:

> [Public] distrust manifests itself in many ways, but has stimulated *the public* and *civil society stakeholders* to be more interested and more vocal about the way *governments* and *companies* do what they do. … RRI [Responsible Research and Innovation] is about creating a shared understanding of the appropriate behaviours of *the EC*, *governments*, *business* and *NGOs* which is central to building the trust and confidence of *the public* and *other stakeholders* in the safe and effective systems, process and products of innovation. (Sutcliffe 2011, 6)

There are a number of different actors described as being relevant to or participating in RI: the public, civil society stakeholders, governments, companies, the EC, business, and NGOs. Each of these categories is a large one, to say the least, which renders the substance of the paragraph rather vague (are all publics uniformly "more interested and more vocal" with regard to governments and companies? How exactly should RI ensure "shared understandings" between the EC and NGOs? Or between governments and business?). Beyond this, however, it is striking that the individual is de-emphasised in this discussion. And in fact this is a general theme in the literature we have examined: the actors in RI are construed in terms of large, general categories, with little reference to particular, specific individuals or indeed to individual action at all. While the *idea* of citizenship, and the responsibility of science, seem important, individual citizens and scientists are occluded. This lacuna is something we will return to in the conclusion.

3.3.5 The Scope of RI

Our third interest in this research has been to explore the scope of RI, looking at what kinds of fields of action are assumed within discussions around it. Again, searching through documents discussing RI results in a long list of areas into which it is understood to reach (see Table 3.1): the "innovation process", "corporate responsibility", "regulation", a "sustainable economy", "global markets", "business", and "modern society". As with the actors of RI, then, the categories at play are extremely general ones. Though there are a few more specific frames drawn upon—particular nations, for instance, or a specific research funding theme—there is again little sense of specificity or of individualised action. The following extract, from the webpage of a research project on RI, demonstrates some of these dynamics:

> Researchers in *cutting edge fields* are increasingly being asked by funders and regulators to conduct responsible innovation in order to increase the social and economic *benefits* and effectively manage the *risks* of their work. They are expected to engage with the *environmental, health and social impacts* of the technologies they are developing, deliver benefit and identify and mitigate risks in the process.[11] (Emphases ours)

[11] Taken from http://www.medical-technologies.co.uk/support-for-innovation/responsible-innovation/.

This assessment of RI is actually one of the more specific within the literature that we have collected: here, RI has a specific focus (scientific research into medical technologies) and associated set of actors (the website lists "patients, consumers, families, activists, producers, funders and regulators" as well as the "researchers" mentioned above). The text also constructs a set of overlapping worlds within which RI is imagined as acting. The researchers work in "cutting edge fields", conjuring up a scientific and technical domain within which they go about their everyday business. "Funders and regulators" exist outside of this, but place certain demands upon it (in the form of requests to "conduct responsible innovation"); other worlds are suggested by the need to increase "social and economic benefits", "manage risks", and engage—presumably from within the semi-cloistered world of science—with "environmental, health and social impacts". While we are able, then, to build a picture of the overlapping domains implicated in RI—science, policy, society, the natural environment—and to gain some sense of the actions and effects that it should enable (the delivery of "benefits" rather than "risks"), the focus on this remains fuzzy. It remains unclear, for instance, how particular individuals or groups should act, with what scope, in which domains. This is even more the case for discussions of RI which refer to even larger domains—the "innovation process", for instance, or "European regulation". As with discussion of the actors of RI, a tendency to frame responsibility as a practice carried out by broad categories with broad effects results in a fundamental de-individualisation of the process.

3.4 Conclusion

Our discussion thus far has described the context to this research, briefly outlined the characteristics of three different national contexts in terms of imaginations of the science-society relationship, and explored the ways in which responsible innovation is being articulated within them. Somewhat to our surprise, we found that, as yet, RI is not strongly inflected by the histories and civic epistemologies of particular nations; instead, it remains a primarily international discourse which is largely present in the form of European policy, academic literature, and, less frequently, as a cognate for corporate social responsibility (though remaining largely decoupled, in practice, from debates around this). Once we start to unpick the meanings attributed to it, and the actors and actions portrayed as implicated in it, RI becomes even more difficult to pin down. Though it seems to be tied to an essentially linear model of science and technology, in which both the process and outputs (the 'beneficial products') of scientific research are, through RI, imbued with responsibility, both the actors implicated in it and the fields in which they are assumed to operate are exceedingly general ('the government', 'Europe', 'citizens'). As such, RI appears to be a fundamentally de-individualised process.

In closing we want to reflect on the implications of these points, and to return to some of the questions raised throughout the chapter. What, for instance, does it mean that RI appears to rely so heavily on a linear model of innovation—and one in which the ultimate endpoint of scientific research and technological development is 'beneficial products'? And what is the significance of the fact that RI, in its current articulations, rarely seems to be concerned with individuals?

As we suggested above, the use of a linear model of innovation opens up questions as to how easy it is, or could be, to actually intervene in the direction and outcomes of innovation. If science and technology are viewed as straightforwardly leading to societal benefits, and RI as being about honing this flow in some way so as to make it rather more sensitive to societal needs, then a number of lines of thought are effectively closed down—for instance the possibility of questioning not only exactly what is best for society, but whether technoscience is the best route to this (cf. Ferrari and Nordmann 2010; Stirling 2008).[12] Similarly, the emphasis on "marketable products" (von Schomberg 2011, 9) or "commercial use" (Technology Strategy Board 2012, 1) continues to tie RI into a regime in which technoscience is primarily oriented towards markets. As with Shamir's (2008) examination of CSR, it is a version of responsibility that has been adapted to (perhaps even designed for) neoliberal governmentalities. More radical versions of responsibility—for instance those that emphasise notions of care (Pellizzoni 2004), or which propose more disruptive interventions into the technoscientific complex (Wynne 2007)—are thus occluded by current debate.

These dynamics situate RI on the macro-scale: its actors are policy organisations, countries, governments, or societies; and their field of action comprises entire innovation pathways, national regulatory systems, or 'the environment'. That RI is primarily being imagined as working at this level raises a number of further questions (some of which, at least, are empirical ones). What does and should RI mean for individual researchers? At present the RI literature has little to say on this question, focusing instead on policy actors or, at times, 'researchers' as a homogenous community who should, for instance, engage in dialogue or consider ethical questions. It remains unclear, then, how RI is in practice entangled with citizenship and responsible behaviour at the level of individual scientists (see Chap. 4 for further evidence of and reflection on this). Just as publics are often muzzled and disempowered by broad gestures towards participation (Wynne 2007), we might speculate that current discourse on responsibility is shifting agency away from individual scientists-as-citizens and focusing it upon the institutions of democracy (government, regulators, research funders) and of the market (watchdogs, businesses, share- and stakeholders). Scientific citizenship, and the practice of responsibility, then, become something rather passive—something that is perhaps more about the idea of citizenship than the actions of citizens themselves. As such, there is currently something of a sense of being able to

[12]In addition, the accuracy of the linear model is itself highly questionable: innovation rarely flows straightforwardly from science to technology to product, but rather forms a complex web in which many different factors interact (cf Joly and Kaufman 2008; see Chap. 2). Relying on this model as a way of structuring responsible innovation seems likely to create expectations that cannot be met.

have your cake and eat it in the discourse around RI. It promises everything, to everyone—safety, profits, contented publics, innovation and growth, ethical rigour—without making specific demands of anyone. This vagueness of discourse means that all sorts of things can be promised without any need to articulate who should do what, where.

Understanding the degree to which this de-individualisation and removal of agency from researchers is, in fact, experienced at the level of bench research is in part an empirical question. Are researchers 'on the ground' aware of these macro moves, and if so, what do they make of them? The limited research that has been done on these kinds of questions (the exploration, for instance, of researcher familiarity with the European Code of Conduct on Responsible Nanoscience and Nanotechnologies Research; Kjølberg and Strand 2011) has suggested that to a large extent these debates remain firmly at the macro-level, rarely troubling those who are meant to be enacting them in the laboratory. Empirical work in progress on the SSR project will continue to elucidate these questions, exploring the ways in which researchers and research managers imagine and put into practice the notion of responsibility, and asking whether the gap between international policy and laboratory practice is being bridged. As such it will, hopefully, contribute to contemporary thinking around what it means to operationalise responsible innovation in science and technology.

Acknowledgments We would like to thank Cecilie Glerup and Raffael Himmelsbach for their comments on and help with the preparation of this manuscript.

References

Anderson, Benedict. 2006. *Imagined communities: Reflections on the origin and spread of nationalism*. Verso Books.

Bauer, Martin W., Nick Allum, and Steve Miller. 2007. What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda. *Public Understanding of Science* 16(1): 79–95.

Bell, Larry. 2008. Engaging the public in technology policy: A new role for science museums. *Science Communication* 29(3): 386–398. doi:10.1177/1075547007311971.

Benz, Arthur, and Yannis Papadopoulos (eds.). 2007. *Governance and democracy: Comparing national, European and international experiences*. London: Taylor and Francis.

Besley, J.C. 2010. Imagining public engagement. *Public Understanding of Science*. doi:10.1177/0963662510379792.

Brown, Mark B. 2009. *Science in democracy: Expertise, institutions, and representation*. Massachusetts: MIT Press.

Burchell, Kevin, Sarah Franklin, and Kerry Holden. 2009. *Public culture as professional science: Final report of the scope project (Scientists on public engagement: From communication to deliberation)*. London: LSE.

Davies, Sarah R. 2013. Constituting public engagement meanings and genealogies of PEST in two U.K. studies. *Science Communication* 35(6): 687–707. doi:10.1177/1075547013478203.

Davies, Sarah R., Cecilie Glerup, and Maja Horst. 2014. On being responsible: Multiplicity in responsible development. In *Responsibility in nanotechnology development*, ed. Simone Arnaldi, Arianna Ferrari, Paolo Magaudda, and Francesca Marin, 143–159. Netherlands: Springer.

Department for Innovation Universities and Skills. 2008. A vision for science and society: A consultation on developing a new strategy for the UK. London. Available at: http://interactive.dius.gov.uk/scienceandsociety/site/.
DG Research. 2011. *DG research workshop on responsible research and innovation in Europe*. Brussels.
Drenth, P.J.D. 2006. Responsible conduct in research. *Science and Engineering Ethics* 12(1): 13–21.
Etzkowitz, H., A. Webster, C. Gebhardt, and B.R.C. Terra. 2000. The future of the university and the university of the future: Evolution of ivory tower to entrepreneurial paradigm. *Research Policy* 29(2): 313–330.
Fairclough, Norman. 2003. *Analysing discourse: Textual analysis for social research*. Abingdon: Routledge.
Falkner, Robert, and Nico Jaspers. 2012. Regulating nanotechnologies: Risk, uncertainty and the global governance gap. *Global Environmental Politics* 12(1): 30–55.
Ferrari, Arianna, and Alfred Nordmann. 2010. Beyond conversation: Some lessons for nanoethics. *NanoEthics* 4(2): 171–181. doi:10.1007/s11569-010-0098-3.
Fischer, Frank. 2009. *Democracy and expertise: Reorienting policy inquiry*. Oxford: OUP.
Frankel, M.S. 2000. Scientific societies as sentinels of responsible research conduct. *Proceedings of the Society for Experimental Biology and Medicine* 224(4): 216–219. doi:10.1046/j.1525-1373.2000.22424.x.
Glerup, Cecilie, and Maja Horst. 2014. Mapping 'social responsibility' in science. *Journal of Responsible Innovation*, 1–20. doi:10.1080/23299460.2014.882077.
Gregory, Jane, and Simon Jay Lock. 2008. The evolution of 'public understanding of science': Public engagement as a tool of science policy in the UK. *Sociology Compass* 2(4): 1252–1265.
Groves, Chris, Lori Frater, Robert Lee, and Elen Stokes. 2011. Is there room at the bottom for CSR? Corporate social responsibility and nanotechnology in the UK. *Journal of Business Ethics*. doi:10.1007/s10551-010-0731-7.
Guston, David H. 1999. Evaluating the first U.S. consensus conference: The impact of the citizens' panel on telecommunications and the future of democracy. *Science Technology and Human Values* 24(4): 451–482. doi:10.1177/016224399902400402.
Guston, David H. 2010. The anticipatory governance of emerging technologies. *Journal of the Korean Vacuum Society* 19(6): 432–441.
Guston, David H., and Daniel Sarewitz. 2002. Real-time technology assessment. *Technology in Society* 24(1–2): 93–109. doi:10.1016/S0160-791X(01)00047-1.
Harder, Peter, Carl Bache, Mogens Flensted-Jensen, Mikael Rørdam, and Nina Smith. 2010. *Samfundets Bevågenhed - Universiteternes Ansvarlighed. Hvad Kan Forskerne, Universiteterne Og Forskningsrådene Selv Gøre?* Forskningspolitisk årsmøde 2010. Det Kongelige Danske Videnskabernes Selskab.
Hellström, Tomas. 2003. Systemic innovation and risk: Technology assessment and the challenge of responsible innovation. *Technology in Society* 25(3): 369–384. doi:10.1016/S0160-791X(03)00041-1.
Higher Education Funding Council for England, Scottish Funding Council, Higher Education Funding Council for Wales, and Department for Employment and Learning, Northern Ireland. 2010. Research excellence framework impact pilot exercise: Findings of the expert panels. *A report to the UK higher education funding bodies by the chairs of the impact pilot panels*. Bristol: HEFCE.
Horst, Maja. 2007. Public expectations of gene therapy: Scientific futures and their performative effects on scientific citizenship. *Science Technology and Human Values* 32(2): 150–171.
Horst, Maja. 2012. Deliberation, dialogue or dissemination: Changing objectives in the communication of science and technology in Denmark. In *The development of public communication of science and technology studies—A comparative approach*, edited by Michel Claessens, Bernard Schiele, and Shi Sunke. Berlin: Springer.

Horst, Maja, and Alan Irwin. 2010. Nations at ease with radical knowledge: On consensus, consensusing and false consensusness. *Social Studies of Science* 40(1): 105–126.

House of Lords. 2000. *Third report: Science and society*. London: The Stationery Office, Parliament.

Irwin, Alan. 2006. The politics of talk: Coming to terms with the 'new' scientific governance. *Social Studies of Science* 36(2): 299–320.

Irwin, Alan, Torben Elgaard Jensen, and Kevin E. Jones. 2012. The good, the bad and the perfect—Criticizing engagement practice. *Social Studies of Science.* doi:10.1177/0306312712462461.

Jackson, R., Fiona Barbagallo, and Helen Haste. 2005. Strengths of public dialogue on science-related issues. *Critical Review of International Social and Political Philosophy* 8(3): 349–358.

Jasanoff, Sheila. 2005. *Designs on nature: Science and democracy in Europe and the United States*. Princeton: Princeton University Press.

Jasanoff, Sheila, and Sang-Hyun Kim. 2009. Containing the atom: Sociotechnical imaginaries and nuclear power in the United States and South Korea. *Minerva* 47(2): 119–146. doi:10.1007/s11024-009-9124-4.

Joly, Pierre Benoit, and Alain Kaufmann. 2008. Lost in translation? The need for upstream engagement with nanotechnology on trial. *Science as Culture* 17(3): 225–247.

Jones, Kevin E. 2004. BSE and the Philips report: A cautionary tale about the uptake of 'risk'. In *The governance of knowledge*, 161–186. New Brunswick and London: Transaction Publishers.

Jones, Richard. 2008. When it pays to ask the public. *Nature Nanotechnology* 3(10): 578–579. doi:10.1038/nnano.2008.288.

Kearnes, Matthew B, and Arie Rip. 2009. The emerging governance landscape of nanotechnology. In *Jenseits Von Regulierung: Zum Politischen Umgang Mit Der Nanotechnologie*, edited by S Gammel, A Losch, and Alfred Nordmann. Berlin: Akademische Verlagsgesellschaft.

Kjølberg, Kamilla Lein, and Roger Strand. 2011. Conversations about responsible nanoresearch. *NanoEthics* 5(1): 99–113. doi:10.1007/s11569-011-0114-2.

Kleinman, Daniel Lee. 1995. *Politics on the endless frontier: Postwar research policy in the United States*. Durham: Duke University Press.

Law, John. 1994. *Organising modernity: Social ordering and social theory*. New York: Wiley.

Macnaghten, Phil, and Julia S. Guivant. 2011. Converging citizens? Nanotechnology and the political imaginary of public engagement in Brazil and the United Kingdom. *Public Understanding of Science* 20(2): 207–220. doi:10.1177/0963662510379084.

Macnaghten, Phil, and Richard Owen. 2011. Good governance for climate-engineering. *Nature* 479: 293.

Mahlouji, H., and N. Anaraki. 2009. Corporate social responsibility towards social responsible innovation: A dynamic capability approach. *International Review of Business Research Papers* 5(6): 185–194.

McCallie, Ellen, L Bell, T Lohwater, J Falk, Jane L. Lehr, B.V. Lewenstein, C Needham, and B Wiehre. 2009. *Many experts, many audiences: Public engagement with science and informal science education: A CAISE inquiry group report.* Washington DC: Center for Advancement of Informal Science Education (CAISE).

Mejlgaard, Niels. 2009. The trajectory of scientific citizenship in Denmark: Changing balances between public competence and public participation. *Science and Public Policy* 36(6): 483–496. doi:10.3152/030234209X460962.

Midttun, A., Gjolberg, M., Kourula, A., et al. (2012). Public policies for corporate social responsibility in four nordic countries: Harmony of goals and conflict of means. *Business & Society*. Available at: http://bas.sagepub.com/cgi/doi/10.1177/0007650312450848. Accessed 26 Feb 2013.

Miller, Clark A. 2008. Civic epistemologies: Constituting knowledge and order in political communities. *Sociology Compass* 2(6): 1896–1919. doi:10.1111/j.1751-9020.2008.00175.x.

Miller, Jon D. 1998. The measurement of civic scientific literacy. *Public Understanding of Science* 7(3): 203–223. doi:10.1088/0963-6625/7/3/001.

Miller, Steve. 2001. Public understanding of science at the crossroads. *Public Understanding of Science* 10(1): 115–120.
Morris, Jeff, Jim Willis, Domenico De Martinis, Bjorn Hansen, Henrik Laursen, Juan Riego Sintes, Peter Kearns, and Mar Gonzalez. 2010. Science policy considerations for responsible nanotechnology decisions. *Nature Nanotechnology* 6(2): 73–77. doi:10.1038/nnano.2010.191.
Nye, David E. 2004. *America as second creation: Technology and narratives of new beginnings*. Massachusetts: MIT Press.
Owen, Richard, David Baxter, Trevor Maynard, and Michael Depledge. 2009. Beyond regulation: Risk pricing and responsible innovation. *Environmental Science and Technology* 43(18): 6902–6906. doi:10.1021/es803332u.
Owen, Richard, and Nicola Goldberg. 2010. Responsible innovation: A pilot study with the U.K. Engineering and Physical Sciences Research Council. *Risk Analysis* 30(11): 1699–1707. doi:10.1111/j.1539-6924.2010.01517.x.
Pellizzoni, Luigi. 2004. Responsibility and environmental governance. *Environmental Politics* 13(3): 541–565. doi:10.1080/0964401042000229034.
Roco, Mihail C. 2001. From vision to the implementation of the U.S. national nanotechnology initiative. *Journal of Nanoparticle Research* 3(1): 5–11.
Roco, Mihail C., Barbara Harthorn, David Guston, and Philip Shapira. 2011. Innovative and responsible governance of nanotechnology for societal development. *Journal of Nanoparticle Research* 13: 3557–3590. doi:10.1007/s11051-011-0454-4.
Royal Society. 1985. *The public understanding of science*. London: The Royal Society.
Shapira, P., and J. Wang. 2010. Follow the money. *Nature* 468(7324): 627–628.
Shelley-Egan, Clare. 2009. Mapping ethics in the real world of nanotechnology. In *Size matters, legal and social aspects of nanobiotechnology and nano-medicine*, ed. J.S. Ach, and C. Weidemann, 147–155. Berlin: LIT Verlag.
Shamir, Ronen. 2008. The age of responsibilization: On market-embedded morality. *Economy and Society* 37(1): 1–19. doi:10.1080/03085140701760833.
Stilgoe, Jack. 2012. Taking care of the future—The imperative of responsible innovation. *People and Science* 22.
Stirling, Andy. 2008. 'Opening up' and 'closing down': Power, participation, and pluralism in the social appraisal of technology. *Science Technology and Human Values* 33(2): 262–294. doi:10.1177/0162243907311265.
Sutcliffe, H. 2011. *A report on responsible research and innovation*. http://www.matterforall.org/pdf/RRI-Report2.pdf.
Technology Strategy Board. 2012. *Responsible innovation framework*. London: Technology Strategy Board.
Thorpe, Charles, and Jane Gregory. 2010. Producing the post-fordist public: The political economy of public engagement with science. *Science as Culture* 19(3): 273–301. doi:10.1080/09505430903194504.
Vallentin, S., and Murillo, D. 2012. Governmentality and the politics of CSR. *Organization* 19(6): 825–843.
Von Schomberg, Rene. 2011. *Towards responsible research and innovation in the information and communication technologies and security technologies fields*. Brussels: European Commission Services, Directorate General for Research and Innovation.
Wilsdon, James, and Rebecca Willis. 2004. *See-through science: Why public engagement needs to move upstream*. London: Demos.
Wynne, Brian. 2006. Public engagement as a means of restoring public trust in science—Hitting the notes, but missing the music? *Community Genetics* 9(3): 211–220.
Wynne, Brian. 2007. Public participation in science and technology: Performing and obscuring a political conceptual category mistake. *East Asian Science Technology and Society An International Journal* 1(1): 99–110.

Chapter 4
Exploring Responsible Innovation as a Guiding Concept: The Case of Neuroimaging in Justice and Security

Marije de Jong, Frank Kupper, Anneloes Roelofsen and Jacqueline Broerse

Abstract Although responsible innovation (RI) is to change the scientific system to meet the grand challenges of our time, its criteria are still unclear. This study explores meaning negotiation on RI by academic actors in both formal and informal discourse. In the formal discourse on RI, we identified characteristic shifts, including engagement of societal stakeholders, anticipation and adaptability, broadening, and new insights on impacts and regulation. However, the intended group of scientists themselves were not visibly involved in the formal discussion. Therefore, as a case study, we studied the informal discourse of Dutch scientists in the field of functional neuroimaging relevant to the domain of justice and security. Our findings show that RI is unfamiliar to scientists. We suggest that RI as a guiding concept is in need of operationalization within the specific context in which it is used. A point of entry for such a process is the role responsibility taken on by scientists. Resistance can be expected as RI can be equivocated with a limitation on the autonomy of science. To avoid evasive practices, extra efforts are needed to involve the scientists in a co-constructive process to operationalize RI.

Keywords Responsible innovation · Neuroimaging · Security · Justice · Science and technology

M. de Jong (✉) · F. Kupper · J. Broerse
Athena Institute, Faculty of Earth and Life Sciences, VU University Amsterdam,
de Boelelaan 1085, 1081 HV Amsterdam, The Netherlands
e-mail: marije.de.jong@vu.nl

A. Roelofsen
KWF Kankerbestrijding, Delflandlaan 17, 1062 EA Amsterdam, The Netherlands

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_4

4.1 Introduction

From the year 2000 onwards, 'responsible innovation' (RI), also known as 'responsible research and innovation', has gained increasing attention in the science policy domain in the USA, the European Union, the UK, Belgium and the Netherlands (Barben et al. 2007). Public funding bodies have begun to require grant applicants to identify societal and ethical issues in their proposals in an attempt to stimulate research and innovation activities that take social impacts into account and facilitate desirable outcomes (Rip 2005; Owen and Goldberg 2010). This trend represents a political redirection of science, the natural sciences in particular. This emphasis on RI is consistent with the pivotal role bestowed on science-driven innovation in the European strategy to overcome sustainably the present financial and economic adversity (EC 2010, 2011a). Nevertheless, it is still hard to pinpoint what is the status of RI. Should we consider it a transcendent philosophy or an inspiring ideal? Is it a fundamental concept, an ultimate manner of behaving, or a perfect process? If RI is to be effective as a guiding principle in science, there needs to be a greater, common understanding of what it means in terms of concepts and methodologies. von Schomberg (2012) of the European Commission's Directorate General for Research, Governance and Ethics proposed the following general working definition of RI:

> [It] is a transparent, interactive process in which societal actors and innovators become mutually responsible to each other with view on the (ethical) acceptability, sustainability and society desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technologies advances in our society.)

The term "working definition" suggests that, despite the increasing emphasis on RI from the policy domain, the term has not yet been "stabilized" in the policy domain as an empirical phenomenon. From the perspective of actor network theory (ANT) (Law 1992; Callon 1986; Latour 1987) and social construction of technologies (Bijker et al. 1984), the conceptual delimitation and meaning of RI is still under negotiation. A central tenet in constructivist theory is that there are no unambiguous truths because objective reality cannot be separated from human perceptions. Instead, meaning is socially constructed and negotiations involve the construction and deconstruction of statements, until this leads to stabilization (Bergquist et al. 2001). Meaning is thus an internal construction of a social group (Brown and Duguid 1996).

The elusiveness of RI makes it difficult to make sense of its substance. This study is concerned with the way in which the meaning of RI is being negotiated by actors in the academic arena. The negotiation of meaning of RI is examined in two arenas of academic discourse: a formal and an informal one. First, the formal academic discourse as it takes place in academic documents was examined. Within constructivist theory, documents play an important role because they are, at the same time, both immutable and mobile (Latour 1990). Documents inevitably surpass the reach of their respective authors and publishers, potentially transgressing social boundaries and entering new arenas. Through constructions of peer review and refereeing, academic documents provide a shared medium for collective sense-making (Bergquist et al. 2001; Brown and Duguid 1996). This process of collective sense-making involves academics for whom this is a subject area of interest, and not necessarily the scientists whose practices are potentially targeted by the greater emphasis on RI from the policy arena. Our starting point therefore is to look at the way RI is conceptualized in the general academic literature. Studying meaning negotiation in academic writings does require attention to both explicit and more implicit formulations, as the body of formal literature on RI was only to a little extent primarily concerned with RI.

Second, we examine the informal discourse among scientists, whose future practices might be altered by the redirection of science. In this, we take a case study approach, focusing on the application of functional neuroimaging in the domain of justice and security. There are several reasons for using a case study approach focusing on the case of functional neuroimaging in the domain of justice and security. First of all, the addition of a case study will provide us with an additional methodological means to look at the discourse on RI, yielding either complementary or confirming data. Second, a large part of the Dutch research in the field of functional neuroimaging relevant to the domain of justice and security is funded by the program "Brain and Cognition: societal innovation".[1] The focus on societal innovation and the explicit inclusion of private or public social partners in the projects indicates that the redirection of science becomes discernible here. Third, we suspect that there is a context-specific element to RI, as what can be considered "responsible" can depend on a variety of contingencies. For example, the prevailing norms and values in a certain field of application, historical developments or social interactions within and between the parties involved. Functional neuroimaging, as it has been developed and used over the past twenty years in the fields of cognitive neuroscience, experimental psychology and neuropsychology, monitors brain function to understand the relationship between brain structure and mental functions. As such, it touches upon fundamental human experiences, such as thoughts and emotions, and beliefs about human nature and identity. Its application in the domain of security further heightens associated ethical issues and societal concerns. Areas of concern relate to, for example, changing ideas about criminal responsibility, determinism and human agency which could paralyze the foundations of the justice system (Goodenough and Tucker 2010; Greely 2007); or privacy issues in relation to the brain. Neuroimaging in itself combines a visual and a scientific component which makes it already fascinating and controversial. As many preceding visualizing technologies, it can make "real" what was previously unseen. Petchesky (1987) has illustrated this for the visual representations of the unborn fetus offered by ultrasound sonograms and their impact on the controversy surrounding the abortion issue. Furthermore, technologies mediate our quality of life, moral actions and decisions in uncertain ways (Verbeek 2011). If we assume RI to be context-specific, it also follows that we want the involved actors

[1] www.hersenenencognitie.nl/contents/1038?locale=en.

themselves to negotiate its meaning. This last point is not only based on a democratic point of view, but also on a pragmatic one: If the redirection of science associated with the framework of RI is to change the scientific system, this will eventually also necessitate a change in the behavior of individual actors. Therefore it is sensible to study how scientists themselves negotiate the meaning of RI in their informal discourse within their own practice. This is what we aim for in the research project we are part of, "Neurosciences in Dialogue",[2] which focuses on the responsible use of neuroimaging in society, among which the domain of security and justice (the other two domains in this project are health care and education, for the latter see also Chap. 8). The framework of RI seems promising as a guiding principle in this project in which researchers and societal stakeholders are involved in an interactive learning process to stimulate developments towards shared desirable applications with few, or at least manageable, negative impacts. As practitioners investigating the value of RI as a guiding concept in a project on the societal embedding of neuroimaging in the domain of justice and security we will want to know how the scientists at the forefront of these developments make sense of such a concept. Both the formal and informal discourse is part of a novel process of negotiation of meaning as the framework of RI which originated in the policy arena is being applied in the academic arena (see Chap. 3 for a more detailed picture of how RI is imagined in policy).

The aim of this chapter is to explore RI as a guiding concept in a project on societal embedding of neuroimaging by looking at RI meaning negotiation in formal and informal discourse. We will conclude with recommendations on how to operationalize RI as a guiding concept, and with warnings for evasive practices that can be expected in this process.

4.2 Methodology

Two research methods were used: a literature review to consider the formal discourse in general academic documents; and interviews to examine the informal discourse among neuroscientists.

4.2.1 Literature Review

4.2.1.1 Data Collection

A systematic literature search was conducted using academic literature databases (Google Scholar, Web of Knowledge, Scopus, Proquest and EBSCOhost)

[2]This research project at the VU University Amsterdam is funded by the Dutch NWO program "Responsible Innovation".

to identify articles published prior to August 2012. References were collected by using the keywords “responsible innovation” in the titles (Google Scholar), topic and title (Web of Knowledge), title, abstract and keywords (Scopus), in all text fields (Proquest including only peer reviewed journals and EBSCOhost including all databases but excluding trade publications). When results did not yield peer-reviewed academic journal articles, it would be included if it concerned: white papers whose author has previously published peer-reviewed articles; contributions to international conferences. In the latter case, authors were asked for the most recent version of their writings. After removal of duplicates, 97 articles were identified (See Appendix). A total of 26 articles were excluded for various reasons: articles were not available to the authors (full-text or abstract) (−3); lacked “responsible innovation” in the body text (−9); were printed in journals lacking peer-review (−5); or did not provide enough context to interpret the concept of RI (−9). The remaining 71 sources, which included full text articles, abstracts and presentations, were analyzed for the context in which RI was conceptualized.

4.2.1.2 Data Analysis

All documents were read and coded inductively by one researcher using qualitative analysis software (Atlas.ti 6). Analysis of these codes by the first and second author led to the formation of code categories (RI conceptualizations general, RI conceptualizations product, RI conceptualizations process, RI innovator characteristics, RI challenges, RI’s targeted sectors and technology types) after which all documents were recoded by the first author (in MaxQDA 10[3]). All documents describing particular sectors or technology types were grouped in sets. These sets, as well as author affiliations, were converted to document “variables”. The sets and variables were used for further analysis using the functions of “code relations”, “typology table”, “crosstabs” and “variable lists”.

4.2.2 Interviews

4.2.2.1 Data Collection

A total of twenty qualitative, semi-structured interviews were conducted with scientists to explore how meaning is negotiated on RI[4] in 2011–2012. Semi-structured interviews allow for an in-depth discourse, while providing enough structure to acquire data on pre-established topics.

[3]The change in the software used was not a methodological choice but a pragmatic one, with respect to user friendliness.

[4]In Dutch ‘maatschappelijk verantwoord innoveren’, which can be translated as “societally responsible innovation”.

There are very few scientists in the Netherland working on neuroimaging in the domain of justice and security. We also included a couple of scientists that work on research concepts that could well be of interest of the domain of justice and security, but who were not formally involved in projects targeting this particular domain. In a separate extensive study of the international academic literature on neuroimaging in the domain of justice and security, we identified important research concepts. These concepts were then matched to scientists using neuroimaging technologies and based in the Netherlands. In selecting the interviewees, we aimed to cover the diversity of concepts as much as possible for maximum variation sampling (Patton 1990). The interviewees differed in academic position: most were professors (50 %), others were researchers with a post-doctoral position (40 %) or researchers working on their PhD (10 %). They were affiliated to nine different universities in the Netherlands, and worked in the disciplines of cognitive neurosciences, psychology and neuropsychology, psychiatry and forensic psychiatry, radiology and neurology.

In addition to providing an overview of the state of the art of neuroimaging in the field of justice and security, the study of the literature on neuroimaging in the domain of justice and security allowed for scientifically informed interviewing (Laudel and Gläser 2007). The scientists were asked what RI means to them and when an innovation could be considered responsible in their view. The intention was to use a minimum degree of steering to allow for new viewpoints to emerge. The interview questions were a part of broader semi-structured interviews focusing on the scientist's research: ideas about the future of their research field; the benefits and barriers they expect to come across; and their thoughts on worrying applications of neuroscience, as well as how to mitigate them. The interviews lasted between 45 and 90 min, and were carried out at the workplace of the participant.

In this case-study approach, the interviewed scientists were selected when they conduct neuroimaging research with potential applications in the field of justice and security, as this is important to the wider project this study is part of. The findings can therefore not be formally generalized to other populations of scientists. However, the themes that surface concern the wider scientific community as well, and the findings can prove as a starting point for theory development (Flyvbjerg 2006; Yin 2002).

4.2.2.2 Data Analysis

All interviews were audio-recorded with consent and transcribed verbatim to secure an accurate account of the discourse. For each interview, a summary was written and sent to the participant to check for the correct interpretation by the researcher, thereby enhancing internal validity of the interview data. The transcripts were inductively coded, using qualitative analysis software (MAXQDA 10). During the coding, extra attention was paid to words indicative of normative aspects of science, duties, and own roles, for example "have to", "must", "should", "role", "task", "as a scientist/researcher" and "responsible/responsibility". During the analysis, other words also surfaced as being important, for example "political choices" and "added value". Throughout the analysis, the authors sought divergent views in order to acquire rich descriptions of scientist's conceptualizations of RI.

The results were then compared to the results obtained from the literature review. However, it is worth noting that these are very different kinds of data and caution needs to be exercised in such comparisons. The selection criteria were very different for the academic authors and the interviewed scientists. The first were selected on the basis of their familiarity with RI, the latter on their involvement in neuroimaging. The academics contributing to the formal literature have, by definition, given RI serious thought, while the interviewed scientists may not have thought about RI before. The production of formal literature is a process of months, entailing the interaction between an author, editor and reviewer about the content, but it is often also a co-production by colleagues within the same or adjacent fields. In writing the academic articles, the respective authors have already entered into a process of meaning negotiation on RI. On the other hand, the interviewed scientists are asked to formulate impromptu answers in relative isolation from their peers and in the social setting of an interview.[5] A scientifically informed interviewer is then better capable of creating similar conditions to mimic this process of meaning negotiation with peers, for example by creating a common language for the interviewee to take the interviewer seriously enough to give in-depth answers. However, responses may still lack detail, as scientists might draw answers from pre-existing repertoires that are plausible, appropriate and sufficient to them, yielding stereotypical answers (van Lente 2012). However, this is in itself important data as it nonetheless gives insight into dominant ideology or mainstream thinking (Diefenbach 2009). Furthermore, even if no meaning is attached to "responsible innovation" in their daily practice, interviews can still provide insights into scientists' initial attitude to this concept originating from the policy domain and potentially invading their future practice.

When this study was started, it was chosen to look for "responsible innovation" as a search query. In the meantime, we became aware that "responsible research and innovation" (RRI) is another formulation that is gaining momentum in more recent texts. It could be that responses from the interviewed scientists could have differed if RRI was introduced instead of RI. We may have ignored similar concepts in adjacent fields described with different terms. In our approach we focused on the discourse in meaning negotiation. However, we did not look at dynamics in communication, or at the practices associated with the discourse. This could provide valuable additional information about meaning negotiation on RI.

4.3 Results

4.3.1 Negotiation of Meaning in the Formal Context

Despite the fact that the policy domain considers the framework of RI relevant to science, the authors of the 71 sources were usually not directly involved in the natural

[5] All but one interviews were conducted with a single scientist.

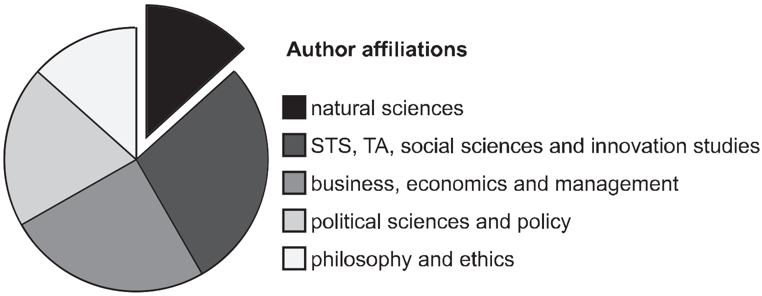

Fig. 4.1 Proportions of the author affiliations (The pie-chart does not give absolute percentages, as for several sources authors had multiple affiliations)

sciences. Only in 1 of 9 articles the authors were primarily affiliated to the natural sciences (see the slice lifted out of the Fig. 4.1. None of them were concerned with neuroscience). Most authors were involved in science and technology studies (STS), innovation studies and social science (1 in 4), business, economics and management (1 in 5), political sciences and policy (1 in 6) and philosophy and ethics (1 in 9).

Different aspects of RI are the subject of meaning negotiation within the formal literature. First, the nature of RI is an area of attention. Second, process and product dimensions of RI surface, as well as the individual innovator. Third, there is a variation of sectors and technology types where RI is deemed relevant. Fourth, the challenges facing RI were emphasized. Each of these aspects is described below.

4.3.1.1 The Nature of RI

Various authors find the concept "broad" (e.g. Ferrari and Nordmann 2010) or "passive" (Pavie 2012). In spite of this described lack of clarity, certain shifts resurface within the gathered literature that are supposed to bring about RI, including engagement of societal stakeholders, the broadening of considerations and disciplines during innovation processes, new insights on societal impacts, and processes of anticipation and adaptation.

Engagement of societal stakeholders

RI is characterized by inclusion of societal stakeholders throughout the entire innovation process, not only downstream decision-making (e.g. Barré 2011). Engagement with the public and other stakeholders is thereby observed to be moving 'upstream', shifting the focus from technology development to science and scientific agenda setting. Engagement of non-scientific stakeholders was also talked of as a process in which responsibility for the innovation is shared between the different parties involved. The ones inviting the other stakeholders do not remain the sole locus of responsibility.

Broadening

In the literature, RI is seen as 'broadening' traditional innovation processes, involving more types of innovations, a wider group of scientific disciplines, a more diverse group of societal stakeholders, and a widening of the issues taken into account. This was a dominant theme within the literature. First, a number of authors consider that more types of innovations should be considered as responsible innovations, not just purely technological ones. Although a considerable part of the literature focuses on technological innovations, RI was also said to involve other types of innovations in order to avoid a technological fix for societal problems that are not technological in nature. Behavioral, organizational or cultural innovations were mentioned in this context (e.g. van den Hove et al. 2012; Ingham et al. 2010; Van Oudheusden forthcoming). Second, a wider group of scientific disciplines should be involved during the research process. The literature generally refers to the involvement of social scientists and the humanities in innovation projects, typically dominated by the natural sciences (e.g. Kiran 2012; Reddy et al. 2011). Third, a more diverse group of societal stakeholders with a stake in the innovation process should be involved. The literature cites both non-scientific stakeholders (lay people, citizens, customers, patients) and professional stakeholders (NGOs, industry, insurance companies, expert practitioners, public health agencies, government). Fourth, a widening of the issues to be considered during innovation processes is also seen as part of RI. In current processes of research and innovation, techno scientific and economic issues are high on the agenda. More attention should be paid to wider issues related to innovation processes and their resulting applications, such as ethical, social, societal, environmental, scientific, health, legal, cultural and political issues. Ethical, social, societal and environmental issues received the most emphasis.

Impacts and regulation

Many articles argue that RI involves consideration of societal impacts, although the realization of potentially positive impacts received more emphasis than the prevention of negative impacts (risks). What constitutes a positive impact was found to vary, ranging from innovations having a human purpose or fulfilling a need (e.g. Groves et al. 2011), to outcomes being socially or environmentally sustainable (e.g. Venier 2011) or robust (e.g. Ozdemir et al. 2011). Ultimately, innovation should contribute to the 'public' or 'collective good' (Guston 2006).

The literature described both systemic and technology-based impacts. Risks were generally formulated at a product level, focusing on the outcome of the technology, and discussed in conjunction with (risk) governance and the precautionary principle which is increasingly dominant in Europe, Although not an important topic of conversation in the identified sources, it is important to note here that there is quite some interpretative flexibility how deal with scientific uncertainty under the framework of the precautionary principle. In its weakest reading, it entails that a lack of decisive evidence of harm cannot be used as a reason not to regulate; a more far-reaching interpretation entails that the burden of proof that an application is not harmful lies with the promoters of the development and has a lower threshold for harm and damages (Jonas 1984; Sunstein 2002).

A regulatory framework is seen here as a key enabler for strategic decision-making, and thus towards leading to positive impacts and preventing risks (e.g. D'Silva et al. 2012). However, some legislation and regulations were described as being counterproductive, for example product liability law (Schwartz 1992) and embryo protection regulation (Dondorp and de Wert 2011). Authors argue for more flexible forms of legislation and regulation (e.g. soft law, co-regulation, self-regulation and voluntary reporting schemes) and more adaptive and anticipatory governance (e.g. Ferrari and Nordmann 2010; D'Silva et al. 2012; Barré 2011; Kaza 2006). Nevertheless, public agencies and governmental bodies need to retain an important facilitative role (e.g. te Kulve and Rip 2011).

Anticipation, adaptability and the temporal dimension

RI has an explicit temporal dimension as it takes a longer perspective, earlier on in the process. The temporal dimension of RI also itself in anticipating potential impacts and considerations early on in the process, and intervene on the basis of this acquired knowledge in the design stage. Upstreaming is therefore also observed here. The action in this feedback loop is also called 'steering' (e.g. Ozdemir et al. 2011) or 'modulation' (e.g. Owen and Goldberg 2010; Fisher 2011). Modulation, which expresses a similar idea as the also used concept of 'integration' (e.g. Schuurbiers 2011; Hellström 2003), is especially seen in the context of the research lab. First and second order learning (learning on the level of expectations and outcomes, or on the level of the underlying values and assumptions, respectively) among researchers is then aimed to take place to the benefit of public matters related to research activities (e.g. Schuurbiers 2011; Goorden et al. 2008). Both modulation and steering set forth an adaptive dimension of RI.

4.3.1.2 Process, Product and Innovator Characteristics

The four mentioned characteristics of the nature of RI show that RI was not only conceptualized on the level of the (product) outcomes of concrete applications. It was also substantiated with regard to the process of research and innovation and the individual innovators. Process and product related aspects of RI were most frequently put forward. The former is most prevalent, and pays attention to all stages of the innovation process, but especially the design phase. There should be a receptive attitude to new information from other perspectives and viewpoints in decision-making, thereby altering the trajectory. Meaning negotiation on the product aspects of RI includes more moral judgments about the application and its impacts. There is a focus on the purpose of the product. Not only moral rules, but especially principles and values are put forward as criteria for the product and impacts. The product and its impacts are shaped by contingent factors. Values are more flexible in accommodating this uncertainty than rules.

Noteworthy is that there was less focus on the characteristics of the individual innovators, compared with product or process aspects of RI. In those articles where innovator characteristics are linked to RI, behavioral, relational and cognitive aspects of innovator characteristics are put forward. For example, being independent (e.g. Stinner et al. 1992) and having self-control (e.g. Pavie 2012) contribute to RI by enabling decision-making in freedom. Innovators should also take on liability (e.g. Armstrong et al. 2012), be transparent and reciprocal in their interactions with other stakeholders (e.g. Chervenak and McCullough 2006), be reflexive and act according to internalized values (e.g. Goorden et al. 2008).

4.3.1.3 RI's Targets

RI is being discussed on different levels, and substantiated for a variety of sectors (e.g. science, engineering, health, industry, education, finance) and technology types (e.g. nano, ICT, obstetrics and digital innovations, but not neuroimaging) (Shelley-Egan 2010; Tyl et al. 2011; Guston 2007; Chervenak and McCullough 2006; Duke 1978; Stahl 2011; Armstrong et al. 2012). Most of the sources concerned science and its related technologies, indicating that this is the primary target for the framework of RI. Within these sources, formal discourse primarily targeted changes on the technical system level or the level of applications. For the former, RI is seen to necessitate transformation of the technology development embedded in its wider system of, for example, practices and institutions. For the latter, the focus of RI was on a responsible product, with respect to their implications and risks. Individual scientists themselves were not the primary target of demands for RI. Interestingly, the formal discourse extends to quite different sectors or application types, notably the previously mentioned finance sector and related financial products, thereby showing the potential of RI to incrementally widen its reach.

Depending on the sector or technology type, different aspects were emphasized in RI meaning negotiation, which might be indicatory of RI having context-specific aspects. A central theme in the literature on nanoscience and nanotechnologies were alternative regulation models that can be employed early in the development to enable the responsible development of nanoscale technological applications. Typically mentioned risks are the safety of products containing nanoparticles during its life cycle for humans and environment (e.g. Ishizu et al. 2008). Promises of potential revolutionary benefits are problematized, as this may invite a discourse of hype, unrealistic expectations and disappointment (e.g. Shelley-Egan 2010). In articles on healthcare and clinical research, however, the focus was on problems with regard to scientific rigor and the dealings with human subjects as innovation processes lack the clear guidelines of regimented clinical trials. On a different note, financial innovations were typically linked on the product level to sustainability against the backdrop of the financial crisis (Armstrong et al. 2012).

4.3.1.4 Aims of RI

The aims of RI were framed with respect to various challenges that need addressing and the obstacles that are urgent to be overcome. For one, RI is to avoid negative impacts as seen in the past. Worries about technology-induced risks and controversies surface by referring to previous or ongoing crises, and the use of dramatic words such as disasters, failures and chaos (e.g. van den Hove et al. 2012). Second, RI is hoped to bring about a rebalancing between financial and economic performance of innovation on the one hand, and alternative valued performance, such as societal benefits and environmental protection, on the other hand (e.g. Tihon and Ingham 2011). In trying to better integrate social and environmental issues with financial and economic performances, reflexivity on social and environmental issues should be of guidance rather than slowing the process down or restraining action. Ideally, RI should open up new opportunities for innovation and work as a 'positive force'. A third major challenge is the asymmetry between the speed of innovation and the assessment of the impacts of the innovative products or the regulation thereof (e.g. Owen 2009). This is manifested in a time lag between research and innovation, the resulting products, its impacts on society and the regulation of these impacts. During this time lag, the technology can become entrenched in society, which makes it hard (or even impossible) to retract. Furthermore, in this period damages to humans or the environment can occur. Fourth, due to the global nature of innovation, irreversibilities and unintended consequences are not bound to the innovation's location of origin. Fifth, RI is to handle uncertainties and ignorance with respect to outcomes of the innovation process (e.g. Barré 2011). For emerging technologies especially, it is not only that it is uncertain what the details and the extent of the possible adverse effects are, or that knowledge is incomplete. There is ignorance about the range of the possible adverse effects themselves and knowledge is unsure. Lastly, RI is thought to increase public trust and acceptance, which is key `for successful development of technologies (e.g. Ishizu et al. 2008). In this line of argumentation, 'doing' RI can lead to a competitive advantage, a better reputation or a higher acceptance of a technology, thereby minimizing chances of failure.

This indicates that the RI framework is to prevent techno-disasters of the past; to deal with the uncertainties, ignorance and unintended consequences associated with research and development and its irreversibilities when intervening downstream; and to yield more successful innovations. RI is expected to achieve this, firstly, when research system is broadened in multiple ways: for example the issues considered during the technology development, most importantly towards societal and ethical issues, and the inclusion of social scientists and professional and societal stakeholders during the research and development process. Secondly, more attention is needed for the distribution of impacts, and for the framing of impacts in terms of a contribution to the collective good, instead of economic and techno scientific terms only. Thirdly, it would require 'upstreaming' with respect to the engagement with publics and other stakeholders and to anticipating and intervening in technology development. Notably, in applying these shifts necessary for RI, not only (product) outcomes of concrete applications should be considered, but also the process of research and innovation. Besides these recurring characteristics of RI, different aspects of RI were highlighted for different contexts. Although this could be indicative of different groups of authors negotiating meaning of RI in different contexts, it could also point toward RI having a context-specific element as well. Furthermore, RI was not only seen to be relevant to the application level, but also on the system level. Interestingly, the negotiation of meaning taking place in literature was hardly done by natural scientists themselves, the ones currently at the receiving end of the redirection of science.

4.3.2 Informal Negotiation of Meaning

As mentioned above, neuroimaging and the wider field of neuroscience did not yield articles on RI, which means that neuroscientists were not visibly involved in the formal discussion within RI. However, as the scientists' research system might change under the influence of the RI framework, the way this group understands RI may play a role in the successful implementation of the RI framework in practice.

4.3.2.1 The Nature of RI

RI did not seem to be a concept with which the scientists were familiar. Most scientists had trouble elaborating on what RI means to them as a concept, as indicated for example by long pauses before answering, a low diversity in RI aspects proposed per scientist, low articulateness, and the elaboration on its ambiguity instead. Some scientists tried to answer by dissecting the wording of RI ("maatschappelijk verantwoord innoveren" or "societally responsible innovation" as it is used in Dutch). These scientists felt comfortable mentioning societal benefits and the novelty aspect of "innovation", but then had difficulties elaborating on the "responsible" part. However, some scientists were quite articulate and specific and mentioned a variety of aspects they thought relevant for RI.

Quite some scientists were skeptical about RI (20–30 %). It was said, for example, that people should not be 'too idealistic', that the notion of RI "does not correspond with the real world", or that the terminology is "nothing more than words managers or politicians use" when wanting to force scientists to deliver quick results in a perverse way.

Engagement of societal actors

The engagement of societal actors and other stakeholders during the research and innovation process was scarcely linked to RI by the scientists. In fact, inclusion of non-scientific stakeholders was not unequivocally supported. Some estrangement from society was voiced by a skeptic, in conveying that if society would finally start paying attention to what science is saying, maybe science could start listening to society. The inclusion of practitioners at the grassroots level was lauded, but opinions varied quite dramatically about including other types of societal stakeholders, such as the public or patients.

> How do you expect these people, society, to have a voice in the innovation process? (…) Then money would go to sophisticated vending machines, or nicer TVs. Society couldn't care less that we develop things that are for the benefit of medicine. People who play soccer every weekend, and hundreds of thousands of them watching it on TV. [Soccer players] earn more than those who are interested in innovation and trying to help others. Thát is society. They weigh it that way. So I do not expect that much of society as a voice in innovation.

Nevertheless, other scientists were open to working with other stakeholders.

> Yes, [the interaction with society] should be continuous, but it is not. I mean, it only takes place during certain stages of your research. That you found something out, and you want to know what people think of it.

Broadening

The dominant theme of broadening was only partly found in the informal negotiation of meaning by the scientists. Broader-than-scientific dimensions were the societal benefits of knowledge production and applications, as well as ethical considerations of the research practice and applications. Other types of broadening, including other scientific disciplines, the relevance of non-technological innovation, or other types of impacts and considerations, were hardly mentioned as being part of RI. Technical, scientific and economic considerations had a considerable weight in the discourse.

Impacts and regulation

The societal benefits of applications were also seen to depend on the context of implementation. On the one side it was argued that one has to carefully analyze the context of application and tailor the application to ensure its optimal and correct use. Examples are to provide training for the people who will have to use the technology, and to add extra material features so that the device will be correctly placed on the human body. On the other hand, it was argued that a certain innovation could be seen as ethically correct in one field of application, such as (mental) health care, but not in another, such as neuromarketing. However, this type of argumentation often included a techno-neutral type of reasoning in which only humans determine the technology's use and benefits, and neglects non-human forms of agency. These scientists felt that this was not their choice to make but that it is the responsibility of policy-makers. Hence, a division of moral labor was encountered here, indicating that scientists see only a marginal role for themselves in RI. On a similar note, scientists argued that an innovation can be considered responsible if it provides an evidence-base, which also implies that innovation is seen as a value-neutral enterprise.

> But the government should apply informed decision making. The government must make decisions based on the best available information. As a scientist, you should contribute to this. And I consider that responsible innovation.

Anticipation, adaptability and the temporal dimension

The academics' formal meaning negotiation stressed the importance of long-term visions with respect to impacts, and therefore the need of anticipation and adaptation beginning in its early stages and continuing throughout the process. Anticipation and adaptability are thus important concepts when locating RI upstream. The interviewed scientists, however, located RI predominantly downstream: RI has to do with the implementation of a technology and not with the process before it. Even then, you can anticipate problematic issues with respect to the implementation of a technology. The scientists found this hard to do, due to the high uncertainty and ignorance of applications of research and innovation activities as well as the application's downstream impacts. This issue was sometimes deflected by putting trust in science as an institution or the scientist as a professional. For one, the importance of the good intentions of embarking upon a certain scientific endeavor was stressed. Secondly, it was stated that science has an inherently societal role, for example by way of knowledge production and understanding, and that science is not performed as "l'art pour l'art", but in the hope that it will be applied one day.

When talking about the issue of RI, it was located downstream. However, this does not necessarily mean that the same can be said for the scientist's *practice*. For example, when talking about a non-related subject, one scientist discussed how he involves grassroots level practitioners upstream, in formulating research questions, and deems this of high importance. It might not be unfamiliar to the scientists, but the framework of RI is just not seen in this light.

4.3.2.2 Process, Product and Innovator Characteristics

In contrast to what we found in the formal discourse, product dimensions were spoken of more than process dimensions of RI. The societal benefit of applications was again an important example of a product dimension of RI. The process element of including other stakeholders during the research and innovation process was scarcely linked to RI. Interestingly, innovator characteristics did surface here, in the form of researcher responsibilities. Especially the non-skeptics related RI to responsibilities they see for themselves as scientists and to ideal type researcher characteristics. These responsibilities and characteristics differed among the scientists. A dominant type of argumentation was in line with the Mertonian norms of 'pure science' and the ethos of the independent Republic of Science (Polanyi 1962). RI was linked to research that should be executed as well as possible by careful and critical analysis and interpretation. Social processes within the research group, helping and discussing research with each other, were said to enable making research choices. Research can then provide an evidence-base of knowledge and understanding, which was also seen as responsible. In the process, scientists should be transparent and open, and should not succumb to raising unrealistic expectations in the competition for research funding or publications in prestigious journals. However, they sometimes find this task difficult.

> You can't write everything that you find important. I had written such a piece, that it is a good thing that (…) we can only [measure] thoughts that the patient really wants [to share]. But that is then taken out.(…) The scientist is responsible to reflect upon things (…) I might warn [for certain uses].(…) But that is often seen as not so important.

Another typical stance of scientists was their perceived duty to educate the public, especially about the possibilities and the limitations of technologies via the popular press. Noteworthy is the prevalence of the deficit model among scientists with respect to public communication. Hence, a social contract is formulated in which the public has a right to expect that scientists will contribute to the public's understanding, knowledge and education. But the public is not to interfere in the process of science and try to steer science's impacts as research benefits from serendipity. A fear of infringement upon the sovereignty of science was exemplified by the shared concern of the scientists of the possibility that this notion of RI will constitute a brake on science. When talking about a potentially problematic direction of research:

> I don't think there should be barriers to science, or…. Of course, there are ethical limits to what you can do. But basically you would be able to think in that direction.

Secondly, one strand of argumentation revolved around not harming human research subjects in the quest for knowledge, especially if it concerns vulnerable patient groups, such as those with 'locked-in syndrome'. Other examples were that researchers should be aware of bias and fallacies in their decision-making, and the duty of scientists to look at more aspects of their research than technological and economic ones: ethical aspects in particular. However, it was also mentioned that (some) ideal type characteristic and researcher duties were endangered by (financial) pressures on researchers.

The duties and characteristics of researchers got more attention than in the formal meaning negotiation process described previously. But the dominance of the ethos of the independent Republic of Science shows that these responsibilities are for an important part placed within a framework that strives for excellence rather than impact. Societal and ethical dimensions are not (yet) integral part of it.

4.3.2.3 RI's Targets

As the scientists were asked to put the framework of RI onto their own research or research field, other sectors or technology types were not discussed as possible fields of application. Considering the difficulties encountered in conceptualizing RI, and the dismissal of the concept of RI altogether by some, raises the point that scientists often did not see their research or research field as a particularly suitable target for RI.

Considering the wide range of fragmented conceptualizations of RI, we could not identify a coherent and focused perspective on what the framework of RI would mean for the domain of justice and security in particular. The only consensus seemed to be on the level on which RI was to be approached, being the application level.

4.3.2.4 Aims of RI

The unfamiliarity of the interviewed scientists was illustrated by a low degree of negotiation at the level of the aims of RI. Rationales for RI remained within their comfort zone. As said before, technical, scientific and economic considerations had a considerable weight in the discourse. Cost-benefit analyses and the accountability for the use of monetary resources were prevalent rationales in RI informal meaning negotiation.

4.4 Discussion

Technology development is a dynamic ongoing process. Technologies themselves, their embedding and their impacts are shaped by complex co-construction processes by the different stakeholders involved (for example scientists, engineers and industry). This makes it difficult for society to hold anyone accountable in case of surprises, negative consequences and long-term impacts. RI can be seen as an attempt to reframe responsibilities in research and technology development. But how is this attempt substantiated? This study gives insight into how the concept of RI is being negotiated in formal academic discourse on the one hand, and during informal discourse of scientists using neuroimaging for concepts relevant to the domain of justice and security.

4.4.1 Difference Between the Formal and Informal Discourse

In the formal discourse on RI to be found in the literature, certain characteristic shifts were identified including engagement of societal stakeholders, anticipation and adaptability, broadening, and new insights on impacts and regulation. The formal discourse on RI suggests a range of changes with respect to activities, product criteria, mindset and the architecture of the scientific system. Scientists might be expected to not only engage professional experts, but also societal stakeholders throughout the entire research and development process, also in the early phases. They may have to start framing impacts of (prospective) applications in terms of a contribution to the collective good, instead of the still prevailing economic and techno scientific terms. They might be viewed in how well they anticipate and adapt the research and development process at any stage if insights acquired via anticipation yields a reason to do so. They may be expected to have a state of mind that is continuously reflexive towards wider issues relevant during the research process, to include at least ethical and societal considerations. The scientific system might change because of the RI framework. The social sciences and the humanities in particular can become involved in their research process. The role of the individual scientists within this range of possible changes remains somewhat unspecified.

The informal discourse, based on scientists in the field of neuroimaging, shows that the term RI is still unfamiliar to the scientists concerned. This can become problematic if RI and its associated formal lingo would proliferate. Their scientific system could alter under influence of incentives fitting the framework of RI. Although the formal discourse does not specifically target the scientists themselves, they can find that the rules of the game will start to change. For example, (future) research funds can demand that the scientists should involve other players in their research projects, notably from the social sciences and the humanities, who are currently relatively alien to them. Within the group of participants, meaning negotiation on RI is not yet taking place and their understanding rarely reflected the complexity found in the formal literature. There is thus a wide conceptual gap between the formal and informal discourse.

As the interviewed scientists did not easily relate to the concept of RI, there is in all probability a need to translate it to and specify it for the context they are familiar with. RI can carry considerable different connotations for scientists. This does not necessarily mean that they wholly disagree with the procedures and purposes of RI as a concept. Identification is needed where this lack of overlap is a mere question of wording and where these scientists hold a different perspective regarding RI. RI as a guiding concept is thus in need of operationalization within the specific context in which it is used. Conversely, authors involved in the formal meaning negotiation of RI might want to take up the challenge to explore the role responsibilities of individual scientists within the framework of RI, as well as those of other actors involved in innovation processes.

4.4.2 How New Is RI?

Before continuing to how this operationalization could be done, it is also important to consider whether this operationalization of RI is in fact necessary, or that we are dealing with old wine in new bottles. Looking at the formal discourse on RI, one cannot help but notice that RI sounds similar to existing methods aimed at a “better technology in a better society”, such as (constructive) technology assessment and value-sensitive design (VSD) (Schot and Rip 1997; Van de Poel 2009). For example, qualities ascribed to technology assessment (TA) resurface in RI negotiation of meaning, including anticipatory learning, taking into account unplanned or unanticipated effects, the broadening of the debate and design process, and the engagement of (professional) experts, stakeholders and citizens (van Est and Brom 2012). This might not be surprising as the authors now writing about RI are from the disciplines that are traditionally linked to TA, shaping and being shaped by TA, impacting the way TA is understood, performed and institutionalized (van Est and Brom 2012).

Maybe, RI should mainly be seen as an umbrella term for all activities related to the societal/political quest for relevancy of science. Policy aims have been directed on sustainability, translation from bench to bedside and now socially robust technologies. Activities have become fragmented, accompanied by arising separate disciplines of constructive TA (CTA), participatory TA, and ethical TA to name a few, developments in the neighboring discipline of VSD, and research groups having their own focus and methodologies within these disciplines. Possibly, RI is nothing more (or less) than an attempt to bring union in activities with a similar intention. In this light, the negotiation of meaning on RI is a reordering of the concepts in this scattered field. What seems to be coming out on top in the process is the focus on impact as a starting point of research and development. The results of this study point towards RI targeting acceptable and desirable impacts from the beginning. This is more pronounced in the policy realm, where RI has been linked to the 'grand challenges' the world is facing in our time, formulated in Horizon 2020 (EC 2011b; Owen et al. 2012). Instead of considering outcomes of research and innovation processes as serendipitous, the impact is the starting point. This 'strategic turn' is not without consequences as it necessitates a renegotiation of the contract between science and society (Borup et al. 2006; Hessels 2010; Irvine and Martin 1984; Rip 2004; Guston and Keniston 1994). Scientists can no longer rely on a position in which they independently produce 'reliable knowledge' or 'evidence', on which policy makers can base normative decisions on how to put the knowledge to use. Where CTA aims to bridge the gap arising from divisions of moral labor, RI would break open traditional moral labor divisions (Rip 2012). In this sense, RI could be seen as revolutionary and not just evolutionary to existing TA practices. And thus that RI is worthy of the effort of operationalisation as it can have an added value as a guiding concept.

4.4.3 Operationalisation of RI

This research suggests that stimulating a process of meaning negotiation on RI among scientists provides a valuable starting point for such an operationalization. Contextualization also does justice to the democratic quality of RI in a performative sense in providing criteria that arise from the practices themselves. It is increasingly recognized that "criteria are inherently immanent and cannot be picked a priori to guarantee outcomes" (Gomart and Hajer 2003). It also corresponds with developments in the field of ethics, towards pragmatic ethics, in which we see a shift from justification to a context of discovery (Keulartz et al. 2004).

The operationalization of an ambiguous concept as RI calls for a constructive approach that facilitates problem structuring. This will allow the substantial exploration of RI in a specific context, while at the same time facilitating a process of transferring this knowledge by transaction (instead of transmission). To prevent reification of hegemonic positions of certain stakeholders, problem-structuring entails a collective effort based on a plurality of insights. An approach that fits these criteria is the Interactive Learning and Action approach (ILA) (Bunders 1990; Broerse 1998; Bunders and Broerse 1991; Broerse and Bunders 2000). Recently, Roelofsen (2011) and Kloet (2011) have shown that the ILA approach can result in mutual learning among scientists and non-scientific stakeholders in the field of ecological genomics. Problem structuring takes place on the levels of expectations and outcomes as well as values and assumptions, and can pave the way for the identification of shared visions and concrete matches between research and practice.

The findings of the present research indicate that suitable points of entry for such a process are scientist role responsibilities, hypes and promises. The results of this study show that the interviewed scientists intuitively link their role responsibilities to RI as a concept. Although the method of interviews with individual scientists might have favored a line of argumentation of individual responsibilities, it is an interesting avenue to take. However, the occupational role responsibility matching the idea of the Independent Republic of Science, expressed by the interviewed scientists, does not sit comfortably in the new model of "science for society with society". According to this new contract, scientists should produce 'socially robust knowledge', the validity of which is determined by a broader group of experts, including 'lay experts' (Gibbons 1999; Swierstra et al. 2013). When the social contract changes, role responsibilities of scientists corresponding with the contract will also need renegotiation, for example with respect to scientists pro-actively addressing societal concerns. Furthermore, although problems relating to hypes and promises in science and research funding policy were not recognized by the interviewed scientists as an element of RI, they were quite articulate and passionate about it. They often voiced their worry about the detrimental effects of hypes and promises, with respect to attracting funding on the one hand and presenting a balanced view of the opportunities and limitations of the technology on the other. Moreover, they see it as a duty of scientists that they have to deal with that in a sound way. This could be a point of entrance in interactively renegotiating the professional identity of scientists in a new model of science. Besides this pragmatic argument why role responsibilities are a point of entry for the operationalization of RI in a context-specific way, we would also like to draw attention to a more substantial argument. In the meaning negotiation in the scientific literature two ethical families can be readily recognized. With respect to the process of RI, there is a focus on duty-driven deontological ethics, requiring the adherence to certain norms and principles. Regarding meaning negotiation on the product of RI, consequentialistic ethical theory can be identified. However, ethical agency can be best understood if also virtue ethics is taken into account. Considering the intrinsic uncertainty of technology development, one also needs virtuous innovators who aspire to make a valuable contribution to society (Pandza and Ellwood 2013). Within processes of science-driven innovation, scientists themselves are central "innovators", whose characteristics are then of importance. But the same goes for parties stimulating innovations, such as actors from industry, research funding organizations and governments.

One of the reasons to choose for a case study approach was to discover some context-specific elements for RI for the application of neuroimaging in justice and security. Although the scientists do see responsibilities for themselves, they were more often on a general level. Their concerns for hypes, for instance, is one that is relevant to neuroimaging, but also to other research domains. We think argumentation stayed on a general level because, on the one hand the scientists were not familiar enough with the concept of responsible innovation to be really specific about it, but on the other hand because the field of neuroimaging in justice and security is quite young, compared to medical neuroimaging for example. We therefore think it is prudent to use interactive methods to explore what can be considered responsible for the context of a specific application in the domain of justice and security, by bringing together the relevant actors involved, such as scientists, professional experts, and policy makers.

4.4.4 Resistance to RI

As transpires from the interviews, resistance can be expected from scientists, as normative position-taking in early stages of research and development can be equivocated by them with a limitation on the autonomy of science, and as such, as a brake on science itself (see also Broerse et al. 2010a; Caron-Flinterman et al. 2005). Science can be seen to have fallen victim to its own success in sectors as healthcare and agriculture. This has incited the wish to codetermine science and development directions from the outside, thereby eroding institutional and epistemic borders in the process (Hessels 2010; Borup et al. 2006). Scientists are increasingly expected to reach beyond the borders delineating their expertise and interact with increasingly wider and heterogeneous networks of potential stakeholders.

Resistance is not only fueled by a perceived threat to the autonomy of science. The interviews also make clear that it can also be due to a lack of appreciation of experiential knowledge by scientific stakeholders, as previously described by Caron-Flinterman et al. (2005). The added value of the participation of the public was found to be contested. Interestingly, the two opposites fear of scientific disasters and science's success, both in effect constrain the autonomy of science. However, note that ethical and societal considerations do not necessarily only limit scientific practice. They can also inspire new lines of research. The same goes for interactions with societal stakeholders (Broerse et al. 2010b; Caron-Flinterman et al. 2005; Popay and Williams 1996). It is up to practitioners involved in projects that employ RI as a guiding concept to facilitate that scientists experience the value of thinking about societal and ethical issues with non-scientific stakeholders. However, the structural involvement of societal stakeholders which is central to RI will require a system innovation (Caron-Flinterman et al. 2007; Elberse 2012). The culture, structure and practice of the larger research system will need to undergo fundamental change. The engagement of societal stakeholders will require new skills with respect to the interaction itself, but also regarding the feedback of the outcomes into the ongoing research and development process. This will require room to adapt within the research system.

On a first glance there also seems to be a disparity on where RI should be located, upstream or downstream. In the meaning negotiation in the academic literature, the trend is towards upstreaming. The interviewed scientists voiced that it is located downstream, at the implementation stage. However, we argue here that the dichotomy is not as stark as it seems. For example, the inclusion of non-scientific professional experts in the earliest phases of research appears to be supported. Apparent dichotomies can be more a matter of wording than of practice. However, scientists can indeed be quite skeptical towards other types of upstream engagement, notably the engagement of publics. Here, the role of practitioners in projects employing RI as a guiding principle would be to facilitate (alternative) experiences of the value of public engagement, as a first step towards its institutionalization in an early stage. Interestingly, within the policy domain there appears to be the mindset that RI is less relevant for basic research (Schuurbiers 2012). This can be a tricky exclusion, as for example most of the interviewed researchers considered themselves basic researchers. This argument was used by some to deny RI being applicable to their research practice. Besides the question whether a distinction can truly be made between basic and applied science, and where to draw this line (Godin 1998; Scott et al. 1994), it is also important to note here that it runs counter to the recent trend in TA from a narrow focus on specific technological artifacts towards the wider system level or even the trans-technological level (Allenby and Sarewitz 2011).

4.4.5 The Hype of RI?

Having watched how most of the interviewed scientists struggled with the term "responsible innovation" when it was introduced to them, one cannot help but wonder whether we could have arrived at the essence of RI sooner by asking scientists different kinds of questions than asking how they would conceptualize RI. In other words, should practitioners in projects employing RI as a guiding principle confront them with the term "responsible innovation" as such, or is RI more to be seen as a concept to be kept in mind by the practitioner? This is a fair question, which we can only counter with one pragmatic argument. RI is in vogue among public funders at the EU level, as well as the national level in the Netherlands. This means that it can be beneficial to scientists (and policymakers for that matter) if they are acquainted with the concept itself when applying for this type of grants. If RI is the way to go, operationalization should not stay implicit, as this would give a greater risk on a hype of the concept followed by disappointment. To avoid RI ending up as a void concept, or mere "manager's language" as one scientist put it, there is a need for practicing RI, measuring and evaluating RI and for the development of (policy) instruments to do so. Just as technologies can fail when the solutions they provide do not match the problems experienced by society, hyped concepts can fail as well, thereby losing its (potential) power to mobilize concerted action. RI might fall victim to its own ambiguity. Just like natural scientists sometimes get reproached for their "pet technologies", we should prevent RI from becoming a similar "pet concept". It is striking that the scientists themselves seem little involved in its meaning negotiation, considering how this concept might change their research system. Evasive practices by scientists are likely to occur if RI is forced upon them top-down. The time has come for scientists and other stakeholders to engage with each other in order to learn about relevant societal and ethical aspects of their work, in a co-constructive and problem-structuring process to operationalize RI.

Acknowledgments We thank the anonymous reviewer for Jacqueline Broerse the helpful comments and suggestions.

Appendix

See Fig. 4.2.

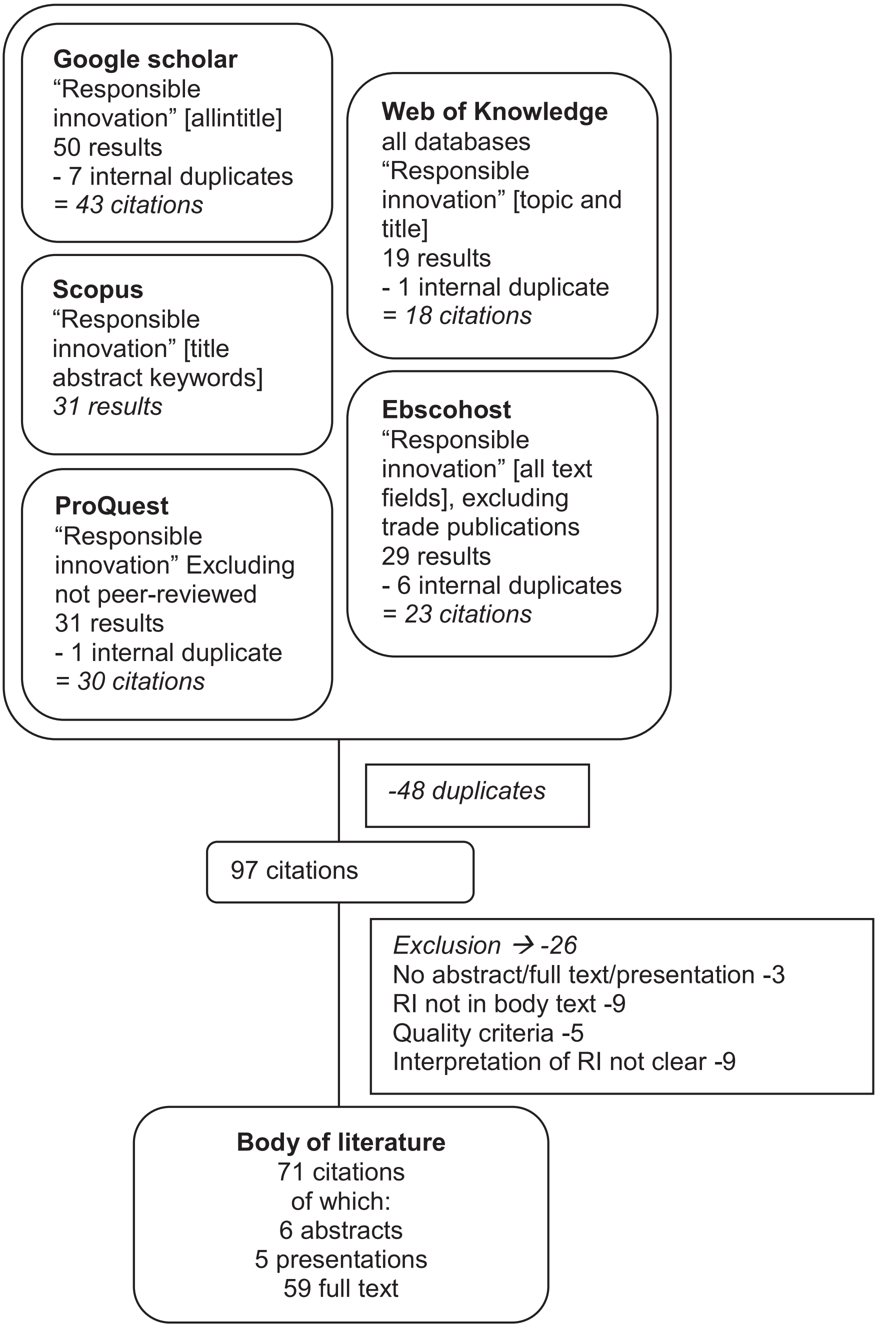

Fig. 4.2 Systematic literature search flow chart

References

Allenby, Braden R., and Daniel Sarewitz. 2011. *The techno-human condition*. Cambridge: MIT Press.

Armstrong, Margaret, Guillaume Cornut, Stéphane Delacôte, Marc Lenglet, Yuval Millo, Fabian Muniesa, Alexandre Pointier, and Yamina Tadjeddine. 2012. Towards a practical approach to responsible innovation in finance: New product committees revisited. *Journal of Financial Regulation and Compliance* 20(2): 2–3.

Barben, Daniel, Erik Fisher, Cynthia Selin, and David H. Guston. 2007. Anticipatory governance of nanotechnology: Foresight, engagement, and integration. *The Handbook of Science and Technology Studies* 3: 979–1000.

Barré, Rémi. 2011. Responsible innovation: From concepts to practice. Reflecting on a Franco-British workshop. *Natures Sciences Sociétés* 19(4): 405–409.

Bergquist, Magnus, Jan Ljungberg, and Ulrika Lundh-Snis. 2001. Practising peer review in organizations: A qualifier for knowledge dissemination and legitimization. *Journal of Information Technology* 16(2): 99–112. doi:10.1080/02683960122785.

Bijker, Wiebe E., Thomas P. Hughes, and Trevor J. Pinch (eds.). 1984. *The social construction of technological systems: New directions in the sociology and history of technology*. Cambridge: MIT Press.

Borup, Mads, Nik Brown, Kornelia Konrad, and Harro Van Lente. 2006. The sociology of expectations in science and technology. *Technology Analysis and Strategic Management* 18(3–4): 285–298.

Broerse, Jacqueline E.W. 1998. *Towards a new development strategy: How to include small-scale farmers in the technological innovation process*. Delft: Eburon.

Broerse, Jacqueline E.W., Janneke E. Elberse, J. Francisca Caron-Flinterman, and Marjolein B.M. Zweekhorst. 2010a. Enhancing a transition towards a needs-oriented health research system through patient participation. In *Transitions in health systems: Dealing with persistent problems*, ed. J.E.W. Broerse, and J.F.G. Bunders. Amsterdam: VU University Press.

Broerse, Jacqueline E.W., and Joske F.G. Bunders. 2000. Requirements for biotechnology development: The necessity for an interactive and participatory innovation process. *International Journal for Biotechnology* 2(4): 275–296.

Broerse, Jacqueline E.W., Marjolein B.M. Zweekhorst, Annemiek J.M.L. van Rensen, and Monique J.M. de Haan. 2010b. Involving burn survivors in agenda setting on burn research: An added value? *Burns* 36(2): 217–231.

Brown, John Seely, and Paul Duguid. 1996. The social life of documents. *First Monday* 1 (1–6).

Bunders, Joske F.G. 1990. Biotechnology for small-scale farmers in developing countries. Amsterdam: VU University Press.

Bunders, Joske F.G., and Jacqueline E.W. Broerse. 1991. Appropriate biotechnology in small-scale agriculture: How to reorient research and development. Wallingford: CAB International.

Callon, Michel. 1986. Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieu Bay. In *Power, action and belief: A new sociology af knowledge?*, ed. John Law, 196–233. London: Routledge and Kegan Paul.

Caron-Flinterman, J. Francisca, Jacqueline E.W. Broerse, and Joske F.G. Bunders. 2005. The experiential knowledge of patients: A new resource for biomedical research? *Social Science and Medicine* 60(11): 2575–2584. doi:http://dx.doi.org/10.1016/j.socscimed.2004.11.023.

Caron-Flinterman, J.Francisca, Jacqueline E.W. Broerse, and Joske F.G. Bunders. 2007. Patient partnership in decision-making on biomedical research: Changing the network. *Science, Technology and Human Values* 32(3): 339–368. doi:10.1177/0162243906298354.

Chervenak, Frank A., and Laurence B. McCullough. 2006. Scientifically and ethically responsible innovation and research in ultrasound in obstetrics and gynecology. *Ultrasound in Obstetrics and Gynecology* 28(1): 1–4. doi:10.1002/uog.2825.

D’Silva, Joel, Douglas K.R. Robinson, and Clare Shelley-Egan. 2012. A game with rules in the making—how the high probability of waiting games in nanomedicine is being mitigated through distributed regulation and responsible innovation. *Technology Analysis and Strategic Management* 24(6): 583–602. doi:10.1080/09537325.2012.693671.

Diefenbach, Thomas. 2009. Are case studies more than sophisticated story telling? Methodological problems of case studies mainly based on semi-structured interviews. *Quality and Quantity* 43(Article): 875–894. doi:10.1007/s11135-008-9164-0.

Dondorp, Wybo, and Guido de Wert. 2011. Innovative reproductive technologies: Risks and responsibilities. *Human Reproduction* 26(7): 1604–1608. doi:10.1093/humrep/der112.

Duke, Daniel L. 1978. Toward responsible innovation. *The Educational Forum* 42(3): 351–371. doi:10.1080/00131727809336323.

EC. 2010. *Europe 2020 flagship initiative: Innovation union.* COM(2010) 546. Brussels: European Commission.

EC. 2011a. *From challenges to opportunities: Towards a common strategic framework for EU research and innovation funding.* Green Paper COM(2011) 48. Brussels: European Commission.

EC. 2011b. *Horizon 2020—The framework programme for research and innovation*. COM(2011) 808 final. Brussels: European Commission.

Elberse, Janneke E. 2012. Changing the health research system: Patient participation in health research. 's-Hertogenbosch: BOXpress.

Ferrari, Arianna, and Alfred Nordmann. 2010. Beyond conversation: Some lessons for nanoethics. *NanoEthics* 4(2): 171–181. doi:10.1007/s11569-010-0098-3.

Fisher, Erik. 2011. Public science and technology scholars: Engaging Whom? *Science and Engineering Ethics* 17(4): 607–620. doi:10.1007/s11948-011-9331-x.

Flyvbjerg, Bent. 2006. Five misunderstandings about case-study research. *Qualitative Inquiry* 12(2): 219–245. doi:10.1177/1077800405284363.

Gibbons, Michael. 1999. Science's new social contract with society. *Nature* 402(6761): C81–C84.

Godin, Benoît. 1998. Writing performative history: The new New Atlantis? *Social Studies of Science* 28(3): 465–483. doi:10.1177/030631298028003004.

Gomart, Emilie, and Maarten Hajer. 2003. Is that politics? For an inquiry into forms in contemporary politics. In *Social studies of science and technology: Looking back, ahead*, eds. Bernward Joerges, and Helga Nowotny, 33–61. Dordrecht: Kluwer Academic Publishers, Dordrecht: Springer.

Goodenough, Oliver R., and Micaela Tucker. 2010. Law and cognitive neuroscience. *Annual Review of Law and Social Science* 6(1): 61–92. doi:10.1146/annurev.lawsocsci.093008.131523.

Goorden, Lieve, Michiel Van Oudheusden, Johan Evers, and Marian Deblonde. 2008. Lose one another … and find one another in nanospace. 'nanotechnologies for tomorrow's society: A case for reflective action research in flanders (NanoSoc)'. *NanoEthics* 2(3): 213–230. doi:10.1007/s11569-008-0043-x.

Greely, Henry T. 2007. Neuroscience and criminal justice: Not responsibility but treatment. *Kansas Law Review* 56: 1103–1138.

Groves, Chris, Lori Frater, Robert Lee, and Elen Stokes. 2011. Is there room at the bottom for CSR? Corporate social responsibility and nanotechnology in the UK. *Journal of Business Ethics* 101(4): 525–552. doi:10.1007/s10551-010-0731-7.

Guston, David H. 2006. Responsible knowledge-based innovation. *Society* 43(4): 19–21.

Guston, David H. 2007. Toward centres for responsible innovation in the commercialized university. Public science in liberal democracy: The challenge to science and democracy. Toronto: University of Toronto Press.

Guston, David H., and Kenneth Keniston. 1994. Introduction: the social contract for science. In The fragile contract. University Science and the Federal Government, ed. David H. Guston, and Kenneth Kenniston, 1–41. Cambridge: MIT Press.

Hellström, Tomas. 2003. Systemic innovation and risk: Technology assessment and the challenge of responsible innovation. *Technology in Society* 25(3): 369–384. doi:10.1016/s0160-791x(03)00041-1.

Hessels, Laurens K. 2010. Science and the struggle for relevance. PhD thesis, Universiteit Utrecht, Utrecht.

Ingham, Marc, Françoise de Viron, and Alain Tihon. 2010. Sustaining responsible innovation through responsible finance: Exploring the strategy process and alignment at Triodos Bank. Paper presented at the EBEN conference, March 1.

Irvine, John, and Ben R. Martin. 1984. *Foresight in science: Picking the winners*. London: Frances Pinter.

Ishizu, Saori, Mizuki Sekiya, Ken-ichi Ishibashi, Yumi Negami, and Masafumi Ata. 2008. Toward the responsible innovation with nanotechnology in Japan: Our scope. *Journal of Nanoparticle Research* 10(2): 229–254. doi:10.1007/s11051-007-9306-7.

Jonas, Hans. 1984. *The imperative of responsibility: In search of an ethics for the technological age*. Chicago: The University of Chicago Press.

Kaza, Greg. 2006. Regulation of financial derivatives in the US code. *Derivatives Use, Trading and Regulation* 11(4): 381–386.

Keulartz, Jozef, Maartje Schermer, Michiel Korthals, and Tsjalling Swierstra. 2004. Ethics in technological culture: A programmatic proposal for a pragmatist approach. Science, Technology and Human Values 29(1): 3–29. doi:10.1177/0162243903259188.

Kiran, Asle H, H. 2012. Does responsible innovation presuppose design instrumentalism? Examining the case of telecare at home in the Netherlands. *Technology in Society* 34(3): 216–226. doi:10.1016/j.techsoc.2012.07.001.

Kloet, Roy R. 2011. Realizing societal ambitions in innovative research programs: The case of the dutch ecogenomics consortium. PhD thesis. Oisterwijk: BOXpress.

Latour, Bruno. 1987. *Science in action: How to follow scientists and engineers through society*. Cambridge: Harvard University Press.

Latour, Bruno. 1990. Drawing things together. In Representations in scientific practice, ed. Michael E Lynch, and Steve Woolgar, 19–68. Cambridge: MIT Press.

Laudel, Grit, and Jochen Gläser. 2007. Interviewing scientists. *Science, Technology and Innovation Studies* 3(2): 91–111.

Law, John. 1992. Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Systems practice* 5(4): 379–393. doi:10.1007/bf01059830.

Owen, Richard. 2009. A new era of responsible innovation. *Planet Earth* (autumn):14–15.

Owen, Richard, and Nicola Goldberg. 2010. Responsible innovation: A pilot study with the U.K. engineering and physical sciences research council. *Risk analysis: An international journal* 30(11): 1699–1707. doi:10.1111/j.1539-6924.2010.01517.x.

Owen, Richard, Phil Macnaghten, and Jack Stilgoe. 2012. Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy* 39(6): 751–760. doi:10.1093/scipol/scs093.

Ozdemir, Vural, Samer A. Faraj, and Bartha M. Knoppers. 2011. Steering vaccinomics innovations with anticipatory governance and participatory foresight. *OMICS: A Journal of Integrative Biology* 15(9): 637–646. doi:10.1089/omi.2011.0087.

Pandza, Krsto, and Paul Ellwood. 2013. Strategic and ethical foundations for responsible innovation. *Research Policy* 42(5): 1112–1125. doi:http://dx.doi.org/10.1016/j.respol.2013.02.007.

Patton, Michael Quinn. 1990. *Qualitative research and evaluation methods*, 2nd ed. Newbury Park: Sage Publications, Incorporated.

Pavie, Xavier. 2012. The importance of responsible-innovation and the necessity of 'innovation-care'. ESSEC Working Paper 1203.

Petchesky, Rosalind Pollack. 1987. Fetal images: The power of visual culture in the politics of reproduction. *Feminist Studies* 13(2): 263–292.

Polanyi, Michael. 1962. The republic of science: Its political and economic theory. *Minerva* 1(Autum): 54–73.

Popay, Jennie, and Gareth Williams. 1996. Public health research and lay knowledge. *Social Science and Medicine* 42(5): 759–768. doi:http://dx.doi.org/10.1016/0277-9536(95)00341-X.

Reddy, Panga Jaipal, Rekha Jain, Young-Ki Paik, Robin Downey, Adam S. Ptolemy, Vural Ozdemir, and Sanjeeva Srivastava. 2011. Personalized medicine in the age of pharmacoproteomics: A close up on india and need for social science engagement for responsible

innovation in post-proteomic biology. *Current Pharmacogenomics and Personalized Medicine* 9(3): 159–167. doi:10.2174/187569211796957557.
Rip, Arie. 2004. Strategic research, post-modern universities and research training. *Higher Education Policy* 17(2): 153–166.
Rip, Arie. 2005. Technology assessment as part of the co-evolution of nanotechnology and society: The thrust of the TA program in NanoNed. Paper presented at the conference on nanotechnology in science, economy and society, Marburg, 13–15 January.
Rip, Arie. 2012. Technology assessment as a site for STS knowledge and insights. Paper presented at the presentation at the WTMC workshop: Assessing technology assessment, Soeterbeeck, Raventstein, The Netherlands, 2 November.
Roelofsen, Anneloes. 2011. Exploring the future of ecogenomics: Constructive technology assessment and emerging technologies. PhD Thesis, Ridderprint, Ridderkerk, The Netherlands.
Schot, Johan, and Arie Rip. 1997. The past and future of constructive technology assessment. *Technological Forecasting and Social Change* 54(2–3): 251–268. doi:10.1016/s0040-1625(96)00180-1.
Schuurbiers, Daan. 2011. What happens in the lab: Applying midstream modulation to enhance critical reflection in the laboratory. *Science and Engineering Ethics* 17(4): 769–788. doi:10.1007/s11948-011-9317-8.
Schuurbiers, Daan. 2012. See ethics as a stimulus and not as an obstacle (Zie ethiek als impuls en niet als belemmering). *LEV*. November.
Schwartz, Victor E. 1992. Innovation and our product liability system: Let us end the conflict on incentives. *Business Economics* 27(4): 15.
Scott, Peter, Michael Gibbons, Helga Nowotny, Camille Limoges, Martin Trow, and Simon Schwartzman. 1994. *The new production of knowledge: The dynamics of science and research in contemporary societies*. Sage Publications Limited.
Shelley-Egan, Clare. 2010. The ambivalence of promising technology. *NanoEthics* 4(2): 183–189. doi:10.1007/s11569-010-0099-2.
Stahl, Bernd Carsten. 2011. IT for a better future: How to integrate ethics, politics and innovation. *Journal of Information, Communication and Ethics in Society* 9(3): 140–156. doi:10.1108/14779961111167630.
Stinner, Deborah H., Ivan Glick, and Benjamin R. Stinner. 1992. Forage legumes and cultural sustainability—Lessons from history. *Agriculture, Ecosystems and Environment* 40(1–4): 233–248. doi:10.1016/0167-8809(92)90095-s.
Sunstein, Cass R. 2002. The paralyzing principle. *Regulation* 25: 32.
Swierstra, Tsjalling, Niki Vermeulen, Johan Braeckman, and Roel van Driel. 2013. Rethinking the life sciences. *EMBO Rep*. doi:10.1038/embor.2013.30. advance online publication.
te Kulve, Haico, and Arie Rip. 2011. Constructing productive engagement: Pre-engagement tools for emerging technologies. *Science and Engineering Ethics* 17(4): 699–714.
Tihon, Alain, and Marc Ingham. 2011. The societal system and responsible innovations: Freeing sustainable development from a deadlock. *Journal of Innovation Economics* 8(2): 11–31. doi :10.3917/jie.008.0011.
Tyl, Benjamin, Jérémy Legardeur, Dominique Millet, André Falchi, and Bertrand Ranvier. 2011. A new approach for the development of a creative method to stimulate responsible innovation. In *Global product development*, ed. Alain Bernard, 93–104. Berlin Heidelberg: Springer.
Van de Poel, Ibo. 2009. Values in engineering design. In *Philosophy of technology and engineering sciences*, ed. Anthonie Meijers, 973–1006. Handbook of the philosophy of science. Amsterdam: Elsevier.
van den Hove, Sybille, Jacqueline McGlade, Pierre Mottet, and Michael H. Depledge. 2012. The innovation union: A perfect means to confused ends? *Environmental Science and Policy* 16: 73–80. doi:10.1016/j.envsci.2011.11.006.
van Est, Rinie, and Frans Brom. 2012. Technology assessment, analytic and democratic practice. In *Encyclopedia of applied ethics*, ed. Ruth Chadwick, 306–320. Oxford: Elsevier.

van Lente, Harro. 2012. Navigating foresight in a sea of expectations: Lessons from the sociology of expectations. *Technology Analysis and Strategic Management* 24(8): 769–782.

Van Oudheusden, Michiel. forthcoming. Broadening 'Innovation' through technology assessment in Flanders, Belgium: The Case of Nanotechnologies.

Venier, Sylvia. 2011. BIRD platform for responsible innovation takes wing. *Biometric Technology Today* 2011(3): 5–7. doi:10.1016/s0969-4765(11)70055-8.

Verbeek, Peter-Paul. 2011. Mediated morality. In *Moralizing technology: Understanding and designing the morality of things*, 1–20. Chicago; London: The University of Chicago Press.

von Schomberg, René. 2012. Prospects for technology assessment in a framework of responsible research and innovation. *Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden*:39.

Yin, Robert K. 2002. *Case study research: Design and methods*, 3rd ed. Thousand Oaks: Sage.

Part II
Approaches—Implementing and Reflecting on Responsible Innovation

Chapter 5
Responsible Management of Social Experiments: Challenges for Policymaking

Zoë Robaey and Arno Simons

Abstract This paper assumes that the introduction of new technologies takes the form of social experiments and asks how such experiments can be managed responsibly. While social experimentation in itself is not an entirely new phenomenon, modern societies are increasingly describing themselves *as* experimental societies. Uncertainty and ignorance are seen as problems of modernity to which a continuous learning approach provides the solution. From an ethical perspective, social experimentation poses entirely new challenges, *inter alia* because outcomes often cannot be anticipated beforehand but have an immediate impact on society. We identify six values behind morally responsible social experimentation and set them against existing policy approaches dealing with the uncertainties involved in introducing new technologies into societies. We draw conclusions on how current policy approaches could better manage the introduction of new technologies in a responsible manner, emphasizing a lack in the value of justice.

Keywords Responsible innovation · Innovation policy · Social experimentation · Justice

Z. Robaey (✉)
Department of Values, Technology and Innovation, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, The Netherlands
e-mail: z.h.robaey@tudelft.nl

A. Simons
Institute for Ecological Economy Research, Potsdamer Str. 105, 10785 Berlin, Germany
e-mail: arno.simons@posteo.de

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_5

5.1 Responsibility in the Introduction of New Technologies

It has often been argued that the introduction of new technologies in society takes the form of social experiments (cf. Martin and Schinzinger 1983; Krohn and Weingart 1987; Krohn and Weyer 1994; Levidow and Carr 2007; Jacobs et al. 2010; Van de Poel 2011). The argument is based on the observation that the outcomes of technological innovation processes cannot fully be apprehended beforehand. Whether and how a new technology works and which side-effects it has can only be determined in practice. Learning about the functioning of new technologies thus takes place under real-world conditions and implies a trade-off between potential gains and harmful side effects. This is problematic in so far as potential failures and harmful side effects of new technologies cannot be contained within safe boundaries, as in classical laboratory experiments. Instead, such side effects are felt directly by the society into which the technology is introduced—sometimes even by more distant societies, as the case of Chernobyl has demonstrated. Introducing new technologies into society therefore raises a number of ethical concerns, but conceiving of these processes as social experiments also allows developing new perspectives on how to deal with potential harmful side effects.

Risks are known probabilities of expected events. This allows one to deal with them to the best of one's knowledge. However, uncertainty and ignorance both remain a field of struggle for decision-makers, as they represent, respectively, unknown probabilities of expected events and things we just cannot anticipate (Felt et al. 2007). How shall one act when one knows neither the consequences nor the probabilities of one's action? This question has been troubling our modern times as it points to the limits of knowledge and to the unintended consequences technologies may have—technologies that are supposed to make our lives better. Before we know it, we find ourselves in a conundrum where we develop technologies to fix problems but worry whether we may not just be bringing about the next one.

In the face of the unknown, our best judgment therefore lies in moral values. Relying on moral values allows a fundamental, shared, and stable base for moral judgment. In fact, important institutions rely on moral values, such as the Treaty of Lisbon, which emphasizes freedom, equality, and democracy, among many others, as core values. With questions of new technologies, values serve as guideposts for acting in the face of uncertainty and ignorance. If we cannot know or anticipate everything, then at least we can act according to a set of values and principles that make our actions responsible. This need for morally responsible action also takes an increasingly important place in research and development as the recent agenda for responsible research and innovation (RRI) within the European Union (European Commission 2012) shows. The ideas we present in this chapter underline the need for RRI and contribute to further defining what RRI can entail. In this context, we ask: given that new technologies potentially bring along a number of uncertainties and ignorance (or unintended negative side-effects), how can we responsibly manage the introduction of new technologies in society, and do current approaches exhibit moral values underlying responsibility?

Within this frame, we bring together two strands of literatures that have been dealing with (social) experimentation in different spheres. First, we draw from insights in biomedical ethics on what is important to take into account when experimenting with human lives, for maintaining health and increasing well-being. We link these insights to another strand of literature that deals with the introduction of technologies as social experiments. To do this, we first discuss why the notion of social experimentation has become central in the context of modernity and we then look into the values behind morally responsible experimentation in the field of biomedical ethics, trying to translate them to social experimentation. For the core of our analysis, we look at existing policy approaches that have been developed specifically for dealing with the experimental and uncertain nature of introducing new technologies in society, namely the precautionary principle, participatory technology assessment, and adaptive management.

As mentioned by Koops in the introduction to this volume (Chap. 1), policy tools and approaches constitute ways of realizing responsible innovation. They are themselves regulatory innovations, and studying them allows extracting practical ethical criteria from on-going practice (cf. Mann et al. 2014a, b; Voß and Simons 2014). Indeed, while responsible innovation presents a "landscape [that] is very diverse" (Chap. 1), similarities can be extracted at a more abstract level. If certain values are present but implemented differently, responsible innovation can be realized both as a process and a product, and this, in a context-sensitive manner. In Chap. 3 of this volume, Davies and Horst study those different landscapes and point to the de-individualization of responsibility. Providing more practical ways of describing moral responsibility as we aim to do in this chapter might greatly inform how agents can be responsible and thus help to avoid this de-individualization. Furthermore, as Chap. 9 by Haen and colleagues underlines, existing approaches and tools need to be improved, and there is space for innovation striving towards values that enhance and support responsible innovation.

Our task here is to set existing approaches against the values we identify as being central to moral responsibility in social experiments with new technologies and to compare them with each other. We draw conclusions on where current approaches score well and where they show deficits when it comes to moral responsibility for the introduction of new technologies in society.

5.2 Social Experimentation

Why do we need to talk of social experiments, or social experimentation when talking about introducing new technologies in society in the first place? Living in a techno-scientific world means that science and technology have, on the one hand, helped to solve major problems and brought about great benefits to societies, and on the other hand, have created a whole range of new problems. And since many of these new problems are unknown until they occur, the term social experiment is justified when talking about the introduction of new technologies in society. Knowledge on the real impacts of technologies can often only be gathered after or during introducing that technology into society. Many debates could be had on the use of the word experiment here, but what we mean by it is simply to do or try something with a core aim to learn about its effects. The notion of 'social' experimentation denotes the situation that the experiment is carried out 'in the wild,' in society that is.

With the scientific and technological revolutions, scientific and technological progress and innovation first seemed to become more predictable, if not fully controllable. 'Modernity' became equated with rational control and the promise of technocratic rule. But over the last three or four decades, such hopes have partly faded away. While some contend we have entered a 'second modernity' (Beck 1992), where technologies become increasingly complex and less controllable and bring along new hazards, others claim that we "have never been modern" in the first place (Latour 1993), that we have been lying to ourselves and constructing an illusion of control. This debate underlines a changing relationship between technology and society.

The changing relationship has two facets. On the one hand, technologies have indeed become more complex, and one can only fully know of their negative consequences once they occur (Collingridge 1980). On the other hand, accidents like the ones in Chernobyl and Fukushima have shown that one cannot control everything. In this context, social experimentation promises to provide a strategy of engaging with new technologies while learning about and dealing with their risks as soon as they occur (e.g. Van de Poel 2011; Krohn and Weyer 1994; Krohn and Weingart 1987).

However, a social experiment does not only entail that uncertainties and ignorance will reveal themselves as technology is used in practice. A social experiment also means that we are consciously taking chances with new technologies. We want to do social experiments because we strive to minimize negative side-effects. Considering the introduction of new technologies in society a social experiment, allows making adjustments and minimizing unwanted side-effects. In the recent years, it also seems our decision-makers are embracing the lens of the experiment, as "politics itself is changing as it acquires a more experimental style that—with respect to the European Union—is variously called 'experimentalist governance', 'regulatory experimentalism', or 'collective experimentation'" (in: Schwarz and Krohn 2011).

This experimental lens brings about new ethical challenges. At the crux of the experiment is the realization of the limits of knowledge and the recognition of things that are uncertain and ignored and therefore need to be learned about and corrected for, as the experiment unfolds. In this context, Van de Poel (2011) suggests a set of conditions for morally responsible experimentation using notions of biomedical ethics (Beauchamp and Childress 2001) in the context of a social experiment.

There are many parallels to be drawn between the notion of social experiments as we present it and the idea of RRI. Von Schomberg (2012, p. 49) defines RRI as "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to 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)." Indeed, the mutual responsiveness suggests an on-going feedback loop between the developers of technologies and society; there is a need to learn from each other, as in the social experiment. Also, RRI is strongly embedded in values that determine this responsibility. A wide spectrum of values underlies ethical acceptability, sustainability and societal desirability. The next section will highlight which values form the baseline and need to be fulfilled for a social experiment to be morally responsible. The values we will describe are not foreign to the ones contained in the meaning of RRI, however they provide a purposive base for evaluation.

5.3 Values for Morally Responsible Management of Technologies in Societies

What can be observed from the list of provisional conditions for morally responsible experimentation proposed by Van de Poel, see Table 5.1, is that they derive on the one hand from environmental management and on the other hand from considerations of biomedical ethics. A closer look at the conditions allows linking them to one or more of the four principles for ethical experimentation on humans: beneficence, non-malevolence, justice and autonomy (Beauchamp and Childress 2001)—we refer to these as higher values. The conditions link also to instrumental values of learning and intervening, as these allow attaining desired ends. We notice that several conditions are linked to learning, thus emphasizing the quality of the

Table 5.1 Provisional conditions for responsible experimentation (Van de Poel 2011)

1.	Absence of other reasonable means for gaining knowledge about hazards
2.	Monitoring
3.	Possibility to stop the experiment
4.	Consciously scaling up
5.	Flexible set-up
6.	Avoid experiments that undermine resilience of receiving 'system'
7.	Containment of hazards as far as reasonably possible
8.	Reasonable to expect social benefits from the experiment
9.	Experimental subjects are informed
10.	Approved by democratically legitimized bodies
11.	Experimental subjects can influence the set-up, carrying out and stopping of the experiment
12.	Vulnerable experimental subjects are either not subject to the experiment or are additionally protected
13.	A fair distribution of potential hazards and benefits

notion of an experiment, which has, as ultimate goal, to learn about the world in order to improve it (by having the ability to act in time to minimize undesired side effects).

Before providing a discussion of these values with examples from the conditions for morally responsible experimentation, we want to underline criticisms that have been addressed to biomedical ethics. Mostly, what we refer to as higher values in this text are commonly used as principles in biomedical ethics. The problem with principles is that they are often used differently in practice and might even be contradictory in certain cases, therefore not providing a guide for moral actions (Clouser and Gert 1990). We therefore do not suggest to use beneficence, non-maleficence, autonomy and justice as principles that can be checked off a list of moral adequacy. Rather, we use these as higher values, i.e. as things we ought to strive for.

Here we provide a short discussion of these higher values with examples. Also, while conditions may link back to more than one higher or instrumental value, we choose to link them to their most obvious value for the sake of brevity and clarity. Since these are the main values associated with the conditions, it will not affect our argument later on. We will also point to overlap between using these values and RRI. Similarities can also be distinguished with other chapters of this volume such as Chap. 12 by Setiawan and Singh, where five dimensions of responsible innovation are described (anticipation, reflexivity, responsiveness, deliberation and participation), or Chap. 9 by Haen et al., where the emphasis is put on responsiveness and feedback loops with regards to public participation. The common direction of all these ideas points to the clear presence of RRI in policy that remains yet to be clearly defined. The set of higher values described below might help towards describing unifying elements amongst applications and definitions.

<u>Beneficence</u>: Beneficence, simply put, is a higher value that tells us to "Do good". For instance, a medical experiment shall be carried out only if benefits are expected for the patient. In the same sense, the introduction of a new technology in society that has high potential benefits but also high stakes makes sense only if those potential benefits are expected for society as condition 8 stipulates, "it is reasonable to expect social benefits from the experiment." In RRI, this could be equated with the need for social desirability, i.e. does it bring something good to society.

<u>Non-malevolence</u>: Non-malevolence is the pendant of beneficence in the sense that is not sufficient to do good; this doing good should also "Do no harm". In medical ethics, an experiment should not further harm the patient and avoid unnecessary suffering. In the same sense, introducing a new technology for societal benefits should not happen if we *know* the harm to be great. A condition that reflects this higher value is condition 6, which stipulates, "avoid experiments that undermine resilience of receiving system," in other words: do not carry out experiments that will harm. Condition 7 also fulfills this higher value; indeed, the containment of hazards *as far as possible* is a way to avoid doing harm. The use of methods for technology assessment, or the precautionary principle underlines similar concerns for non-malevolence.

Justice: While there is a wide body of literature on justice, its meaning in biomedical ethics boils down to a question of fairness in distribution and access when looking at access to treatment or choice of sample groups for instance (Beauchamp and Childress 2001). In the technological context, it might mean, as an example, to take into account the social status of communities where power plants are built. Conditions 12 and 13 represent well the value of justice: "vulnerable experimental subjects are either not subject to the experiment, or are additionally protected" and "a fair distribution of potential hazards and benefits." In the framework of RRI, values of the European Treaty should be upheld, which include social justice and equality between women and men (von Schomberg 2012). These also illustrate concerns of access and distribution.

Autonomy: Autonomy means that individuals should be able to decide for themselves what happens with them, in the medical context. Someone who is autonomous could also be called a moral agent, i.e. the agent is capable of evaluating and assessing actions and deciding, even acting, upon this deliberation. In the social experiment, this implies that members of society that are affected are aware and have a say in the set-up of the experiment, as is mentioned in conditions 9, 10 and 11: information needs to be provided to participants, also democratic institutions are involved and subjects, or participants, have a say in the experiment. In the definition of RRI, transparency of the process is an important dimension. Transparency could be seen as an instrumental value for autonomy.

Learning: Learning is probably the most central element of the social experiment. Indeed, the whole reason why this is called a social experiment is because of the uncertainties and ignorance that are inherent to the introduction of new technologies in society. Also learning, in all its shape, is the only way to deal with things one doesn't know. Conditions 1 and 2 are example of how learning is present: "absence of other reasonable means for gaining knowledge about hazards", underlining the necessity of carrying out the experiment to learn, and "monitoring", showing the conscious gathering of any novel information that might inform the experimenters on the course of the introduction of the new technology in society.

Intervening: Intervening is linked to the autonomy of the agent. Indeed, intervening is the capacity of subject to act in, modify or adapt the experiment. Intervening is only possible when the autonomous agent has learned and revises her decisions. Conditions 3, 4, and 5 are an example of intervening through stopping, scaling-up and modifying the set-up. As mentioned earlier in this paper, RRI is also about learning and intervening, by creating feedback loops between developers and society.

The fulfillment of these conditions indicates how such a social experiment can be carried out responsibly since it fulfills higher values that contribute to the rightness of an experiment as well as instrumental values which insure that the end, to minimize undesired and unknown side effects of a desirable technology, is met. It is important to note that the term "moral responsibility" has so far been left undefined purposefully since in the framework suggested by Van de Poel, moral responsibility emanates from fulfilling the conditions.

5.4 Comparison

What justifies policy action in the face of the partially and totally unknown? A variety of practices are used to gather knowledge, evaluate alternatives and make an informed decision on a technology. Risk Assessments (RA), Cost Benefit Analyses (CBA), Environmental Impact Assessments (EIAs) are varying degrees of this information gathering process, with the exception of EIAs that rely on a justification of a certain course of action based on available information and therefore contain value judgment as opposed to RA and CBA which present themselves as objective evaluations. EIAs differ in that sense; often a jury of expert delivers a *final* opinion, thus providing for a process, which is dynamic up to the point where the project is approved. Indeed EIAs are meant to "assist in making the decision whether or not a proposal should be approved" (Harvey and Clarke 2012) and therefore do not relate to how risky and uncertain technologies are actually managed once approved (cf. Perdicoulis et al. 2012). While these approaches have their merits, they all lie in the realm of deciding on a technology and not as much managing its uncertainties and ignored consequences, which is the focus of this paper.

What we aim to do in this paper is to reflect on the values behind commonly used approaches to deal with uncertainty and ignorance in order to see whether they meet what we define as a baseline for moral responsibility above. This does not mean that these approaches are used in a vacuum, or independently of each other. Interestingly, these approaches are used as methods under RRI, especially for the precautionary principle and the range of methods that may fall under participatory technology assessment.

Therefore, we investigate policy approaches that both purposefully leave space for dealing with uncertain and ignored consequences as well as deal with the management-side of introducing technologies. Specifically, this paper will look at the precautionary principle (PP), participatory technology assessment (pTA) and adaptive management (AM). It is important to note that these approaches are not necessarily used with a preference over each other or exclusively of each other in the real world. The analysis at hand tries instead to point out what distinguishes them in their conceptual design in order to compare them to the values we identified for carrying out morally responsible social experiments.

Before presenting these policy approaches, it is important to note that they do not come out of a vacuum but are the result of years of policy experience and scholarly work. Also, we wanted to explain how we went about deciding *what* to look at and *why* in the midst of extensive scholarly work and policy translations. First of all the three approaches we discuss are all used in the context of the European Union. This provides for relative unity in terms of governance style, or governance culture. Secondly, we had to make such a choice in order to have things to compare to each other. Otherwise, we would have a very abstract discussion on notions, which while valuable would not provide an answer to our question, namely on how to responsibly manage the introduction of new technologies in society. This question indeed begs for a look at applications. Last but not least, a possible objection would be to say that there is a world of difference between how these approaches are described on paper and how they are actually applied in the real world. This is indeed a very valid limitation to this paper, but the evaluation of values in the field could very well be seen as the next step to the moral assessment of policy approaches we suggest here.

5.4.1 Three Approaches

The precautionary principle is perhaps the oldest idea to deal with scientific uncertainty. Without going into much detail of its history, it is important to point out that it originates from a few decades back. As several reviews of the PP point out, it comes from German environmental policy in the 70s with the *Vorsorgeprinzip*, quite literally the precautionary principle as it is called today. A little bit down the road, the PP consolidates itself in 1992 in principle 15 of the Rio Declaration (cf. Sunstein 2003; Stirling 2007) under this form:

> In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation. (UNEP)

The precautionary principle is an important and widespread policy approach to deal with scientific uncertainty. However, the PP may be written up or interpreted differently in different documents, laws or contexts (EU COM 2000), and its application ranges from a paralyzing principle to a too permissive one (Sunstein 2003). Moreover, the PP is found to be useful in the management of risk and not in its assessment (Stirling 2007), therefore making it a prime subject for our investigation. There are many versions of the PP in policy terms, and for this paper, we focus on the EU definition and guidelines on how to implement it as presented in its 2000 Communication on the PP, which we will explicate below.

Participatory technology assessment represents a range of methods used to deliberate about technologies and their assessments. Be it consensus conferences, future scenarios, publi-forums, the aim of pTA is to create dialogue and find solutions taking the interests of different social groups into account. pTA came about as a result of realizing the limits of a technology assessment based only on technical considerations and as mentioned earlier, the limits of scientific knowledge, which is seen as necessary but non-sufficient in this approach (PACITA), or in other words, as a response to modernity "in creative ways, new views of problems and preferences in the course of assessment" (Hennen 1999, p. 311). Since there are as many interpretations of pTA as there are technologies in their contexts, the EU mandated a study on pTA to evaluate commonalities to the method and produce recommendations for its implementation, EUROPTA (Klüver et al. 2000). Although pTA is essentially used before a technology is introduced in society, we argue that pTA relates to the management of new technologies as it does not necessarily relate to a technology in particular but to broader societal goals and how to achieve them using technologies, and this is done in a very open way in terms of subject, methods, procedure and participants (unlike EIAs which is why we did not include EIAs in our analysis). The analysis at hand looks at the conclusions and recommendations of EUROPTA as guidelines on what this approach entails.

Adaptive management is a concept stemming from ecology and environmental management (Holling 1978). It has been further described as a case of real-world experiments (Gross et al. 2005). While it is commonly used in the US and Canada as a policy approach to deal with natural resources (US DOI; CEAA), it has only recently been used in Europe for conservation purposes in Natura2000 areas. Indeed, while we found no official European AM management guidelines, we found a European natural resource management network, Eurosite, partnering up with a US think tank, the Foundation of Success (FOS), a US based think-tank specializing in adaptive management all over the world, but also in Europe for N2000 areas.[1] Therefore we chose to use FOS's guideline as the one used in a European context (Salafsky et al. 2001).

Now that we provided the reader with the basic background information on the tree approaches under study, we will proceed into more detail, this time following the values described in the previous section. For simplification, all references made to the PP are from the European Communication on the PP, to pTA from the EUROPTA report and to AM from the FOS tool.

5.4.2 Beneficence and Non-malevolence

It is useful to consider these two higher values simultaneously as they represent two sides of one coin, as they are related in the intentions and consequences of an action, but not opposed to each other: do good *and* do no harm. In order to look into whether the policy approaches under study answer to the higher values of beneficence and non-malevolence, it is particularly relevant to look at the triggers that will favor using each of the approaches, or under which conditions a particular approach is desired.

Indeed, "the precautionary principle is relevant *only* in the event of a potential risk, even if this risk cannot be fully demonstrated or quantified or its effect determined because of the insufficiency or inclusive nature of the scientific data" (EU COM 2000, p. 13). But what defines those cases? The EU communication defines when such triggers occur: after having identified potentially negative side effects, evaluated these under scientific scrutiny and figured out which of these risks have one or more elements of uncertainty, then the PP shall be applied. While

[1]Capacity building for Adaptive Management in Europe, presentation by Neil McIntoshfrom Eurosite available at: http://www.europarc.org/uploaded/documents/324.pdf.

the EU communication defines uncertainty as the knowledge of an event without its associated probability, there seems to be no place for ignorance. We will return to this observation a bit later in the text. From these triggers, it seems that the use of the PP does not primarily appeal to seeing whether a product (or a technology) can be good or bad, but rather whether its uncertainties can be manageable or not, and thereby acceptable or not. Nevertheless, products such as “drugs, pesticides or food additives”, i.e. substances deemed “potentially hazardous at a certain level of absorption”, undergo the “principle of prior approval” (p. 20), in which the burden of proof to show that the products are not dangerous relies on the ones wanting to bring them to the market, so in a sense, showing that it is a not a *bad* product. This does not necessarily mean that it is a *good* product. In addition, ‘the principle of prior approval’ does not apply to all products, in which case, calling for more scrutiny would need to come from other actors in the social experiment.

The PP therefore does not preclude that technologies can be beneficent, but it also does not call for the technology to *be* beneficent. The PP wants to avoid negative side effects rather than to actively promote products that do good. In this way, we argue that the PP answers to non-malevolence rather than to beneficence.

In contrast, pTA and AM show to answer more to both benevolence and non-maleficence at the same time. pTA is carried out as a complement to TA: as the EUROPTA report points out, while TA “speaks truth to power”, pTA aims at “finding solutions together” and “generating dialogue” (Klüver et al. 2000, p. 15). Also, pTA is carried out not as the PP on a particular technology, or product, but rather looks at a technology that *aims* to do something within a system. For instance, discussions around how to better manage traffic, insure clean water, and use biotechnology for modifying plants and animals all have a particular goal that involve using technologies to make *something* better. At the same time, pTA challenges the idea that because a technology *can* make *something* better it *actually* makes it better. Herein are the triggers for using pTA. Firstly, involving a broad number of stakeholders allows defining what a technology means to different groups, and there, the dialogue starts. This is an answer that pTA provides to what the EUROPTA report calls inequality, in terms of access to knowledge, plurality of values and voice (p. 22). Secondly, the level of uncertainty associated with new technologies constitutes the other layer requiring assessment. These triggers can be linked to non-malevolence and justice, but we will focus on non-malevolence for the moment. Indeed, if some groups feel strongly about modifying animals’ genetic material, or about changing the order of traffic in a neighborhood, the introduction of these new technologies might be harming them. In addition to how people might feel or perceive a technology, uncertainties underline the need for scrutiny on its implementation in society. Therefore, pTA looks at technologies that aim at bettering societies, hence it is benevolent, and at creating dialogue and finding solutions to avoid doing harm, hence it is also non-malevolent.

AM is used in settings explicitly dealing with uncertainty and ignorance. Also, the focus of AM is the ecosystem, or the socio-eco-technical system, in which technology, society and environment interact and influence each other. The triggers for using AM are therefore not centered on the technology but rather around the system. If a system is complex and unpredictable (Salafsky et al. 2001, p. 25), these qualities warrant the application of AM. The goals of AM will change depending on the context, the problem and how the involved stakeholders have decided to formulate it (p. 35). They may be linked to societal welfare as well as to environmental health. However, the original intention behind AM is "biodiversity conservation and sustainable management of natural resources" (p. 1). This idea of conservation links to non-malevolence: do not harm the system, and so does the idea of sustainable management. But it also answers to beneficence to a certain extent because sustainably managing it will not only prevent from harm, but will also do *good* for the system for now and in the future.

5.4.3 Learning

Across the three approaches under study, we observe a range of attitudes towards the instrumental value of learning. Also, we choose to discuss it here right after the higher values of benevolence and non-malevolence because learning is not only central to the idea of a social experiment, but also to these approaches. We will describe learning along three main points: what the learning entails, how learning amongst actors in the social experiment happens and is distributed, and how learning about uncertainties and ignorance is tackled.

Firstly, in all three approaches learning entails not only the accumulation of knowledge through different means, but also the incorporation of this knowledge into actions. In the case of the PP, scientific experts do the research that feeds in the learning process and that will help decision-makers determine what kind of actions shall be taken or not taken, in the EU communication, this is called the principle of *examining scientific developments* (EU COM 2000, p. 19). This process is on-going as long as uncertainties persist. Also, there is another form of learning, namely, the PP call for consistency and non-discrimination, which provide for using similar measures in similar cases. In pTA, learning happens mostly as a societal learning process (Klüver et al. 2000, p. 172). This does not exclude the accompanying scientific investigation, but it is not at the center of the learning process for pTA. Another interesting feature of learning in pTA is linked to institutional learning (p. 174). Indeed, learning takes the shape of designing better tools and institutions, or adapting them from other cases, or countries, to facilitate participation. Last but not least, in AM, learning is paramount to the approach. The tool on AM talks itself about how to set-up an experiment in the system for the purpose of testing assumptions and to consider failures, in order to adapt and learn (Salafsky et al. 2001, pp. 44–45). In AM learning therefore entails the use of a scientific method as a management tool, and iteration and adaptation result thereof. Learning does not stop at the scientific inquiry for AM, it also applies to the institutions managing it.

Secondly, in all three approaches, learning could be described as multidirectional, i.e. different stakeholders learning from each other, and to different extents.

Starting with the weakest level of multidirectional learning, the PP calls for experts to do research and advice decision-makers. The only place other actors have are citizens, who may bring attention to problematic products that have not gone through the principle of prior approval, that is if they learn about it. In pTA, learning is stronger, more multidirectional, as stakeholders are highly involved and experts can learn from their concerns and experiences, while stakeholders can also learn to better understand a technology and what it entails. This multi-directional learning occurs through deliberation (Klüver et al. 2000, p. 170). However, one may issue the following critique to pTA as in that it serves as a tool for convincing rather than deliberating. In AM, learning is a process that is defined with all relevant stakeholders as mentioned above. Indeed, goals are formulated as a group, and this in itself involves deliberation. The actual experimentation to learn about the system results from the common definition of the goals. In that sense, it is more than multidirectional learning but more of a common learning. Tensions may arise in defining common goals, but these can be overcome. Also, AM explicitly aims at gathering data, and analyzing it in order to share it with others who might be working on similar projects (Salafsky et al. 2001, p. 80).

Thirdly, all three approaches have differing ways of dealing with uncertainty and ignorance. In a way, all three approaches deal with scientific uncertainty by waiting for scientific advances to specify the unspecified, however in AM, this scientific undertaking is internal to the project as opposed to referring to external experts. The PP deals with ignorance by not dealing with it, and pTA may put emphasis on ignorance depending on the problem at hand. In addition, AM deals with uncertainty and ignorance in the most active way by wanting to find them out, setting-up an experimental design and monitoring to be ready to receive surprises.

Overall, it seems that learning as an instrumental value is most present in AM, then in pTA and finally least present in the PP.

5.4.4 Autonomy and Intervening

We decided to look into autonomy and intervening at the same time because the instrumental value of intervening completes the higher value of autonomy. It is not enough to be considered a moral agent if that same agent is not given the space and means to act. These two values are also particularly important in the conceptualization of new technologies as a social experiment as there is more than one experimenter, so more than one actor can intervene. It is therefore highly interesting to look into whether such an idea is already present in the approaches under study.

The EU communication on the PP makes reference to the Aarhus Convention regarding access to information and justice. As stated in the communication “The Commission has confirmed its wish to rely on procedures as transparent as possible and to involve all interested parties at the earliest possible stage” (EU COM 2000, p. 16). This remains a vague description with regard to the autonomy different actors may have. Decision-makers may have information and more power to act, scientists may have more information and very little power to act, and citizens may find themselves with some access to information (it is not specified how the information is to be accessed) and very little power to act. In that sense the PP answers poorly to the values of autonomy and intervening but at least addresses them in some way.

One would think that pTA would be at the top for these values since the emphasis is on the stakeholder and participation. While we find it fulfills the values of autonomy and intervening to a higher extent than the PP, some pitfalls remain. Indeed, even if deliberation is key to pTA, which strongly supports the value of autonomy, the instrumental value of intervention is left with little importance. Indeed, the report states: "Participatory TA should explicitly be established in order to improve public discourse on, and political opinion forming about, science and technology, with the aim of supporting policy-making with relevant processes and inputs. Participants should not expect to get a decision-making power-base from participatory TA, unless the existing power-structure is represented among the participants" (Klüver et al. 2000, p. 172). This emphasizes the limitations of all actors in the social experiment.

As mentioned earlier, one of the strengths of AM seems to lie in the definition of common goals amongst several actor groups (Salafsky et al. 2001, p. 35). In terms of autonomy, this means all actors share the same level of information and seem to be on equal footing. Also, since the experimental set-up is defined from the goals, one could extrapolate that all actors have the capacity to intervene. However, one should be prudent in such an assertion so while it may be closer to pTA in terms of autonomy, it is probably similar in terms of intervening.

5.4.5 Justice

Interestingly enough, justice is a value poorly explored in all three approaches. The PP mentions justice with respect to the Aarhus convention which has to do with access to knowledge, but this does not address issues of distribution. pTA does have concerns with inequalities but talks about a fair representation of viewpoints and access to knowledge (Klüver et al. 2000, p. 22). All in all, justice issues in pTA and the PP have to do with the procedure but not the distribution and access to benefits of a technology, or additional protection of vulnerable people from negative consequences, which is what we were investigating.

And last but not least, AM mentions that because of the experimental nature of the approach, ethical considerations should be taken into account,

> Adaptive managers must also beware of the potential impact of their experiments on people's lives. In many cases, people's welfare and lives depend on the natural resources generated by ecosystems. It is thus important to consider the implications of withholding treatment that is presumed to be beneficial or trying "second best" practices in control groups — a problem akin to the ethical challenges faced by medical researchers experimenting with human subjects (p. 47).

Table 5.2 Summary table: assessment of values in the three approaches

	Beneficence	Non-malevolence	Autonomy	Justice	Learning	Intervening
PP		+++	+		+	+
pTA	+++	+++	++		++	+
AM	++	+++	++	+	+++	+

No symbol means it is not present, + means somehow present, ++ partly fulfilled, +++ mostly fulfilled

This underlines a concern about justice in the sense that it deals with the distribution of benefits an experiment may produce. In that sense, AM is the only approach that answers to justice as defined earlier. However, there are no recommendations on how to proceed on that front. Table 5.2 provides a schematic overview of our analysis in order to provide a quick overview and shows where existing approaches leave room for improvement for responsible innovation.

5.5 Conclusion

The impetus for this paper came from the realization that introducing new technologies in society could be conceived of as a social experiment, thus extending the scope of responsibility beyond the research and development phase and into the management of the technology once it is out in the world. Responsible innovation also implies the responsible introduction of a technology in society. We therefore asked: How can we responsibly manage the introduction of new technologies in society, and do current approaches exhibit moral values underlying responsibility? We took inspiration from biomedical ethics, drawing from values used when carrying out experiments in the medical context to transfer them to a technological context. Using the higher values of beneficence, non-malevolence, autonomy and justice and adding instrumental values of learning and intervening for the nature of the experiment, we find interesting differences. Before going over these again, we would like to remind the reader that a deficiency in these values does not mean an approach is bad or irresponsible. It only means that in order to provide for more responsibility in the social experiment with new technologies, more can and should be done. Deficiencies can be compensated for either by further developing the approaches, or by using complementary approaches.

The good news from our analysis is that all approaches seek to avoid harm to societies—non-malevolence is fulfilled. However, this does not mean that a new technology will actually do good, as is the case for the PP. Also, all approaches acknowledge the experimental nature of this introduction by fulfilling to different extents the instrumental value of learning, AM being the strongest, followed by pTA and then by the PP. Subsequently, we looked into autonomy and intervening as complementary values and found that it is partially fulfilled. This indicates, however, the need for a stronger democratization of science and technology, and may not have to do with the approaches themselves but rather with the political contexts and institutions in which they are embedded. Chapter 9 by Haen et al. in this volume provides a good example of those limitations. Last but not least, the value of justice presented the greatest deficiency. This is an area where social experiments with new technologies can learn from the medical context, who will benefit from what, and who will be protected from what. Chapter 15 by Balkema and Pols in this volume also provides an interesting example of a context in which there is room for more justice within responsible innovation.

All in all, while the current approaches do not provide for a complete responsibility fully reflecting the relevant higher and instrumental values, they do exhibit several of the required values for responsibility to some degree. This analysis therefore opens the door for a discussion on how to continue improving them and innovate more responsibly in society.

Acknowledgments This paper was written as part of the research program 'New Technologies as Social Experiments', which is supported by the Netherlands Organization for Scientific Research (NWO) under grant number 016.114.625. We would also like to thank Ibo Van de Poel, Neelke Doorn, Shannon Spruit, and Jan Bergen for their valuable input in the making of this paper.

References

Beauchamp, Tom L., and James F. Childress. 2001. *Principles of biomedical ethics*, 5th ed. New York: Oxford University Press.

Beck, Ulrich. 1992. *Risk society: towards a new modernity*. London: Sage Publications Ltd.

CEAA. http://www.ceaa-acee.gc.ca/default.asp?lang=En&n=50139251-1. Accessed 27 Feb 2013.

Clouser, K.Danner, and Gert Bernard. 1990. A critique of principlism. *Journal of Medicine Philosophy* 15(2): 219–236.

Collingridge, David. 1980. *Social control of technology*. London: Frances Pinter.

Commission of the European Communities. 2000. *Communication from the commission on the precautionary principle*. Brussels: Commission of the European Communities.

Commission, European. 2012. *Ethical and regulatory challenges to science and research policy at the global level*. Luxembourg: Publications Office of the European Union.

Felt, Ulrike, Brian Wynne, Michel Callon, Maria Eduarda Gonçalves, Sheila Jasanoff, Maria Jepsen, Pierre-Benoît Joly, Zdenek Konopasek, Stefan May, Claudia Neubauer, Arie Rip, Karen Siune, Andy Stirling, and Mariachiara Tallacchini. 2007. *Taking European knowledge society seriously. Report of the expert group on science and governance to the Science, Economy and Society Directorate, Directorate-General for Research, European Commission*. Brussels: Directorate-General for Research, Science, Economy and Society.

Gross, Matthias, Holger Hoffmann-Riem and Wolfgang Krohn. 2005. *Realexperimente: ÖkologischeGestaltungsprozesse in der Wissensgesellschaft*. Bielefeld: transcript Verlag.

Harvey, Nick, and Beverley Clarke. 2012. *Environmental impact assessment in practice*. South Melbourne: Oxford University Press.

Hennen, Leonhard. 1999. Participatory technology assessment: a response to technical modernity? *Science and Public Policy* 26(5): 303–312.

Holling, C.S. 1978. *Adaptive environmental assessment and management*. London: Wiley.

Jacobs, Johannes. F., Ibo Van de Poel, and Patrica Osseweijer. 2010. Sunscreens with titanium dioxide (TiO_2) nano-particles. A societal experiment. *Nanoethics*. doi:10.1007/s11569-010-0090-y.

Klüver, Lars, Michael Nentwich, Walter Peissl, Fritz Gloede Helge Torgersen, Leonhard Hennen, Josée van EijndhovenRinie, Simon van Est, Joss Sergio Bellucci, and Danielle Bütschi. 2000. *EUROPTA: European participatory technology assessment participatory methods in technology assessment and technology decision-making*. Copenhagen: The Danish Board of Technology.

Krohn, Wolfgang, and Peter Weingart. 1987. Commentary: nuclear power as a social experiment-European political "fall out" from the Chernobyl meltdown. *Science, Technology and Human Values* 12(2): 52–58.

Krohn, Wolfgang, and Johannes Weyer. 1994. Society as a laboratory: the social risks of experimental research. *Science and Public Policy* 21(3): 173–183.

Latour, Bruno. 1993. *We have never been modern*. Harvard University Press: Cambridge: Harvard University Press.

Levidow, Les and Susan Carr. 2007. GM crops on trial: technological development as a real-world experiment. *Futures* 39 (4): 408–431.

Mann, Carsten, Jan-Peter Voß, Arno Simons, Nina Amelung, and Till Runge. 2014a. *Challenging futures of biodiversity offsets and banking. Critical issues for robust forms of biodiversity conservation a report based on an interactive, anticipatory assessment of the dynamics of governance instruments, 19 April 2013.* Technische Universität Berlin.

Mann, Carsten, Jan-Peter Voß, Nina Amelung, Arno Simons, Till Runge, Louisa Grabner. 2014b. *Challenging Futures of Citizen Panels. Critical Issues for Robust Forms of Public Participation. A Report Based on an Interactive, Anticipatory Assessment of the Dynamics of Governance Instruments, 26 April 2013. Berlin*. Technische Universität Berlin.

Martin, Mike W., and Roland Schinzinger. 1983. *Ethics in engineering*. Boston: McGraw-Hill.

PACITA: Parliaments and civil society in technology assessment. http://www.pacitaproject.eu/. Accessed 27 February 2013.

Perdicoulis, Anastassios, B. Durning and Lisa Palframan. 2012. Furthering environmental impact assessment: towards a seamless connection between EIA and EMS. Cheltenham: Edward Elgar Publishing Limited.

Salafsky, Nick, Richard Margoluis, and Kent Redford. 2001. *Adaptive management: a tool for conservation practitioners*. Washington, D.C.: Biodiversity Support Program.

Schwarz, Astrid and Wolfgang Krohn. 2011. Experimenting with the concept of experiment: probing the epochal break. Experimenting with the concept of experiment: probing the epochal break. In: *Science transformed? Debating claims of an Epochal Break*, eds. A. Nordmann, H. Radder, and G. Schiemann, 119–134. Pittsburgh: University of Pittsburgh Press.

Stirling, Andrew. 2007. Risk, precaution and science: towards a more constructive policy debate. *EMBO Reports* 8(4): 309–315.

Sunstein, Cass R. 2003. Beyond the precautionary principle. *University of Pennsylvania Law Review* 151(3): 1003–1058.

UNEP. 1992. Rio declaration on environment and development. http://www.unep.org/Documents.Multilingual/Default.asp?documentid=78&articleid=1163. Accessed 27 Feb 2013.

US DOI. http://www.doi.gov/initiatives/AdaptiveManagement/whatis.html. Accessed 27 Feb 2013.

Van de Poel, Ibo. 2011. Nuclear Energy as a Social Experiment. *Ethics, Policy and Environment* 14, (3) 285–290.

Von Schomberg, René. 2012. Prospects for Technology Assess-ment in a framework of responsible research and innovation. In: *Technikfolgen abschätzen lehren: Bildungspot-enziale transdisziplinärer Methoden*, eds. M. Dusseldorp and R. Beecroft, 39–61. Wiesbaden: Springer.

Voß, Jan-Peter and Arno Simons. 2014. Instrument constituencies and the supply side of policy innovation: the social life of emissions trading. *Environmental Politics* 23 (5). Open access 10.1080/09644016.2014.923625.

Chapter 6
The Observatory for Responsible Research and Innovation in ICT: Identifying Problems and Sharing Good Practice

Bernd Carsten Stahl, Grace Eden, Catherine Flick, Marina Jirotka, Quang A. Nguyen and Job Timmermans

Abstract The implementation of responsible research and innovation (RRI) with the aim of ensuring socially acceptable and desirable outcomes of research and innovation activities requires coordinated action by numerous actors. RRI may be conceptualised as a network of interlinking responsibilities, some of which have long been established, others that will have to be defined. Actors in these networks of responsibilities will require knowledge about possible activities, normative foundations and good practice that they currently are unlikely to possess. In order to provide a platform for the exchange of knowledge and good practice that different actors can use, the UK EPSRC-funded project on a Framework for Responsible Research and Innovation in ICT is developing an Observatory meant to be a community-based resource that can provide the resources required by stakeholders of ICT research. This chapter describes the way the system is developed and tested. In reflecting upon the development process of the observatory, the chapter provides insights into how the broader discourse on responsible innovation could benefit from this type of resource.

Keywords Responsible innovation · Observatory · Community

B.C. Stahl (✉) · C. Flick · Q.A. Nguyen · J. Timmermans
Centre for Computing and Social Responsibility, De Montfort University, Leicester LE3 9EP, UK
e-mail: bstahl@dmu.ac.uk

G. Eden · M. Jirotka
Department of Computer Science, University of Oxford, Parks Road, Oxford, OX1 3QD, UK

M. Jirotka
e-mail: marina.jirotka@cs.ox.ac.uk

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_6

6.1 Introduction

Responsible Research and Innovation (RRI) can be understood as a higher level responsibility or meta-responsibility that aims 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. This initial definition shows that RRI is a highly complex social construct that involves a potentially large number of actors and stakeholders who are involved in research and innovation processes. The view of RRI as a meta-responsibility is furthermore consistent with most, if not all, of the current strands of debate and thus viable despite the persistent lack of a generally agreed definition (see Koops, this volume, Chap. 1).

RRI raises a number of conceptual questions and the budding discourse around the concept point to many of these. There are epistemological questions concerning the knowledge of the future and future uses innovation. There are normative questions concerning a shared view of what counts as acceptable and desirable and on what basis such judgments can be made. There are regulatory and legal challenges with regards to governance and institutions of RRI. These and others are discussed in this volume and elsewhere (Owen et al. 2013). In addition to such difficult fundamental questions RRI raises a number of practical issues. Key among them is that most of the actors in research and innovation (R&I), such as researchers, companies and research funders are often not clear what it would mean to act responsibly or which responsibilities are ascribed to them.

The present chapter describes a practical intervention that aims to provide support for these R&I actors. Arising from a UK EPSRC-funded research project dedicated to developing a "Framework for Responsible Research and Innovation in ICT", the chapter describes the process of developing what is currently termed "The Observatory for Responsible Research and Innovation in ICT" or TORRII. The project resulted from a workshop held in late 2010 where the UK Engineering and Physical Science Research Council (EPSRC) invited a number of researchers to give input into the development of its funding strategy in the area of information and communication technology (ICT). Two of the authors of the chapter attended this workshop and agreed to develop a proposal for a project to investigate how the ICT research community currently interprets its responsibilities, to develop a network of interested scholars, and to create the TORRII system. The initial idea of TORRII was to design a web-based platform to enable the sharing of experience and good practice and to serve as a resource for ICT scholars to use when faced with responsibility-related questions. It is intended to be a community-based resource, i.e. contain content that is made by members of the ICT research community and other stakeholders for the benefits of others.[1] The term "observatory"

[1] ICT researchers are a key stakeholder group of RRI in ICT and they are the primary focus of the FRRIICT project.

underlines that the resources within the system are not centrally produced for determined consumption but reflects ongoing discourse and reflects practical experience.

This chapter recounts the step in the development of the system, focusing on the first larger scale user test undertaken for TORRII. In order to show why and how the system was developed, the chapter starts with a discussion of the conceptual underpinnings of RRI which explain the choice of content structure of the system. This leads to a more detailed discussion of the purpose of observatory and the principles and strategy of systems development. The chapter then describes the first major interaction with potential users and how data from this activity was analysed and fed back into the systems development. This leads to a discussion of current opportunities and challenges. The conclusion returns to the broader context and points to a way forward.

The chapter is specific to a particular system arising from a particular project. It therefore contains some of the idiosyncrasies related to this specific context. We nevertheless believe that the chapter touches on an issue of broader importance for the entire discourse of RRI. No matter how RRI will eventually be defined, it is likely to lead to expectations and responsibilities concerning a number of R&I actors. Many of these will be novel and the actors' experience and training will not have prepared them for it. In these cases the different actors will require resources and support. A number of chapters in this volume point to particular challenges of RRI in particular settings and situations. Whether the topic of research and innovation is cognitive enhancement (Maslen et al., this volume, Chap. 7), neuroimaging (Edelenbosch et al., this volume, Chap. 8), food (Haen et al., this volume, Chap. 9) or energy (Correljé et al., this volume, Chap. 10), new challenges will arise that require researchers and other stakeholders to receive support and guidance to help them act responsibly. Some of this support is likely to take the form of a community repository of the type that the Observatory system is aiming to establish. We therefore believe that there is a need for a more general "Observatory for RRI" that gives actors an opportunity to share experiences and good practice. The development of such a broader resource will be able to benefit from the experiences collected in the FRRICT project.

6.2 Conceptual Background

The purpose of the current chapter is to describe a practical intervention that aims at facilitating the definition and realisation of responsibility relationships by a variety of actors who are engaged in a broad array of activities in research and innovation. The chapter nevertheless needs to clarify its conceptual and theoretical position on RRI in order to demonstrate how the practical intervention, the development of the Observatory of Responsible Research and Innovation can fulfill this function. The chapter therefore briefly reviews aspects of the current discourse on RRI that justifies the definition of the term given in the opening sentences. This is then used to identify key aspects of RRI that actors need to be aware of in order to undertake R&I responsibly.

Responsible Research and Innovation has been defined 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)” (Von Schomberg 2011). It is important to realise that RRI does not exist in a vacuum. It is based upon and incorporates numerous existing responsibility relationships. We suggest that it is best be understood as a meta-responsibility that aims to align and, where necessary, modify and develop responsibility relationship with a view to ensuring that novel scientific and technical developments are conducive to the shared and justified aims of society (Stahl et al. 2013).

One field of research and innovation that has a high potential to affect people’s rights and obligations and is therefore considered an important subject of RRI is that of Information and Communication Technologies (ICTs). ICTs are increasingly ubiquitous and can have numerous positive as well as negative consequences. Because of their “logical malleability” (Moor 1985, p. 269), the fact their uses are variable and difficult to predict, ICTs pose particular challenges for RRI. At the same time ICT researchers tend to come from a multitude of backgrounds, which renders it difficult to establish shared norms and expectations. Finally, ICT is one of the so-called “converging technologies”, which means that aspects of ICT are visible in much other research from materials and nanotechnology to cognitive and biological sciences (Grunwald 2007; Roco and Bainbridge 2007). The question of how individuals using ICT in their research can identify RRI-relevant issues and communicate their experience and share good practice is thus a core question that has repercussions in other fields of R&I.

6.3 Identifying Issues and Sharing Good Practice: The Observatory for Responsible Research and Innovation in ICT

The complexity of moral issues and their ethical evaluation in the area of ICT has been recognised since the early stages of the development of digital computing technology (Wiener 1954). It has led to an on-going academic debate linking moral philosophy, social sciences and technology research and development (Bynum and Rogerson 2003; Floridi 2010; Himma and Tavani 2008; van den Hoven and Weckert 2008). The broad range of issues and possible solutions implies that it is difficult for any one individual to understand and evaluate these issues. Moreover, one of the principles of RRI is that research and innovation need to be accompanied by public engagement and exchange with stakeholders (Owen et al. 2012).

For these reasons, the prior research has led to the policy recommendation for development of an “Observatory” that will facilitate interested stakeholders in the exchange of views and experience with regards to RRI in ICT (see www.etica-project.eu). This idea was taken up by the UK EPSRC funded research project "Framework for Responsible Research and Innovation in ICT" (FRRIICT; www.responsible-innovation.org.uk) which, amongst other activities, will design a prototype of such an observatory. The aim is that this Observatory will be a community-owned online resource for ICT researchers and other stakeholders involved in ICT research and innovation.

It is probably not contentious to state that the existence of such a resource would be desirable. What is more difficult to say is what shape it should take, which functionality it should have, how it should be developed and which content it should contain.

6.3.1 Principles Underpinning the Observatory

In order to understand the principles of the Observatory, it will help to see it in the context of the FRRICT project. The FRRIICT project aims to build a researcher network focussed upon ethical issues ICT through a baseline study to understand and analyse current perceptions of ethics in ICT (Eden et al. 2013). This was done by interviewing a representative cross-section of the ICT community.

The overall aims of the project are to:

develop an in-depth understanding of ICT researchers' ethical issues and dilemmas in conducting ICT research;

provide of a set of recommendations and good practice to be adopted by EPSRC and the community;

create a self sustaining 'ICT Observatory' serving as a community portal and providing access to all outputs of the project.

This chapter refers mainly to the final point, but it is important to understand that this is embedded and draws on the other activities of the project. The findings of the other work packages of the project including the landscape research, case studies, and the network discussions will only be relevant if they can be communicated to the entire ICT community. In addition, the area under investigation is highly dynamic with technologies as well as ethical views and perceptions changing over time. In order for the different research aspects of the project, which can only ever constitute snapshots, to remain relevant, a responsive way of capturing, communicating and disseminating is required. The project intends to achieve this by developing the Observatory which will be a repository of findings from the project and the extent to which it may be owned by and remain accessible to the ICT community will be explored. It will be designed according to the needs of the ICT community and is will to contain the following information:

Current and emerging ICTs and likely ethical issues that may arise.
Case studies and application exemplars of ethical issues.
Governance arrangements suitable to address these issues.
A list of experts on any of the issues identified.

The Observatory will be evaluated by the network community and will then be opened up to the entire ICT community in the UK and abroad. It will be a dynamic and community-owned tool developed using agile approaches to provide an innovative, flexible and robust community support (Bennet et al. 2010). It is likely to employ established and novel technologies, including some of the collaborative tools that are summarised under the heading of Web 2.0.

The Observatory will be the operationalised legacy of the FRRIICT project and is designed to render it relevant to individuals within the ICT research community and beyond. It will allow researchers to identify and address ethical problems early, to explore suitable project governance mechanisms and to ensure that ethics reviews are contemplated and passed successfully. Depending upon the needs of the community, the observatory could monitor what is happening in ICT, identify progress across the domain, produce summary statistics and from these produce bespoke reports for specific communities. It may submit data to public bodies for consultations and investigations and as such form a repository for policy development on the unique ethical issues of ICT that could be taken into consideration. It will allow funders, such as the EPSRC, to define standards of RRI that ICT research may need to achieve. A core aspect of the legacy will be to ensure that future generations of ICT researchers understand the relevance of RRI to their work. The Observatory will contribute to this by providing current and pertinent cases and materials that can be used for teaching and education of undergraduate, postgraduate and research students. Furthermore the Observatory will allow researchers to identify areas in need of more detailed investigation and it will show where public engagement on particular issues may be relevant.

The Observatory should be seen in the broader context of RRI currently unfolding across Europe. While it is focused in its initial stages on issues concerning ICT research and development and pays particular attention to the community of researchers, it should be clear that these represent only a limited subset of stakeholders and concerns relevant to RRI. Similar concerns are likely to arise in other scientific disciplines and cover other stakeholders. It is therefore expected that further similar systems will be developed and the overall RRI discourse should ask how these different systems relate to one another. These concerns go beyond the present chapter which focuses on the development of the Observatory prototype as described in more detail below.

6.3.2 Development Approach

In order to ensure that principles of RRI would be incorporated into an open and participative development approach, the FRRIICT team decided to pursue an agile development approach in the design of the Observatory. Such an approach requires frequent feedback from users to allow for targeted development of the system to meet user needs. In practice this raises the questions of who might the users be and how can they be motivated to contribute to the development of the Observatory?

After an initial specification of the system by the FRRIICT team, feedback was sought from potential stakeholders including ICT researchers, researchers using ICT as well as professional bodies, ICT research funders and policy makers.

On the basis of these initial discussions the development team designed a set of initial use cases and application scenarios. These were used to gain a more detailed understanding about what types of activities the Observatory might support and which types of users would require which services.

This conceptual analysis was supported by a review of comparable and possibly competing systems and websites. The following table contains the sources explored to find out which content and functionality each provided (Table 6.1).

A separate document was then developed that compared key aspects of these systems, including themes, target groups, contribution, funding, authors, content source, content types, incentives for contribution, data presentation, look and feel and business model. This analysis was part of the design process and informed decisions about content and presentation of the Observatory.

One key question discussed at this stage concerned the type of content and the structure of entries. Additionally, the early stage included discussions of principles of the Observatory, such as the way in which decisions about design and development were made, the way in which users were to be involved or the distribution of intellectual property related to the system.

The following figure shows an overview of the entities that the project team thought to be likely to be useful to be included in the scope of the Observatory.

On the basis of these discussions an initial prototype was developed that allowed for input of resources and retrieval.

6.4 User Testing

As the Observatory is meant to be a community-based system, it was seen as important to include feedback from potential user communities as early as possible. This caused the predictable problem that the team was trying to test something before it was clearly visible what the system to be tested will eventually represent. This is a problem that this common to all agile development approaches (Bennet et al. 2010). It is not a fundamental problem, however, because the engagement of uses in iterative system design is meant to lead to a prototype that is useful to the user community. In this chapter we concentrate on the first major event that was used to engage with one particular user community, namely digital social researchers.

Table 6.1 List of observatory websites analysed during the development

	URL	Name	Organisation
1.	http://www.unesco.org/new/en/social-and-human-sciences/themes/global-ethics-observatory/	Global Ethics Observatory	UNESCO
2.	http://www.debatinginnovation.org/	Observatory for Responsible Innovation	MINES ParisTech
3.	http://www.onlineethics.org/	Online Ethics Center (OEC)	National Academy of Engineering
4.	http://ethics.iit.edu/eelibrary/	Ethics Education Library	Illinois Institute of Technology
5.	http://www.scu.edu/ethics/	Markkula centrer for applied ethics	Santa Clara University
6.	http://www.chistera.eu/	European Coordinated Research on Long-term Challenges in Information and Communication Sciences & Technologies ERA-Net.	FP7 project
7.	http://www.ethicsguidebook.ac.uk/	Research Ethics Guidebook	Institute of Education, University of London
8.	http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/index.htm	Ethics and research guidance	Medical Research Council
9.	http://dornsife.usc.edu/dilemmas-and-case-studies/	Ethics Resource Center	USC Levan Institute
10.	http://caseplace.org/	CasePlace.org	The Aspen Institute Center for Business Education
11.	http://www.spj.org/ethicscasestudies.asp	Ethics Case Studies	Society of Professional Journalists
12.	http://sisob.lcc.uma.es/sisob/index.php	SISOB	FP7 project
13.	http://www.p3gobservatory.org/repository/ethics.htm	Ethics, Governance and Public Engagement	Public Population Project in Genomics
14.	http://www.ethicsweb.eu/ere/	European Research Ethics Website (ERE)	German Reference Centre for Ethics in the Life Sciences
15.	http://www.changethegame.org/ethics-observatory	Ethics Observatory	Change the Game Open innovation hub
16.	http://www.oecd.org/mena/governance/observatoryonethicscodesandcodesofconductinoecdcountries.htm	Observatory on Ethics Codes and Codes of Conduct in OECD Countries	OECD
17.	http://www.observatorynano.eu/project/	ObservatoryNANO	European Nanotechnology Gateway
18.	http://europa.eu/rapid/press-release_IP-13-113_en.htm?locale=en	Bioeconomy observatory	European Commission/Joint Research Centre

6.4.1 Data Collection

A good way of exploring user needs is to physically assemble them and ask them to engage with the topic. This was facilitated by a grant of £5000 from the Digital Social Research[2] fund for a training event on "Identifying and addressing ethical issues in technology-related social research." The grant allowed the FRRIICT team to invite digital social researchers in the UK to a two day event hosted at the Oxford eResearch Centre. Twenty external researchers attended the event. These were mostly social scientists working on ICT-related issues or researchers with a technical background using social science approaches or methods. Most had encountered ethical issues in their work and were interested in receiving more information and skills for identifying and addressing ethical issues.

From the perspective of the participants the workshop had three main goals that would allow researchers to:

Gain a better understanding of how to identify and describe ethical issues in technology-related social research;

Collaboratively develop solutions to some shared ethical problems;

Use ICT tools (i.e. the Observatory for Responsible Research and Innovation in ICT) to identify problems and find appropriate solutions.

When registering for the workshop participants were asked which experience they had of ethical issues with regards to ICT. This allowed the FRRIICT team to develop a set of four case studies based upon the queries or experiences of the participants. A case study approach to discussing issues of RRI in ICT was deemed appropriate because it allowed participants to apply their various areas of expertise and collaborate on the identification and resolution of ethical issues. The case studies developed for the workshop were fictional but based on prior experience and ongoing research by the FRRIICT project team. The titles cases[3] were:

(A) Using Twitter data for market research for a new iPhone app
(B) Using the UK open data initiative for comparative geo-spatial analysis
(C) Joining an underground forum (e.g. Silk Road) to do research on drug distribution
(D) Developing a digital sensory room for a hospice

The cases were chosen because they offered a wide array of possible issues related to ethics and responsibility. Furthermore, the cases linked to a number of the research interests expressed by the participants.

To facilitate detailed engagement with the cases and allow participants to develop their understanding of the problems, the workshop was organised through group activities with participants placed into four groups, each of which was accompanied by a member of the FRRIICT project team.

[2]http://www.digitalsocialresearch.net/wordpress.

[3]The case studies are available on the Observatory: http://torrii.responsible-innovation.org.uk/case-studies.

The approach described here is in many ways similar to teaching of ethics and ICT through case studies that is used widely (Bynum 2004). There are, however, several aspects that set apart the workshop described here from normal classroom-based teaching. To start with, the participants were active researchers, partly PhD students and partly established researchers employed by universities and research institutions. They were therefore able to draw on much richer insights into ethics and ICT and the cases were constructed specifically with a view to reflecting the ethical issues they had encountered previously. Furthermore, the workshop was less structured than one would expect in a traditional teaching environment. Participants had two full working days to engage with the material. The time slot for the discussion of the first case was allocated three hours, which gave participants the opportunity to engage with the case in depth. The open nature of the discussions also facilitated broader reflection related to different aspects of RRI in ICT. While the guiding questions outlined below did provide some structure, the idea was to engage participants in a broader debate about the operationalising of RRI from identification of ethical issues in a particular case all the way to policy development. Additionally, the workshop referred specifically to the Observatory with participants given the opportunity to explore the system and use the exercise to reflect upon important and desirable functionality that such a resource might offer.

6.4.2 Findings

The four participant groups discussed the case studies with a view to identifying ethical issues and they were asked to consider the following questions:

What are the ethical issues of the case?
What makes them ethical issues?
Who are the stakeholders?
Did the stakeholder analysis change the initial ethical evaluation?
How would the ethical issues you described normally be addressed?
How do you think the ethical issues should be addressed?

In a separate session they were then asked to think about specific solutions and whether these could be generalised by answering the following questions:

Can you suggest a protocol for other researchers in similar cases?
What would you need to develop this further (knowledge, tools,…)
Can you deduce general rules from the case? If so, which ones?

Each group was given one of the case studies during the first day of the event. They were then invited to discuss these questions in detail and present their findings and thoughts to the full group. The groups were then asked to consider which types of tools they would require in order to address the issues identified. On the second day the groups exchanged case studies and undertook a similar, albeit shorter, exercise to ensure that there was cross-fertilisation of ideas between the groups.

The workshop discussions were presented by the individual groups using structured slides that were made available through Google docs. This had the advantage of allowing the participants to see each other's contributions, to focus on shared group outcomes and, from the perspective of the organisers, it was a simpler way of collecting all the outputs without having to separately transcribing them. After the workshop all group discussions were uploaded to a separate NVivo file for further data analysis. In addition, all users were invited to provide feedback at any stage during the event and with regards to all aspects. This was implemented by a global survey. Participants had access to all the different files through a dedicated Google site Web page.

All outcomes and feedback from the workshop were analysed using principles of Grounded Theory (Strauss and Corbin 1997; Charmaz 2006) to allow for a broad coverage of important aspects. Some of the interesting highlights that were gathered from this event include:

1. Participants had no problems identifying ethical issues but often found it difficult to state precisely why something would count as an ethical issue.
2. The distinction between ethical and legal issues proved difficult. Privacy and data protection is a key example where significant amounts of legislation related to data protection supersedes ethical concerns.
3. Asking participants to undertake a detailed analysis by stakeholder helped develop a more fine-grained understanding of ethical issues in a case.
4. Practical concerns with gaining required ethical approval based on biomedical research ethics procedures were often on the forefront of participants' concerns. This could lead to less attention being paid to the more ambiguous and difficult to manage questions related to the ethics of research outcomes.
5. There was very little recourse to explicit ethical theories even though specific types of reasoning corresponding to ethical theories were visible.

These findings are of general interest to the discourse concerning RRI. They suggest that even an audience of individuals who have shown a clear interest in the topic struggle with fundamental conceptual issues. The prior framing of ethics and responsibility in ICT-related research tends to come from the dominant ethics review paradigm, which in turn, are an outcome of biomedical ethical principles (Beauchamp and Childress 2008). While these are firmly established and the principles of informed consent enshrined in biomedical ethics is important, it is not obvious whether and to what degree this paradigm alone should be the guiding principles of ICT research.

Another important aspect of these findings is that practical recommendations and governance suggestions will need to take the context of the specific research question and its organisational embedding into account. This suggests that RRI will need to develop procedural approaches that are context sensitive. A number of approaches exist that attempt to incorporate ethical considerations into research and development such as value-sensitive design and privacy by design. An important question will be to what degree such approaches can be scaled up and made available for research and innovation activities more broadly.

Another key issue is the relationship between ethics and law. This is a question that has entertained legal philosophers for a long time and is unlikely to be settled any time soon. However, our workshop indicated clearly that participants' positions with regards to the acceptability of research and its outcomes are strongly influenced by their views on what is legal. Furthermore, a number of possible governance arrangements that would help operationalise RRI involve legislation, thereby leading back to the question of the relationship of ethics and law with regards to responsibility in research and innovation.

Finally, it is important to note that some of the key characteristics of RRI, namely the question of broader societal acceptability and desirability of a particular piece of research, were generally deemed to be outside of the scope of what the researcher can determine. This means that it is likely to require significant efforts if the ideas of acceptability and desirability are to be embedded into the conceptualisation and practice of RRI.

In addition to these substantive insights into perceptions and possible practices of RRI, the workshop shed light on the question how and to what degree an Observatory might be helpful to participants. The workshop participants were given an opportunity to explore an early prototype of the Observatory, to register as users, and to upload and download materials. The general feedback given by participants was that the Observatory is a principally useful resource, although, at its current stage of development it has faced usability issues. In addition there was relatively little content, which is partly explained by the fact that it is to be a user-driven system and so requires community buy-into be successful.

Overall the workshop proved to be relevant to the participants and provided important input into the evaluation and further development of the Observatory. The case study-based approach was shown to be appropriate as demonstrated by the following feedback by one of the participants: "I found this event very helpful—particularly the group discussion of the case studies. These often seemed straightforward to me at first, but talking with the rest of the participants frequently revealed ethical problems that I had not considered. I will return to my work with a greater appreciation for the complex and subtle nature of many ethical issues, and this will enable me to do a better job identifying them in my own research. I also look forward to following the development of TORRII [the Observatory], which has the potential to be a powerful platform for collaboration and knowledge-sharing, and with which I would very much like to participate." The feedback collected throughout the event furthermore pointed to other areas of interest, such as international and intercultural issues which were deemed to be outside the scope of the UK-based FRRIICT project but which are arguably central to RRI more generally.

6.5 Further Steps

The workshop described above provides a clear indication that the Observatory is viable and that potential users do see its value. While it has demonstrated that the FRRIICT project is going in the right direction, numerous questions remain open.

A core question with regards to the Observatory is what its content should be and how it is to be created. On the one hand the principle of development of the Observatory was bottom-up. This would suggest that content should originate from stakeholder and user communities, primarily from the ICT research communities. On the other hand there remains the problem that many of the issues related to RRI in ICT require detailed subject knowledge such as knowledge of ethics or foresight methods that goes beyond the disciplinary knowledge of ICT researchers. A related question is how content should be solicited. During the year 2012 the FRRIICT project launched a call for case studies where ICT researchers were asked to submit applications for up to £5000 of funding to develop relevant case studies. This call was hugely oversubscribed with more than 40 applications received and eight case studies funded. At the time of writing this chapter there is a second call open which is seeking submissions to different categories of the Observatory's, i.e. case studies, ethical issues, technology descriptions, concepts or solutions. It is hoped that this call will lead to a large number of high-quality contributions that will render the content of the Observatory more interesting and relevant to the intended users.

The funded calls so far have proven to be successful but they raised the question of how future contact can be developed. It also raises the question of the

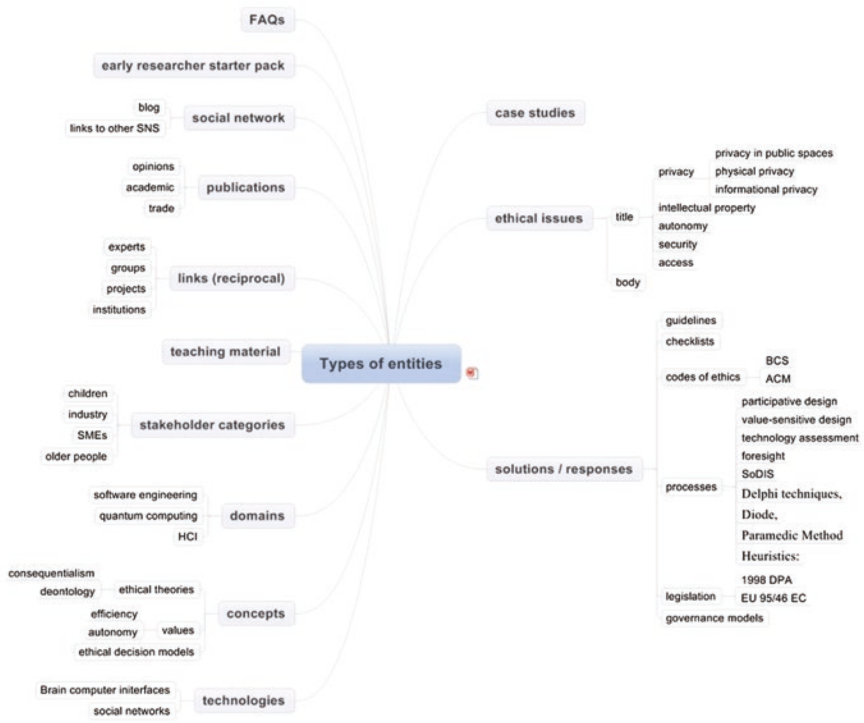

Fig. 6.1 Initial list of entities of the observatory

legacy and future of the Observatory and is currently being discussed within the project team. A key aspect of this is the need for further funding required to both maintain the system and to continue to develop it according to user needs. An initial market analysis suggests that users would prefer a solution where the Observatory remains in a not-for-profit environment. What shape this would take eventually is currently undecided.

A further question that will have a strong influence on the continued success of the Observatory will be how it can be linked to stakeholder and user communities. The system should provide support to users but it would benefit from further institutional embedding, for example through direct references to the Observatory by research funders, who could use it to promote RRI-related policies and practices.

At this stage of the project, the team is concentrating on the further development of the look and feel of the Observatory as well as adding more content. On the basis of user feedback from the workshop described here and from other user engagement activities the team has drawn up a set of wireframes that represent possible appearances of different aspects of the Observatory (Fig. 6.1). Similar wireframes have been developed for other pages and will be presented to users to gain further feedback (Fig. 6.2).

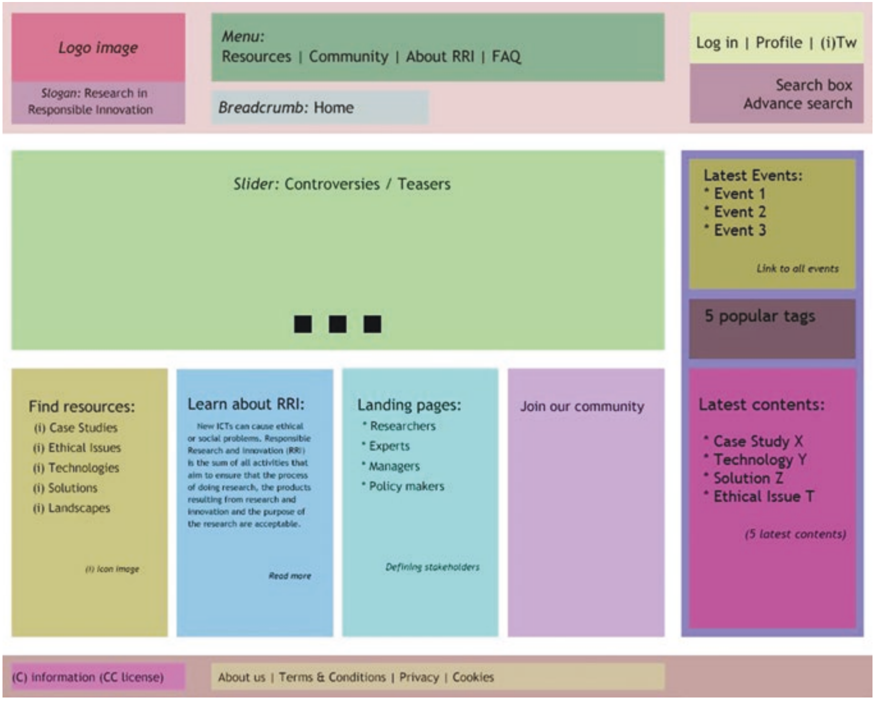

Fig. 6.2 Wireframe of the observatory homepage in next stage of development

6.6 Conclusion

This chapter relies heavily upon one particular event with all the ensuing problems, such as lack of representativeness, self-selection bias, etc. The workshop described here made use of brief and mostly fictitious case studies that may have skewed participants' views. The findings and interpretations can therefore not claim to be representative but they nevertheless provided interesting insights into the possible need for an Observatory to support RRI. This chapter shines light on important aspects of RRI in ICT. On the one hand it shows that there is an interest in the topic and that digital social researchers, one important stakeholder group in the area of ICT research, are keen to learn more about RRI and ways to put it into practice. The Observatory tool that the FRRIICT project is developing was generally viewed as desirable but still raises many questions with regards to content, design and the identification of key user groups.

At the same time the user feedback highlighted some of the issues that RRI in ICT still faces. RRI builds on existing responsibilities and mechanisms of dealing with ethics in ICT, such as professionalism, computer ethics or existing regulations. It is not clear in what way, if at all, RRI goes beyond these existing discourses. Much of the attention to ethics in ICT is generated by requirements to comply with ethics review procedures, which raises the question of the relationship of RRI with such procedures.

The experience of the FRRIICT project so far indicates that researchers are most likely to engage with RRI if they can see the relevance to their activities. Therefore, we suggest that areas of RRI which focus on specific disciplines might be core to the success of the overall RRI discourse.

A final lesson arising from this paper that goes beyond the subject area of ICT is that RRI will need to develop and embed tools that allow actors to enage responsibly in research and innovation. Such tools can take many forms. We suggest that the idea behind the Observatory, namely to develop a bottom up repository that will facilitate stakeholders in the exchange of both experience and good practice will need to spread into different areas for RRI to be successful. We hope that the present chapter has shed some light on possible approaches to building such a system, ways to engage with potential users and some of the practical considerations and problems that one needs to engage with in order to develop such a system.

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

The workshop on which the paper is based was funded by the Digital Social Research Community Activities Funding Scheme Ref no 0003.

The Observatory is available at http://www.responsible-innovation.org.uk. It contains all of the case studies used in event described in this chapter.

References

Beauchamp, T.L., and J.F. Childress. 2008. *Principles of biomedical ethics*, 6th ed. USA: OUP.

Bennet, S., S. McRobb, and R. Farmer. 2010. *Object-oriented systems analysis and design using UML*, 4th ed. Maidenhead: McGraw-Hill Higher Education, c.

Bynum, T.W. 2004. Ethical decision-making and case analysis in computer ethics. In *Computer ethics and professional responsibility*, ed. T.W. Bynum, and S. Rogerson, 60–86. Oxford: Blackwell.

Bynum, T.W., and S. Rogerson. 2003. *Computer ethics and professional responsibility: introductory text and readings*. New York: Wiley, Blackwell.

Charmaz, K.C. 2006. *Constructing grounded theory: a practical guide through qualitative analysis*. New York: Sage Publications Ltd.

Eden, G., M. Jirotka, and B.C. Stahl. 2013. *Responsible research and innovation: critical reflection into the potential social consequences of ICT*. Paris: In Research Challenges in Information Science (RCIS).

Floridi, L. (ed.). 2010. *The Cambridge handbook of information and computer ethics*. Cambridge: Cambridge University Press.

Grunwald, A. 2007. Converging technologies: visions, increased contingencies of the conditio humana, and search for orientation. *Futures* 39(4): 380–392.

Himma, K.E., and H.T. Tavani. 2008. *The handbook of information and computer ethics*. Hoboken: Wiley.

Moor, J.H. 1985. What is computer ethics. *Metaphilosophy* 16(4): 266–275.

Owen, R., P. Macnaghten, and J. Stilgoe. 2012. Responsible research and innovation: from science in society to science for society, with society. *Science and Public Policy* 39(6): 751–760.

Owen, R., M. Heintz, and J. Bessant (eds.). 2013. *Responsible innovation*. New York: Wiley.

Roco, M.C., and W.S. Bainbridge. 2007. Converging technologies. In *Nanotechnology: societal implications*, ed. M.C. Roco, and W.S. Bainbridge, 131–168. Dordrecht: Springer.

Stahl, B.C., G. Eden, and M. Jirotka. 2013. Responsible research and innovation in information and communication technology—identifying and engaging with the ethical implications of ICTs. In *Responsible innovation*, ed. R. Owen, M. Heintz, and J. Bessant, 199–218. New York: Wiley.

Strauss, A.C., and J. Corbin. 1997. *Grounded theory in practice*. New York: Sage Publications Inc.

Van den Hoven, J., and J. Weckert (eds.). 2008. *Information technology and moral philosophy*, 1st ed. Cambridge: Cambridge University Press.

Von Schomberg R ed. (2011) Towards responsible research and innovation in the information and communication technologies and security technologies fields, Luxembourg: Publication Office of the European Union. Available at: http://ec.europa.eu/research/science-society/document_library/pdf_06/mep-rapport-2011_en.pdf. Accessed 1 Sept 2014.

Wiener, N. 1954. *The human use of human beings*. New York: Doubleday.

Chapter 7
With Cognitive Enhancement Comes Great Responsibility?

Hannah Maslen, Filippo Santoni de Sio and Nadira Faber

Abstract Although drugs that enhance the cognition of 'healthy' individuals (e.g. methylphenidate and modafinil) have received attention from ethicists and philosophers, little research has focused on the concrete opportunities they present for particular groups in society. Recent policy discussion has gone as far as suggesting there may be a moral obligation for individuals in high-risk professions (e.g. surgeons, pilots) to take enhancers. This chapter outlines a theoretical framework and methodology for investigating the claims that some professionals: (a) might have a responsibility to enhance and (b) might acquire more responsibilities once enhanced. Our methodology is interdisciplinary—as we examine normative hypotheses alongside psychological data and legal precedent—and practice-oriented—as we ultimately aim to make recommendations for policy and the professionals within its remit. Philosophical analysis exposes the conceptual and normative questions involved in a discussion of enhancement in professional contexts, offering and refining definitions of concepts (capacity, responsibility) and theory about their relationship. Psychological inquiry uses surveys and

We thank Nicole Vincent and an anonymous reviewer for comments on earlier drafts of this paper.

H. Maslen (✉)
Uehiro Centre for Practical Ethics, Suite 8, Littlegate House, 16/17 St Ebbe's Street, Oxford OX1 1PT, UK
e-mail: hannah.maslen@philosophy.ox.ac.uk

F. Santoni de Sio · N. Faber
Department Values, Technology, and Innovation, Section of Ethics/Philosophy of Technology, Delft University of Technology, Jaffalaan 5, 2628BX Delft, The Netherlands
e-mail: f.santonidesio@tudelft.nl

N. Faber
Department of Experimental Psychology, University of Oxford, 9 South Parks Road, Oxford, OX1 3UD, UK
e-mail: nadira.faber@psy.ox.ac.uk

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_7

experimental methods to collect data from lay people and professionals on attitudes and responsibility attributions associated with enhancement. Legal analysis examines the conditions under which professional duties to enhance might emerge and how the law might impose or limit liability.

Keywords Cognitive enhancement · Capacity · Modafinil · Reflective equilibrium · Responsibility · Ritalin

7.1 Introduction

Recent research suggests that some medications such as methylphenidate (e.g. Ritalin) and modafinil (e.g. Provigil), which were originally created to treat conditions like attention deficit hyperactivity disorder (ADHD), narcolepsy, shift work sleep disorder and excessive daytime sleepiness might also improve certain aspects of mental performance when taken by healthy individuals. For instance, methylphenidate has been shown to have modest effects on response inhibition, working memory, attention and vigilance; modafinil has been shown to have modest effects on working memory, episodic memory and attention (Husain and Mehta 2011). Given these effects, might there come a day when people are expected to take such drugs?

We think that this is not an altogether unlikely scenario. Consider for instance the following discussion and recommendations published by Queensland Health, the medical regulatory body of the North-East Australian state, in their Fatigue Risk Management System Resource Pack:

> To meet the needs of patients at any time of the day or night, … doctors and other healthcare workers … often work long hours — throughout the night and on-call over weekends, public holidays and other times of need. This presents us with the challenge of fatigue and its associated risks to staff and patients. To meet this challenge, fatigue risk management must be included in our core business operations. (2009:1)

As a possible solution, the Queensland Health report suggests that doctors could take '[n]aps of less than 30 min in length [t]o provide measurable boosts in alertness and performance' (77) and up to '400 mg of caffeine [which is the] equivalent to about five to six cups of coffee' (78) because '[c]ompared with other psychoactive drugs (e.g. modafinil), caffeine is … more readily available and less expensive' (79). However, given the drawbacks of napping [the report cites 'sleep inertia effects' which involve a 'period of disorientation and performance impairment that is experienced immediately upon waking' (77)], that caffeine is not well tolerated by everyone (78), and that the cost and availability of drugs like modafinil could change with changes in regulation, it is perfectly conceivable that a future report may recommend that such drugs be used.

Further, the report from a recent workshop on 'Human Enhancement and the Future of Work', hosted by the Academy of Medical Sciences, the British Academy, the Royal Academy of Engineering and the Royal Society (2012) considers that there may in fact be a moral obligation for some individuals to use enhancers at work:

> [O]ccupations that require particular patterns of focus could benefit from enhancements that facilitate achieving such patterns. For example, surgeons may need to be able to concentrate for extended periods, whereas other jobs such as air traffic control can require very rapid reactions during periods of relative uniformity. As an extrapolation to this, it is possible that in these high-responsibility occupations enhancement could be seen as a moral obligation, or even demanded by the public. Recent examples of traffic accidents involving passenger coaches draw attention to the drivers of these vehicles as another potential target for such demands. Situations like this will require careful consideration [p38]

The above concerns are of great importance to public policy. However, cognitive enhancement drugs also raise important philosophical concerns. In a paper that explores the relationship between different senses of the word 'responsibility', Vincent (2011) proposes that a central underlying assumption of much of our reasoning about responsibility is the capacitarian idea that responsibility tracks mental capacity. To elucidate this co-variation, Vincent provides a range of examples:

> [I]n lay contexts responsibility is often thought to require such things as the ability to perceive the world without delusion, to think clearly and rationally, to guide our actions by the light of our judgments, and to resist acting on mere impulse. This is, for instance, why children, the senile, and the mentally ill are thought to be less than fully responsible for what they do (i.e. because they lack the right kind and/or degree of mental capacity), why children can acquire more and/or greater responsibilities as they grow up (i.e. because their mental capacities develop as they mature), and how responsibility is reinstated on recovery from mental illness (i.e. because the needed mental capacities are recovered).

Elsewhere, Vincent (2013) elaborates on this idea by arguing that if the capacitarian thesis is right, then it seems that as capacities are enhanced beyond the 'normal' range the people in possession of these capacities might in some sense become 'hyper-responsible'. This might mean that they incur *new* or *greater responsibilities*—that their duties multiply and extend (c.f. also Santoni de Sio et al. 2014a, b). It might also mean that these people should be held *more responsible*—are more blameworthy—when things go wrong. If this theory is correct, then there are obvious implications for professionals who use enhancers: once enhanced, more will be expected of them, both in terms of the set of things they are expected to do and the level of proficiency with which they are expected to do them.

The performance-improving effects of cognitive enhancers on professionals engaged in particularly challenging tasks poses an interesting question: does cognitive enhancement enhance responsibility? The question can be split in two parts: (a) might individuals engaged in certain high-risk professions have a responsibility—a duty—to enhance? and (b) are these individuals, once enhanced, more responsible—more accountable—for what they do?

Intuitively, both questions could receive a positive answer. On the one hand, it seems that, for the same reasons for which professionals operating in high-risk situations may be required to use all the technologies available to reduce the probabilities of mistakes, they might (sometimes) also be required to enhance their mental capacities through a pharmacological intervention (provided this had no or negligible negative side-effects on their health). On the other, as we accept that responsibility is diminished when mental capacities are lowered, it seems that we should accept also that responsibility might be increased through mental enhancement.

7.2 Cognitive Enhancement and Responsibility: Goals and Means of the Inquiry

To comprehensively test these two claims in a way that is useful for making recommendations for policy, a wide-ranging assessment of these hypotheses is required. In order to do so, at least three elements seem to be necessary: a deeper understanding of the concept of 'capacity' and of its relationship to responsibility, an understanding of the views of professionals and the general public on the subject, and speculation on the potential legal ramifications of various positions. Thus, to a certain extent, we follow others in taking the 'participatory turn' identified in Chap. 1 of this volume (Koops 2015). However, we emphasize that the importance of listening to stakeholders in innovation processes does not eliminate the need for robust conceptual and normative analysis (cf. Santoni de Sio et al. 2014a, b). Thus, philosophical, psychological and legal modes of inquiry should ultimately coalesce to reach conclusions on the interrelated facets of the debate. On the basis of these results it will be possible to offer responsible recommendations to policy-makers and innovators on this issue.

Answering the normative and conceptual questions about cognitive enhancement and responsibility therefore requires bringing together multiple modes of inquiry. We need to understand the concepts involved and the relationship between them; we need to understand lay and professional beliefs and attitudes; and we need to understand the legal context, which might restrict or lend support to particular approaches. In fact, it is not clear that the philosophical goal of clarifying concepts and relations can be met through philosophical analysis alone: lay intuitions and reflections from professionals about enhancement and responsibility, and legal concepts of duty and negligence can usefully feed into philosophical inquiry. This approach is known as the method of 'wide reflective equilibrium' and is explained in greater detail below.

Thus, to achieve the two primary goals—philosophical understanding and policy recommendation—there is a need to engage in these three strands of research. *Philosophical* analysis exposes the conceptual and normative questions involved in a full discussion of enhancement in professional contexts; it offers definitions of concepts and theories about their relationship, whilst also seeking to refine or review said definitions and theories; it works through the normative implications of the refined theories for professionals in their respective contexts.

Psychological inquiry uses surveys and experimental methods to collect data on attitudes and responsibility attributions from lay people and professionals. These data serve as a source of ideas for normative analysis, are useful in exposing any philosophical bias, and enable the policy recommendations to be made in full awareness of professional and public opinion.

Legal analysis enriches the conceptual analysis through bringing its own definitions of capacities, duties, responsibility and negligence; it analyses the issues that a claim of negligence relating to a failure to enhance would raise for the court; it draws on any analogous cases to try to ascertain what authority or transferable principles, if any, there might be. The next three sections elaborate on each of the three strands in turn. Our approach can be seen to correspond to the 'product approach' characterized by Koops (2015). Accordingly, we demonstrate the way in which our theoretical framework and methodology would promote responsible policy-making in the context of the pharmaceutical cognitive enhancement of various professionals.

7.3 Philosophical Analysis

Philosophy must assess whether the straightforward extension of capacitarianism that Vincent envisages extends to the enhanced range or, alternatively, whether the relationship is complicated by countervailing considerations. In order to do this, the philosophers should employ the method of wide reflective equilibrium. The method of wide reflective equilibrium is often used to find and justify solutions to moral and political puzzles. It is a deliberative process that begins with a priori moral intuitions, principles and theory, and then tests and revises the attendant beliefs alongside each other, and alongside competing theories and empirical facts about the world with the overall goal of reaching coherence (see Rawls 1971). The initial beliefs about an area of inquiry are revised in this manner until the most coherent and credible set is arrived upon.

The method of wide reflective equilibrium may be particularly fruitful when drawing on data from the psychological strand, using the data as a source of ideas, and also as a yardstick against which to test tentative conclusions. In wanting to know whether lay people attribute responsibility in a way consistent with the capacitarian thesis, the philosophers are not necessarily hoping for or expecting consensus. However, great divergence between lay and philosophical theories of responsibility puts an extra burden on the philosophers to explain why their theory is justified. Insights into attributional biases in lay people can help clarify where there is real rather than symptomatic disagreement.

The capacitarian hypothesis is exposed to two very general conceptual objections: firstly, it might be asked whether the mental *modification* produced by drugs like methylphenidate and modafinil should qualify as a mental *enhancement*, i.e. as an *improvement* of the subject's mental conditions. This objection highlights that it may not always be good to enhance memory, attention, or wakefulness, as these modifications may also hinder the pursuit of other valuable things. For instance, the ability to forget can be as important as the ability to remember, and forgetting is often necessary to cope with and go through stressful of painful personal past experiences. Therefore—so the objection goes—enhancement is not an absolute but relative concept since it cannot be applied without reference to a specific context of evaluation. Whilst the general point of this objection can easily be accepted (see Santoni de Sio et al. 2012), this does not seem to be a problem for our specific hypothesis. The relevant assumption here is not that methylphenidate or modafinil constitutes a source of mental enhancement *in general*, but that these drugs may constitute a form of mental enhancement at least *for certain professionals* (surgeons, pilots and soldiers) *when they are engaged in particularly challenging tasks typical for their professions*. This limited assumption seems to be much less controversial and open to criticism.

Suppose, though, that the most that such drugs could be claimed to do, or that the most appropriate way to describe what they would do, is that they would only treat what is ultimately a dysfunction, rather than to raise people's mental function to better-than-normal levels. For instance, that they might temporarily wipe away fatigue-induced mental cobwebs and sluggishness, and return their user to normal rather than genuinely enhanced levels of mental function. It would then seem that this should count as a form of *treatment*, rather than of *enhancement*. Nevertheless, we take it that in an important sense what such drugs would do should still be described as a form of enhancement, since people who used them (or who had access to them and were prepared to use them) could do things that others could not do as effectively—e.g. late-night cramming for an exam the next morning, actually sitting that exam with few or no hours' sleep, or working exceedingly long hours to meet a deadline. Relative to a baseline of normal performance, which takes in to account naturally-occurring performance degradation due to common maladies like tiredness, such individuals would be enhanced.

Another objection to the capacitarian hypothesis derives from the concern for *authenticity*. In a nutshell, the argument is that the cognitive capacities enabled by a particular drug are not truly the capacities of the enhanced person. If this position is taken seriously, then it could be argued that—*contra* the hypothesis—the enhanced person is actually *less* responsible for the things that he does because, post-enhancement, his behaviour is not (completely) attributable to him, but to an inauthentic self. Whilst this is an objection that needs careful consideration, it must be noted that it is not in itself a fatal blow to the extended capacitarian thesis. The objection is seated in a concern about unnatural, external means of becoming enhanced, which do not exhaust all the possible means of enhancing capacities. The kind of circumscribed mental modification discussed here should not therefore raise particular concerns in this perspective.

More generally, the worry about authenticity and responsibility seems to derive from a naïve view of the nature of responsibility and its relation to the self (see Santoni de Sio et al. in press). According to this view, the reason why people persist over time despite continually undergoing changes is because something about them—some essential core—remains unchanged through different times and situations. On this view, responsibility for actions is legitimately attached to people only when their actions stem from this unchanging part of the self. In this perspective, all relevant modifications of the self are a potential threat to the subject's responsibility. Discussing this view in detail would require more space than this forum permits, but suffice it to say that in our view it is more plausible to accept that humans change—often quite substantially—and to accommodate responsibility within this changing view of the self rather than to insist that responsibility requires an absence of change. The burden of the proof, it seems to us, stands on the side of those claiming that a mental modification *deriving from these drugs* constitutes a responsibility-undermining threat to authenticity. Simply pointing at the presence of a mental modification is not enough.

A further element of complication for the capacitarian approach derives from the ineliminable presence of normative elements in the capacity attribution. On the one hand, it seems that even accepting the idea of there being an important relationship between capacity and responsibility, this idea cannot not be read in the simple sense of there being a *direct ratio* between the two, such that at each modification in the capacity of the subject corresponds an equal modification of his responsibility. That this is not the case can be realized by looking at the *diminished* capacity side, for example in the legal working of the insanity defence. Put very roughly, people that fall above *a certain threshold* of capacity are *ceteris paribus fully* legally responsible for their deeds, no matter *how much* they are intelligent, rational, sensitive and so forth. So something similar might be true on the *enhancement* side. It could therefore be the case that the kind and/or quantity of mental modification allowed by cognitive enhancement turns out to be insufficient to modify the capacities in a way that is *relevant from the point of view of moral and legal responsibility*. This is another important specification that does not affect the validity of the capacitarian approach in itself. It only invites us to characterize it in a more precise way, in order to be able to apply it correctly to the present case (Vincent 2013).

A final challenge to the capacitarian approach derives from problems related to the very notion of 'capacity'. According to a certain view of capacity (Santoni de Sio and Jespersen 2013) roles are decisive not only for *setting the thresholds of morally or legally relevant capacities*, but for *defining* and *attributing* capacities in the first place. On this view, knowing the level of *performance* expected from a certain subject given his past performances or his material and psychological conditions in a given circumstance is not sufficient to determine his *capacities* at that moment. Being able to predict an agent's particular *behaviour* is something different from evaluating their *capacities*. To determine, for example, whether a certain person has the capacity to run 1 km in less than 5 min or to calculate the square root of 121 in a given circumstance, one must know what *kind* of runner or mathematician she is in the first place. Then one should determine the expected performance—in a *normatively-laden sense*—by people occupying *that role in those circumstances*. Only at this point will it be possible to determine whether a particular performance is one to be expected by that individual in those circumstances.

From this perspective, the general capacities of an agent are defined through the reasonable *normative expectations* entailed in the roles that she occupies in a given circumstance. And the (legitimate) attribution of a certain role does not depend only on past performances and material conditions of the actual performance, but also on facts like the fulfilment of certain formal requirements (the existence of various social structures) and procedures (the recognition of one's position within those structures). Although this account of 'capacity' does not necessarily *undermine* the capacitarian idea that responsibility tracks capacity, it does nevertheless *complicate* it, since it posits a two-way relationship between capacity and responsibility. Not only does the scope and content of a given person's responsibilities hinge on their possession of particular capacities (defined according to a purely naturalistic criterion), but their possession of those capacities at a certain time is itself something that depends on what roles and responsibilities they have.

7.4 Psychological Inquiry

Psychology should use empirical methods to collect data on the attitudes and responsibility attributions of groups of two types: *valorisation groups* of surgeons, pilots and military personnel—professionals who might one day be expected to take cognitive enhancers—and the *general public*. The data gathered from both types of group will capture opinions on the responsibilities there might be for certain individuals *to* enhance and the degree to which enhanced individuals are responsible *for* their acts and omissions.

In relation to the valorisation groups, surveys are used to directly investigate the views of the professionals on the subject of enhancement within their respective professions. Samples of people from different professions (e.g., soldiers, surgeons, pilots) and different countries (e.g., Netherlands, UK, Australia) are asked for their opinions on the responsibility *to* enhance in their profession (first wave of surveys) and *if* enhanced (second wave of surveys). These surveys will invoke different senses of responsibility: for example, soldiers are asked about the extent to which they agree that 'responsible soldiers would take substances to improve their performance' ('virtue responsibility'), and extent to which they agree that 'soldiers who do not take substances that improve their performance are blameworthy' ('outcome responsibility/blame'). The data collected from these surveys has two main uses: (1) it provides ideas for normative analysis—views and opinions on responsibility may emerge that the philosophers had not yet considered (these views feed into the reflective equilibrium) and (2) it aids in the overall development of policy recommendations—it is crucial to understand the concrete contexts and the views of those within them to make *reasonable* and *efficacious* policy recommendations for these contexts.

Extending our data collection beyond members of the valorisation groups to assess the opinion of the general public, surveys and experiments are used to explore lay reasoning on both the responsibility *to* enhance and *if* enhanced. The same method as for the valorisation groups is employed: in surveys, participants are asked about their opinion on whether people from different professions (e.g., soldiers, surgeons, pilots) should enhance themselves (responsibility *to* enhance) and how they attribute responsibility *if* enhanced. Again, different senses of responsibility are covered. This way, the opinion of the valorisation groups can be compared to the opinion of lay people, allowing for a systematic comparison. Additionally, behavioural experiments are employed. Experiments involve manipulating one variable to determine if changes in this variable cause changes in another variable. This way, causal conclusions can be drawn. Hence, attributions of responsibility and the factors that influence these attributions can be assessed in detail. Participants' reasoning is addressed indirectly by presenting them with scenarios to assess. Different factors that might influence the attribution of responsibly in relation to enhancement are varied in these scenarios (e.g. outcome of action, how much is at stake). For example, a basic scenario describing a surgeon who has the opportunity to take modafinil before a challenging operation might vary whether participants are asked to imagine this surgeon to be themselves versus another person. This might alter the participants' opinion on whether they think the surgeon should take the enhancer.

This assessment of the reasoning of the 'average person' serves two functions: (1) it serves as a yardstick against which to test preliminary normative conclusions: if the philosophers posit responsibility concepts and a theory about their relationship to enhancement that turn out to *diverge* from lay intuition, then this divergence must be explored and explained: either the normative concepts and theory have to be revised or the lay intuitions have to be shown to be compatible through exposing biases or systematic differences in conceptual definitions. (2) The assessment of the 'average person' allows insight into the attributional biases that may be colouring the lay attributions, for example differences in judging your own versus the behaviour of another person, i.e., actor-observer asymmetries (e.g. Malle et al. 2007). These insights into such biases (cf. Caviola et al. 2014) are important for (a) providing a possible explanation for any divergence between the normative conclusions and lay attributions (hence removing the divergence as a challenge) and (b) predicting public responses to the policy approach normative conclusions suggest: the insights provide information that will help policymakers present and explain potentially controversial enhancement policy to the general public, especially if the 'average person' opinions are different from those of the members of the valorisation groups.

Whilst understanding the particular professional contexts involves asking the right people within that profession—people who will be best positioned to provide a useful perspective on what the implications of enhancement would be for surgery or aviation or military operations etc., understanding lay attributions of responsibility requires gathering data from as large and as representative a sample of the general population as possible. The data gathered from both types of group is crucial to the overall aims of reaching a coherent normative position on the concepts and conceptual relationships involved in cognitive enhancement debate, to situating this position constructively within the professional contexts where the debate is most relevant, and to generating policy recommendations for these contexts; policy which must be justified to the population at large.

7.4.1 Preliminary Findings

We gathered preliminary data from lay people that can shed light both on their opinions regarding a responsibility *to* enhance and their attributions regarding the responsibility *if* enhanced.

To obtain an initial impression of the attitudes lay people have about the use of cognitive enhancement substances by professionals in certain roles we asked the following question on a paper-and-pencil questionnaire: ‘Do you agree with the following statement? People in professions that affect the life and death of other people (e.g. surgeons) have a moral obligation to take substances that enable them to perform as well as possible.’ The answers were to be given on a 7-level rating scale ranging from 0 (‘completely disagree’) to 6 (‘completely agree’). 80 participants (81 % female, 19 % male; mean age 22; mostly University students) took part in that survey in a controlled laboratory setting. Results show that participants clearly reject the claim that there exists a moral duty to enhance in this context: 44 % responded ‘completely disagree’ and 18 % ‘disagree’ to the above statement. Only one participant agreed, and nobody agreed completely. Correspondingly, the overall mean value of all responses is 1.50 (with a standard deviation of 1.7), translating to ‘between completely disagree and disagree’. Results did not depend on participants’ demographical characteristics like sex and age. Hence, the lay intuition in our sample is clearly that there is no moral duty to take enhancement substances for certain professionals. This result is perhaps not surprising given the strong objection the general public has against the use of cognitive enhancement substances (e.g. Bell et al. 2013; for a review, see Schelle et al. 2014).

Obviously, in keeping with the principles motivating responsible innovation, research in this area must be sensitive to the possibility of attitude change, especially as the technology develops and public knowledge of its benefits and risks accumulates. Especially where psychoactive substances are concerned, we can usefully reflect on changes in attitudes over time. For example, attitudes towards caffeine have changed significantly: whilst Sweden now has one of the highest per capita rates of coffee consumption, in the 18th Century coffee consumption was disapproved of and even prohibited (Weinberg and Bealer 2001). At the present time, research tells us that people do not think about cognitive enhancers as comparable to caffeine. For example, although people take cognitive enhancers for reasons similar to those motivating ingestion of caffeine—often in pursuit of the psychoactive effects—people are disinclined to see them as analogous substances (Forlini and Racine 2012). This could be due to a lack of familiarity with the new cognitive enhancers and suspicion surrounding their effects and side-effects. Further, whilst drinking coffee has become a social activity, this is not the case for taking enhancers like methylphenidate and modafinil, the use of which might even bare some social costs for the few who take them (Faulmüller et al. 2013). Of course, familiarity with and attitudes towards enhancers may change over time, as might opinion on the appropriateness of their use in professional contexts. To address the possibility of changing attitudes and values, the ongoing psychological research has to take account of factors that have been shown to determine how acceptable people find the use of cognitive enhancers (e.g. negative side-effects or fairness; Faber et al. 2015; Scheske and Schnall 2012), and to investigate whether attributions of responsibility change when participants are asked to envisage changes to these factors (e.g. availability of cognitive enhancers without considerable side-effects).

To obtain preliminary data on attributions related to the responsibility *if* enhanced, we conducted a first experiment with the student sample described above. In that experiment, participants were completing a performance task. Before they started the task, they were given a glass of juice to drink. Participants were told that this juice might contain certain enhancing substances like vitamins or caffeine. After the completion of the performance task, they were asked to rate their agreement to the following statements on a 7-level scale from 0 ('completely disagree') to 6 ('completely agree'): 'I am responsible for how I performed in the task' and 'The substances contained in the juice are responsible for how I performed in the task'. Overall, participants attributed significantly more responsibility for their performance to themselves then to the potential enhancer: The mean value for themselves is 4.8 (with a standard deviation of 1.1), which translates to 'agree' (with a tendency to 'somewhat agree'). The mean value for the substances is 1.0 (with a standard deviation of 1.0), which means 'disagree'. (Statistical significance of this difference between self versus substances was proven by a t-test, $t(79) = 20.33$, $p < 0.001$.) Interestingly, this pattern does not depend on participants' subjective performance: whether they thought they had done well or badly in the task had no influence on their attributing responsibility for the result to themselves, rather than to the substances they took. This suggests that they did not use the potential cognitive enhancer as an 'excuse' for bad performance.

Responsibility attributions for self versus substances are significantly negatively correlated ($r = -2.2$, $p = 0.027$): the more responsibility they ascribed to themselves, the less they attributed to the potential enhancer (and vice versa). These results were not affected by participants' age, sex, or what the juice they drank actually contained.

In sum, these results imply that in this context people seem to see themselves as responsible for the results of their actions, rather than a cognitive enhancer, and that this is independent of the subjective quality of this result. Moreover, there might be a tendency to distribute a certain fixed amount of subjective responsibility: the more people give themselves, the less an enhancer gets. However, there are reasons why these results should not be taken as more than preliminary indications about lay reasoning. First, this responsibility 'sub-experiment' was part of a larger experiment. Although statistical analyses imply they were not, we cannot rule out that our participants have been affected by the other treatments, e.g. in their ability to concentrate. More importantly, due to the experimental set-up participants were not sure which substances were actually contained in the juice they drank. Although what participants believed they took did not influence the results reported above, results might be different for people who take a substance which they are convinced has an effect on their cognitive performance.

In sum, our first tentative empirical investigations imply that lay people seem not to see an obvious positive relationship between cognitive enhancement substances and responsibility: they do not think that people in certain professions have a moral duty to take such substances, nor do they attribute responsibility for their own performance to an ingested enhancer. Whether these tendencies remain stable, our further empirical investigations have to prove.

7.5 Legal Analysis

The legal analysis of the problem involves a degree of speculation: there are no recorded legal cases involving cognitive enhancers and the idea that professionals such as surgeons and pilots might someday have a legal obligation to take enhancers has been put forward only very recently and very tentatively. However, it is conceivable that, if cognitive enhancers were proven to be safe and effective for surgeons, pilots and other professionals for whom the costs of error are high, the general public may begin to expect these professionals to take them under some circumstances (for an extended discussion on the relation of safety, effectiveness and ethical debate, see Maslen et al. 2014). Where this expectation emerges, so too does the possibility of negligence claims in the event of a failure to enhance. Given this possibility, and the speed at which new pharmaceuticals are being developed and made available, preparatory analysis is needed to identify and examine the legal issues such claims would raise. This analysis is not only important for legal practice, but also for the professionals who could be affected by this practice.

To do this, the lawyers are examining the tort of negligence to consider how a court would assess a claim involving harm allegedly resulting from a professional's failure to take an enhancer. This involves looking at when the law ever imposes obligations on individuals to *do* certain things (i.e. makes them liable for *omissions*), how the standard of care is determined in different professions (and how the attendant expectations evolve), and how causation could be established in a case involving enhancer omission. The lawyers are also looking at any guidance from relevant councils—medical, aviation etc.—that could be used to support or challenge a claim that a professional was negligent in not taking an enhancer. Such guidance often constitutes 'soft law': not following it is not illegal per se but may have disciplinary consequences and can be used as evidence when establishing negligence in court. Finally, cases involving analogous features—for example, where an individual is required to remedy a deficiency in a capacity—are compared in an attempt to extrapolate key principles. Driving with certain medical conditions can generate such requirements.

7.5.1 Preliminary Findings

The legal research so far has focused on surgeons and their potential duties in relation to enhancement. In particular, it has focused on the modest claim that there could be some limited circumstances—where the surgeon is very fatigued and the only person available to perform the surgery—in which it might be thought that the surgeon has an obligation to take a cognitive enhancer (see Goold and Maslen 2014b). In general, English law is very reluctant to impose omissions liability. For example, it does not impose a duty of easy rescue: absent an established duty of care, a person walking past a drowning child is under no obligation to save the child, even if it would be very easy for the person to do so.[1] However, where there is a clear duty of care—as found between a surgeon and her patient—the individual owing the duty of care will be obligated to *do* certain things.

Negligence liability depends on five things being established: (1) that the defendant owed the claimant a duty of care, (2) that the defendant breached this duty of care, (3) that the claimant suffered some harm, (4) that this harm was caused by the defendant's breach, and (5) that the harm was not too remote. Of particular importance for our project is to ask whether a surgeon might ever be in breach of his *duty* by not taking a cognitive enhancer (2) and how it could be established that not taking an enhancer *caused* harm to the patient (4).[2]

In England and Wales, a surgeon breaches his duty if his acts or omissions fall below the minimum standard of care. In cases of alleged clinical negligence, the minimum standard of care is determined by what has become known as the 'Bolam test', arising from the judgment of McNair J in *Bolam v Friern Hospital Management Committee.*[3] According to the test, a defendant 'is not guilty of negligence if he has acted in accordance with a practice accepted as proper by a responsible body of medical [persons] skilled in that particular art'.[4] In the case of our fatigued surgeon, the question for the court would be whether it is standard practice to use enhancers in emergency situations such as when the surgeon is particularly fatigued but the only person able to perform the surgery. The answer at the present moment would obviously be that it is not: the use of cognitive enhancers is not mainstream practice and a body of medical persons who would testify to this could easily be found.

However, what is accepted as proper practice changes over time as new techniques and procedures are introduced. Further, subsequent to the decision in *Bolam* the court now has the authority to independently assess the *reasonableness*

---

[1]Relatedly, see *Barrett v Ministry of Defence* [1995] 3 All ER 87 (CA); *Sutradhar v Natural Environment Research Council* [2006] UKHL 33; [2006] 4 All ER 490 (HL).

[2]The issues of duty and of causation are explored in depth in Goold and Maslen (2014b) and Goold and Maslen (2014a), respectively.

[3]*Bolam v Friern Hospital Management Committee* [1957] 1 W.L.R. 582.

[4]Ibid, at 587.

of the way in which the surgeon acted. In *Bolitho v City and Hackney Health Authority,*[5] the Court held that in determining the standard of care, that the Court must be '… satisfied that, in forming their views, the experts have directed their minds to the questions of comparative risks and benefits and have reached a defensible conclusion on the matter'. This supplements the assessment of prevalence of behaviour with an assessment of its reasonableness. Thus, even if there existed a body of medical people who stated that they would not use cognitive enhancers, if it became apparent that cognitive enhancers cheaply and effectively reduced risk, a judge would *be able* to (although would not necessarily) find the expert's testimony indefensible. However, at the present time, concerns about the risks posed to the surgeon in requiring her to take medication for unapproved purposes, and the general reluctance of the law to intervene on people's bodies make a legal duty to enhance very unlikely.

Despite the current conclusions it should be noted that professional requirements—along with public attitudes, noted above—do change over time, especially as new technologies emerge. Whilst courts are currently unlikely to invoke *Bolitho* to establish a duty for surgeons to enhance, there have been instances of technological change that did result in new duties. For example, American case law includes a couple of 'landmark' cases where medical professional defendants were held liable for their failure to adopt new technologies or procedures, even when near universal custom did not involve using them.[6] In one case (*Helling v Carey*[7]), an ophthalmologist was held negligent for failing to perform a simple pressure test for glaucoma, which the claimant had developed. This judgment was reached despite expert testimony attesting to the fact that the standard practice in ophthalmology was to not require glaucoma tests for patients under the age of forty. In another case, (*Washington v Washington Hospital Center*[8]) a hospital was held liable for failing to use an oximetry monitor which, it was argued, would have provided early detection of the oxygen deprivation which ultimately lead to the claimant's brain damage. This was despite expert testimony claiming that the use of these monitors was not yet widespread and not mandated. Thus, if empirical evidence repeatedly demonstrated the safety and effectiveness of enhancers, views on whether their use should be adopted as a simple precaution may change.

However, it must be remembered that medical practice is not only shaped by peer consensus on proper practice and legal assessments of corresponding negligence. Surgeons should also act in accordance with guidelines such as *Good Surgical Practice* (2008), published by The Royal College of Surgeons. Whilst not legally binding, deviation from these guidelines can at least be used as an argument that some particular way of proceeding was negligent. If cognitive enhancers

[5]*Bolitho v City and Hackney Health Authority* [1998] AC 232.

[6]The following two cases are cited in Greenburg (2009).

[7]*Helling v Carey* (519 P.2d 981 [Wash. 1974]).

[8]*Washington v. Washington Hospital Center*, 579 A.2d 177 (D.C. Cir 1990).

were ever to become a legal requirement for surgeons under some circumstances, it is more likely that this would be as a consequence of explicit statements in guidelines rather a revolutionary court decision.

If there ever were to be a duty for surgeons to take enhancers, in cases where a breach of this duty was established, the court would also have establish that, it was this breach that *caused* the claimant's harm (see Goold and Maslen 2014a). Despite the 'balance of probabilities' standard of proof, establishing causation can be notoriously difficult, particularly when there are multiple competing causes. The standard test of causation—known as the 'but for' test—would involve an assessment of the likelihood that if the surgeon had taken the enhancer, the harm would not have occurred. If it is more likely than not that taking the enhancer would have prevented the harm, then the surgeon may be held liable. However, the effects of not taking an enhancer—essentially, allowing fatigue to persist—are not easy to identify. Whereas, for example, not giving a patient enough oxygen has clear and measurable effects, the specific effects on the patient of a surgeon operating whilst fatigued are indeterminate. Further, there are always inherent risks involved in surgery, which are often blameless if they materialize. It would be very difficult to know whether or not some instances of harm are risks that materialized blamelessly or were consequences of the fatigue that the surgeon omitted to remedy. Much will depend on the precise details of the case: the particular procedure that took place, the level of cognitive and manual dexterity involved, the risks inherent to the procedure, the probability that these risks materialize, and so on.

The legal analysis has compared hypothetical 'fatigued surgeon' cases with various landmark cases in negligence law to draw out similarities and differences in causal structures. This analysis has identified the causal structures most likely to be in operation in potential fatigued surgeon cases. However, it remains questionable whether the court could sufficiently overcome the uncertainty arising from the indeterminacy of the effects of fatigue, and the fact that some harms will have been risks identified as inherent to the surgical procedure.

7.6 Conclusions and Future Research

As has emerged, the philosophical thesis of capacitarianism faces some *prima facie* challenges generated by the psychological and legal findings. The preliminary psychological data suggests that, *at present*, lay people do not seem to see an obvious positive relationship between cognitive enhancement substances and responsibility: they do not think that people in certain professions have a moral duty to take such substances, nor do they attribute responsibility for their own performance to an ingested enhancer. However, as emphasized, the method of reflective equilibrium *anticipates* conceptual and normative disagreement, aiming to reach coherence once all the normative commitments and empirical facts have been considered. Further psychological research will help confirm whether the survey and experimental instruments indeed tap into the postulated concepts and whether any attributional biases are in operation. In the event that the intuitions of lay people do substantively diverge from the tenets of capacitarianism, the philosophers will be able to explore why and reflect on the consequences of this for their theory.

The legal analysis has so far suggested that intuitions about a moral obligation for some professionals to take enhancers may not translate organically into a legal duty in the foreseeable future. Negligence law in England and Wales still tends to use an evaluation of consensus to determine the duties of the clinical practitioner and, unless a decision to issue explicit guidance were made, a potential claimant would be hard-pressed to provide evidence that a surgeon had breached her duty of care by not taking an enhancer. Even if a duty were to be identified, the difficulty involved in proving that the failure to take the enhancer caused the particular harm (even on the balance of probabilities) would generate a significant challenge to the imposition of liability.

However, even if professionals like surgeons were never to be held liable for not taking enhancers, this does not mean that the capacitarian thesis must be abandoned. Capacitarianism principally makes claims about the responsibilities of people *once* enhanced. It also does not speak against the possibility of a *moral* duty to take enhancers if they are proven to be safe and effective. We are not legally required to do all that is morally required of us. Further, the evidential difficulties that would accompany attempts to establish causation might preclude holding people liable even if changes to common practice were one day to weigh in favour of a clinical duty to enhance. Philosophical theses about the relationship between capacities and responsibility, and about what we are morally required to do—whilst benefiting from knowledge of empirical realities, such as what cognitive enhancers do—do not have to make pragmatic decisions about liability based on limited information.

Where emerging technologies are seen to promise great social benefits, research is needed at the early stages of innovation to assess risks and opportunities (current and future), and to identify the likely social and ethical implications of these risks and opportunities. Ideally, we have argued, such research should engage with groups in society who could be affected by or benefit from the particular emerging technology. In the spirit of responsible innovation, research into the effects and implications of these technologies should be sensitive to the evolving landscape of public attitudes and professional duties, whilst providing detailed analysis of the current state of affairs for immediate policymaking.

Cognitive enhancement drugs such as methylphenidate and modafinil have received attention from ethicists and philosophers but, to date, little research has focused on the concrete opportunities cognitive enhancement presents for particular groups in society. Our research has begun to redress this by examining normative hypotheses alongside psychological data and legal precedent, ultimately aiming to make recommendations for policy and the professionals within its remit.

References

Bell, S., B. Partridge, J. Lucke, and W. Hall. 2013. Australian university students' attitudes towards the acceptability and regulation of pharmaceuticals to improve academic performance. *Neuroethics* 6: 197–205.

Caviola, L., A. Mannino, J. Savulescu, and N. Faulmüller. 2014. Cognitive biases can affect moral intuitions about cognitive enhancement. *Frontiers in Systems Neuroscience* 8: 195.

Faber, N.S., T. Douglas, F. Heise, and M. Hewstone. 2015. Cognitive Enhancement and Motivation Enhancement - An Empirical Comparison of Intuitive Judgments. *AJOB Neuroscience* 13: 18–20.

Faulmüller, N., H. Maslen, and F. Santoni de Sio. 2013. The indirect psychological costs of cognitive enhancement. *The American Journal of Bioethics* 13: 45–47.

Forlini, C., and C. Racine. 2012. Stakeholder perspectives and reactions to "academic" cognitive enhancement: Unsuspected meaning of ambivalence and analogies. *Public Understanding of Science* 21: 606–625.

Goold, I., and H. Maslen. 2014a. Obliging surgeons to enhance: Negligence liability for uncorrected fatigue and the problem of causation. *Medical Law Review* Online First.

Goold, I., and H. Maslen. 2014b. Must the surgeon take the pill? Negligence duty in the context of cognitive enhancement. *The Modern Law Review* 77(1): 60–86.

Greenburg, M.D. 2009. Medical malpractice and new devices: Defining an elusive standard of care. *Health Matrix* 19: 423–445.

Husain, M., and M.A. Mehta. 2011. Cognitive enhancement by drugs in health and disease. *Trends in Cognitive Sciences* 15(1): 28–36.

Koops, B.J. 2015. The concepts, approaches, and applications of responsible innovation; An introduction. In *Responsible innovation*, vol. 2, ed. B.J. Koops, et al. Dordrecht: Springer.

Malle, B.F., J. Knobe, and S. Nelson. 2007. Actor-observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology* 93: 491–514.

Maslen, H., N. Faulmüller, and J. Savulescu. 2014. Pharmacological cognitive enhancement—How neuroscientific research could advance ethical debate. *Frontiers in Systems Neuroscience* 8: 107.

Queensland Government. 2009. *Fatigue Risk Management System Resource Pack*, Queensland Health.

Rawls, J. 1971. *A Theory of Justice*, 2nd ed. Cambridge: Harvard University Press.

Santoni de Sio, F., and B. Jespersen. 2013. Function, roles, and human capacity. *Methode: Analytic Perspectives* 2(2): 58–66.

Santoni de Sio, F., H. Maslen, and N. Faulmüller. 2012. The necessity of objective standards for moral enhancement. *AJOB Neuroscience* 3: 15–16.

Santoni de Sio, F., N. Faulmüller, and N.A. Vincent. 2014a. How cognitive enhancement can change our duties. *Frontiers in Systems Neuroscience* 8: 131.

Santoni de Sio, F., P. Robichaud, and N.A. Vincent. 2014b. Who should enhance? Conceptual and normative dimensions of cognitive enhancement. *Humana Mente: Journal of Philosophical Studies* 26: 179–197.

Santoni de Sio, F., Faulmüller, N., Savulescu, J., and N.A. Vincent. (in press). Why less praise for enhanced performance? Moving beyond responsibility-shifting, authenticity, and cheating, towards a nature-of-activity approach. In *Cognitive Enhancement: Ethical and Policy Implications in International* Perspectives, eds. F. Jotterand, V. Dublievié, Oxford: Oxford University Press.

Schelle, K.J., N. Faulmüller, L. Caviola, and M. Hewstone. 2014. Attitudes towards pharmacological cognitive enhancement—A review. *Frontiers in Systems Neuroscience* 8: 53.

Scheske, C., and S. Schnall. 2012. The ethics of 'smart drugs': Moral judgments about healthy people's use of cognitive-enhancing drugs. *Basic and Applied Social Psychology* 34: 508–515.

The Academy of Medical Sciences. 2012. *Human Enhancement and the Future of Work*.

The Royal College of Surgeons. 2008. *Good Surgical Practice*, RCSENG—Professional Standards and Regulation.

Vincent, N.A. 2011. Capacitarianism, responsibility and restored mental capacities. In *Technologies on the stand: Legal and ethical questions in neuroscience and robotics,* eds. B. van den Berg, L. Klaming, Netherlands: Wolf Legal Publishers.

Vincent, N.A. 2013. Enhancing Responsibility. In *Neuroscience and legal responsibility*, ed. N.A. Vincent, New York: Oxford University Press.

Weinberg, B.A., and K.B. Bonnie. 2001. *The world of caffeine: The science and culture of the world's most popular drug*. New York: Routledge.

Chapter 8
Neuroimaging and Personalized Learning: Value Reflection with Societal Stakeholders

Rosanne Edelenbosch, Frank Kupper and Jacqueline Broerse

Abstract The emerging technology of neuroimaging may contribute to personalized learning, the adaptation of teaching methods to individual learning needs. In order to proceed with this application in a socially responsible way, it is necessary to carefully consider the practice of education during the innovation process. In this chapter we discuss the results of focus groups in which we reflected on the opportunities and concerns regarding this application with a selection of societal stakeholders: three focus groups with randomly selected parents of one or more children attending secondary school, three focus groups with randomly selected secondary school teachers and four focus groups with secondary school children attending one particular school. Our analysis shows that a different framing of 'the learning child' and 'neuroimaging' can lead to a different attitude towards the application of neuroimaging for personalized learning. It is important to anticipate these different framings in subsequent structuring of science-society dialogue.

Keywords Responsible research and innovation (RRI) · Neuroimaging · Personalized learning · Focus group methodology · Pragmatic ethics

8.1 Introduction

8.1.1 Background and Research Questions

The field of neuroscience has developed rapidly in the past decades, particularly with regard to neuroimaging. This emerging technology has given rise to potential applications in many different domains, including the domain of education.

R. Edelenbosch (✉) · F. Kupper · J. Broerse
Athena Institute, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, The Netherlands
e-mail: r.m.edelenbosch@vu.nl

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_8

Educational neuroscience or ‘neuroeducation’ offers many possibilities for improved learning, but also meets several concerns, in relation to for example commercial promises (Hardiman et al. 2011) the responsibility of teachers to be critical consumers of neuroscience (Hardiman et al. 2011; Maxwell and Racine 2012), and the sensitivity of neuroimaging data (Heinrichs 2011). Ethical issues have been discussed extensively, for example in a special issue of Neuroethics (2012) focusing specifically on neuroeducation.

An important emerging feature of Responsible Research and Innovation (RRI) is moving beyond conventional ethical review and approval of research, to the institutionalization of research that is inclusive, involving stakeholders at an early stage, and value-sensitive, incorporating multiple values and perspectives (Owen et al. 2012). Researchers within the field of neuroeducation are becoming increasingly aware that this calls for new approaches to ethics, but the facilitation of such an approach can be considered a challenge. First of all, there are many publics and many values that need to be considered. Education in itself is an enterprise with its values in “perpetual dispute” (Sheridan et al. 2006). The field of neuroeducation involves even more different stakeholders, with their own deeply rooted values and perspectives. These include, but are not limited to, scientists from different disciplines, policy makers, educational developers, teachers, students and their parents. As shown by Bohman (1996) it is difficult for different stakeholders to come to a consensus because the values that underlie ethical discussion cannot simply be negotiated or traded off, as that would compromise the stakeholders’ integrity. Second, the ethical complexity regarding the application of neuroimaging to education is enhanced by the dynamic nature of the application. As with many emerging technologies, neuroimaging is subject to an uncertain future: not only do we not know what potential applications the technology will give rise to, we also do not know the social embedding of these applications (Rip 2012). Due to the contingency of both the technology and the application itself, the issue cannot clearly be defined and its context is subject to change. These dynamics, together with the value plurality involved, can make it difficult to reach ethical consensus.

As shown by for example Kupper and de Cock Buning (2010) technological controversies characterized by value pluralism and dynamics can benefit from a pragmatic approach to ethics, as monistic approaches fail to grasp the complexity and diversity of real morally problematic situations. The authors argue that the intrinsic nature of the value underlying a moral claim does not determine the ‘‘rightness’’ of that claim. More important is the extent to which it contributes to the resolution of specific morally problematic situations. According to Keulartz et al. (2004) “the moral core of pragmatism revolves around the possibilities for living and working together” (p. 17). In order to further this goal of cooperation, the process of moral inquiry and deliberation is very important. It should allow for the contribution of all those that have a stake in the matter and decisions to be made on the basis of a careful consideration of all relevant conflicting moral claims (Keulartz et al. 2004).

With the research project Neurosciences in Dialogue, the Athena Institute of the VU University Amsterdam aims to contribute to RRI within neuroscience research. In Chap. 4 of this volume, De Jong et al. have described one part of this project, providing insight into how neuroscientists negotiate the meaning of Responsible Innovation in formal and informal contexts. In this chapter we discuss another component of the project, focusing on the *process* of RRI, in which we ourselves have taken an active role. We take a pragmatic approach to the ethical issues emerging from the field of neuro-education, working step-by-step towards the facilitation of dialogue between stakeholders from science and society. In order to facilitate such a process, the following three questions need to be considered: who should be included in such a deliberation, how should this process be designed so that it allows all stakeholders to contribute and how should conflicting moral claims be approached?

The first question is who should be involved in such an ethical deliberation. The role that future users can and should play in innovation processes has been receiving more attention in the past years (see for example Oudshoorn and Pinch 2005; Rohracher 2003). Racine (2008) argues that within a pragmatic ethical approach to health care, “nurses, social workers and all those who can broaden the ethical perspectives” should be included. Besides scientists and educators, we would stress the importance of including the end-users of this application in this process: the teachers, students and their parents. We agree with Howard-Jones and Fenton (2011) that lay people should be included in ethical deliberation so that researchers become aware of public hopes and concerns and do not leave behind human values. Furthermore, we would add that the end-users of neuroimaging have their own knowledge and experience to bring to the table. Not only does their involvement contribute to the fairness of the process of ethical enquiry, they can also improve the products of deliberation by adding knowledge of the contextual aspects of educational practice (Wynne 1996).

The second question posed relates to how the process can be designed in a way that does justice to different types of knowledge and the contextual interpretations of societal stakeholders. We have decided on an approach in which societal stakeholders reflect separately on neuroscientists' ideas about future applications of neuroimaging in education. As a starting point, we have followed the example of Roelofsen et al. (2010), who engaged societal stakeholders to explore the societal future of ecogenomics. Roelofsen et al. (2010) started out with the identification of a scientific field's guiding visions, images of the future that scientists consider to be desirable. This approach starts from the assumption that guiding visions guide technology development because they influence the research choices that are made (Grin and Grunwald 2000). At the same time, they provide a foothold for societal stakeholders' deliberation. By allowing societal stakeholders to reflect on these visions, more insight can be gained into the issues underlying conflicting moral claims.

The third question is how to carefully consider the different moral arguments that are made by stakeholders. For this, we make use of Schön and Rein's (1995) idea of frame reflection and Kupper et al.'s (2007) interpretation thereof. According to Schön and Rein (1995), difficulties in handling controversy often have less to do with how the problem should be solved and more with how the problem is set. They argue that different positions regarding an issue arise from differences in underlying structures of belief, perception and appreciation, or "frames". Kupper et al. (2007) define frames as "the perspectives that actors use to make sense of a complex reality and that guide their actions". When deliberating about complex or indeterminate problematic situations, stakeholders make a selection of features and relations that are considered important in relation to their own frame (Schön and Rein 1995).

In this chapter we describe the process and results of a deliberation with end-users, societal stakeholders involved in the daily practice of education, on the potential application of neuroimaging for personalized learning in the practice of education. We started with the collection of scientists' guiding visions for this field, followed by a reflection on these visions with end-users. During this process, we tried to gain more insight into the values at stake by disentangling different framings regarding the use of neuroimaging for education. When this application is described as a morally problematic situation, which elements are deemed important and how are these elements interpreted? How does this influence the way in which end-users think neuroimaging should be used in education? By answering these questions, we prepare for a more productive dialogue, necessary for inclusive development and socially responsible implementation of neuroimaging in education.

8.1.2 Guiding Visions: Neuroimaging and Personalized Learning

In order to determine what future visions to reflect on with societal stakeholders, we organized two discussion sessions with neuro-education experts from various Dutch universities. During these sessions, the experts discussed which future applications of neuroimaging in education they would find desirable. The visions described by the scientists related to either the use of neuroimaging on a population level, as a tool for gaining more general insight into how the brain works, or the use of neuroimaging on an individual level, to gain more insight into the brains of specific people. The former vision is discussed in more detail in a forthcoming article on the evidence base of neuroimaging (Edelenbosch et al., in preparation). In this chapter, we focus on the vision researchers had of a transition of neuroimaging from a population level to the level of individuals, and what that would mean for the field of education. During the expert session, there was discussion over intervention also taking place on an individual level, a form of personalized learning, if neuroimaging could be applied in this way.

In 2008 the Netherlands' Study Centre for Technological Trends (STT) published the book Brain Visions. In this book 'personalized learning' is identified as an important line of future research for the neurosciences, in particular the adaptation of teaching to sensitive periods in the development of children's brains, the monitoring and assessment of individual abilities, and adaptive cognitive learning systems (STT 2008, pp. 260–353). The idea of individualized neuroimaging contributing to personalized learning is also reflected in the Dana Foundation's Cerebrum, in which experts in brain research present emerging ideas in neuroscience. Here, Howard Gardner suggests that "in 50 years, our successors will laugh at the notion that there is but a single way to teach and assess. Instead, they will seek the best way to teach this concept or subject to this student and the best way for this student to demonstrate understanding" (Dana Press 2010, pp. 44–45).

This application would be in line with developments of both the neuroimaging technology and the field of education. Neuroimaging holds the promise of contributing to the development of individual diagnosis and personalized treatment, especially in the field of clinical neuroscience (Brammer 2009; MacQueen 2010). This could be relevant to the treatment of learning disorders. For example, in dyslexia, the detection of structural alterations of the brain before a child begins reading could allow for intervention to take place before a dyslexic child starts to get behind in school (Raschle et al. 2011). Furthermore, non-medical examples are also starting to emerge, such as measurement of human 'brain preparedness to learn', which could theoretically be applied to education and training (Yoo et al. 2012). At the same time as these developments have been going on, personalized learning or tailor-made education has been a trend in the field of education in the Netherlands and the UK, at least for the past decade. Adaptive (digital) learning technologies have the potential to assist teachers in education targeted at students' needs (The Royal Society 2011). For example, a recent special issue of Educational Technology and Development (2012) focused specifically on personalized learning.

However, during our expert discussion sessions the idea of neuroimaging contributing to personalized education was a contested ideal. Congruently, a number of concerns were identified in the literature. For example, with regard to personalized education, questions can be raised about the degree of choice and responsibility of teachers and students (Cutler et al. 2007). There is also fear that research about the biological basis of intelligence differences can be misinterpreted as evidence for an individual's or group's worth (Gray and Thompson 2004). Furthermore, the use of neuroimaging technology to generate information about individuals brings ethical complications of its own, for example with regard to privacy (Heinrichs 2011).

8.2 Focus Groups with Societal Stakeholders

In order to gain insight into these issues from the perspective those involved in the practice of education, we have reflected on the expert vision of individualized neuroimaging and personalized learning with different stakeholder groups. We conducted 10 focus groups with potential end-users: 3 focus groups with randomly selected parents of one or more children attending secondary school at different levels (n = 19), 3 focus groups with randomly selected secondary school teachers of different subjects at different levels (n = 23) and 4 focus groups with secondary school children attending 5 VWO (n = 35), the highest level of secondary education in the

Dutch school system. A recruitment agency with a large database in Amsterdam was used to find parents and teachers willing to participate for a small fee (max. 40 euro). All participating students attended one particular school in Amsterdam, and were selected randomly from four secondary school classes. Students were between 16 and 17 years old. See Table 8.1 for an overview of the conducted focus groups.

The activity of focus groups with parents and teachers lasted for two hours. The focus groups with students met for 80 min, due to time constraints in their class schedule. All focus groups consisted of two phases. At the start of the first phase, the facilitator asked the participants to write down their associations with "the adaptation of education to individual needs" on post-it notes and made a round to ask participants to explain their associations in more detail, in order to gain insight into the core features of their explanations. These features were written on new post-it notes and clustered by the facilitator on a flip-over. After all post-it notes had been discussed, the participants were asked to name the different categories.

During the second phase of the focus group, the facilitator used a short power-point presentation to introduce neuroimaging as a tool that could be used in the implementation of personalized learning. The participants were asked to make a drawing based on the associations they had with "the use of a brain scan to optimize learning and instruction for each individual child" on a sheet of paper, after which the drawings were discussed. In concurrence with the participants, the facilitator summarized each drawing on a post-it note of a different colour than the first phase one and added it to the flip chart. Subsequently, the participants were asked to reflect on a scenario that included elements relating to the adaptation of teaching

Table 8.1 Overview focus groups

Focus group	1	2	3	4
Teachers (120 min): n	8	7	8	–
M, f	5, 3	1, 6	4, 4	–
Age	25–58	26–51	27–58	–
Havo/vwo, vmbo	6, 2	5, 2	5, 3	–
Subjects (n)	English, art, history (2), theatre (3), Dutch	English, Dutch, geography (3), physical education, technology	History, French, German, biology, art, technology, social studies	–
Parents (120 min): n	7	5	7	–
M, f	3, 4	3, 2	3, 4	–
Age	32–53	50–58	41–55	–
Child's level:				–
Vwo, havo, mbo	3, 3, 4	3, 2, 3	1, 0, 7	–
Student (80 min): n	7	7	7	7
	3, 4	4, 3	4, 3	3, 4

Note The Dutch School system consists of three levels: vmbo, havo and vwo. 60 % of Dutch students are enrolled at vmbo, which generally prepares students for vocational training. The levels of havo and vwo prepare students for higher education. A secondary school can offer one or more levels of education, and a teacher can teach at different levels to sensitive periods in the development of children's brains and the monitoring and assessing of individual abilities by means of a brain scan (see Box 8.1). The participants were asked to write down new associations on a third coloured post-it note, which was also added to the flip chart. Finally, the participants were asked to write down the conditions under which the use of neuroimaging for personalized learning would be acceptable or desirable. The goal of the second phase of the focus group was to reflect freely on the application of neuroimaging for personalized learning, without being held back by uncertainties and technological limitations.

Box 8.1. Scenario Presented to Focus Group Participants
Imagine it is the year 2030. John, an enthusiastic maths teacher, has many years' experience teaching at a secondary school. At his school, children regularly have a brain scan to (1) determine the child's learning strategy and (2) assess the degree of brain development. This information serves to adapt the teaching strategy to the individual needs of the student. For example, some children learn more when they receive instructions in a certain way, while others learn better when they every three months so that John can adapt the teaching material to what the child needs throughout the year. Find the subject to be very challenging. In addition, children learn at different speeds. The students have a brain scan every three months so that John can adapt the teaching material to what the child needs throughout the year.

All the participants had been asked in advance for their permission to record the sessions, which were subsequently transcribed in full. Two of us (RE and FK) thematically analyzed (Braun and Clarke 2006) the transcripts using Atlas.ti. RE started an iterative process of reading and re-reading, coding, and grouping codes into themes, which were reviewed, defined and reflected on together with FK. This process led to the establishment of themes in the data relating to the way participants framed "learning", "the child", "the brain", "brain scans" and "personalized learning". Subsequently, RE and FK constructed a large matrix in which the previously established themes were related back to the storylines of individual participants. This matrix gave further insight into the different patterns of argumentation and how they related to the way participants viewed the application of neuroimaging for personalized learning. In addition, this cycle of interpretation ensured the validity of the data.

8.3 Reflection on Personalized Learning

8.3.1 Introduction

Here, we present the results of our analysis of the focus groups with end-users. Generally, the different groups used similar argumentation ways of framing and therefore their reflections are presented together. Where slight deviations or accent differences exist between the groups, they will be discussed later.

At the start of the focus group, the facilitator asked the participants to write down their first associations regarding “the adaptation of education to individual students’ needs”. This exercise showed that many of the participants considered personalized learning an appealing idea. Possibilities seen by the participants were for example the adaptation of education to the individual’s competences, interests, learning style, learning pace and motivation. According to the participants, these possibilities would contribute to the realization of different values: paying attention to all students’ unique capabilities could contribute to the value of beneficence because learning would become more efficient, fewer students would drop out of school, and talents would be done justice. Treating all students in the way that suited their specific condition would contribute to the value of equity.

Differences in position towards personalized learning became visible when, in the course of the focus group, participants’ ideas of personalized learning became more concrete. The participants each had their own vision of how personalized learning would work in practice, varying from changing the entire school system, to more computerized education, to hiring teaching assistants to help regular teachers. Some participants were of the opinion that, although the idea of personalized learning was “wonderful”, it would not be realistic because, for example, “classes are too big” or “it is outside the comfort zone of teachers”. In this example, the unit of the classroom and the effect on the teacher were important features in the participants’ framing of personalized learning. The idea of personalized learning was “unrealistic” for some participants because it did not fit into their framing of ideal education.

After discussing the idea of personalized learning in general, we introduced neuroimaging as an instrument to help identify a child’s needs. This made the idea of personalized learning more concrete and intensified the trade-offs participants had to make, bringing differences in framing of certain elements of the application into focus more clearly. In the following sections, we pay attention to two questions central to the participants’ framing of neuroimaging for personalized learning: (1) what should be the approach to the learning child? and (2) what can a brain scan actually measure?

8.3.2 What Should Be the Approach to the Learning Child?

Three areas of tension were identified in participants’ argumentation regarding the approach to the learning child. The first related to the implementation of the values of equity or equality in personalized learning. During the focus groups, participants debated whether all children should be treated equally or unequally, because they all have unique learning needs. A second tension can be seen between personalized learning and social learning, because participants often associated personalized learning with individualized education. Finally, the third area of tension concerns the value of freedom on the one hand, and achievement on the other. Participants deliberated whether the adaptation of education to individual capacities would come at the cost of freedom of choice.

Equity and Equality

During the course of the focus groups, in both the first and the second phases, an argument used by participants was that personalized learning would be tailored to the needs of all children, meaning that all children would benefit. At the basis of this argument are both the principles of equality and equity. The participants who expressed this view argued that, while everyone is equal as a human being, everyone learns differently. This means that some children are disadvantaged by the current system: equal treatment leads to inequality. Therefore, children have a right to a more personalized approach, i.e. an equitable approach attuned to their needs.

At the same time, most participants were still of the opinion that all children need to meet the same basic exam requirements at the end of secondary school, mostly because secondary school is viewed as a preparation for the rest of life, teaching children how to function within society. Some participants went a step further in their argumentation: they viewed children as all having innately equal capacities but needing different learning approaches to attain that capacity. The following quote is from a teacher who explained that not only do all children have the right to a personalized approach, this approach can also contribute to everyone achieving the same results. The teacher framed the brain as being plastic, a network of pathways that is malleable to a very large extent.

> I think that neurologically, there are already things known, for example about dyslexia. We know that people without dyslexia have a pathway from left to right. A dyslectic person does not have that. But by developing another pathway in neurology, you get to the place that you need to get to, and that is what I like, to be able to do that through neurology. Because then a child can still become professor. Because he learns to go via another pathway. That is a very good development. [teacher]

Personalized learningand social learning

In the first phase of the focus group, but even more so in the second phase of the focus group, participants expressed the fear that personalized learning could be at the cost of the social setting in the classroom and the relationships students have with each other and with the teacher. According to these participants, personalized learning focuses on individuals, and the technology of neuroimaging enhances the idea of individuality because it emphasizes that everyone has their own unique style of learning. However, many participants considered it important for children to learn how to work together. For example, one student mentioned:

> Being social is the only thing you don’t learn explicitly as a subject at school, but I think that is one of the most important things you learn. [student]

Being social is not only fun, different participants argued that it is necessary for students’ future careers. Moreover, interaction between people was seen as an important component of learning itself. A question that arises here is how a teacher should decide which form of learning to concentrate on, and what the beneficence is in that case with regard to both the student and the teacher.

It is worth pointing out that, at the start of the focus group activity, when discussing general associations with personalized learning, participants compared the current school system to a factory, the problem being the shortage of individual attention. Personalized learning could be a way of overcoming this situation because it is more personal, referring to the relationship between the child and the teacher. However, when the brain scan was introduced in the second phase of the focus group, some participants felt the "personal" was actually taken out of personalized learning. For example, one parent argued the following:

> A brain scan, I wonder, why do they want this? We all want to keep school personal, and give a child attention. Well, with this you can just send a teacher home. (…) I find it very scary. School is more than only testing. [parent]

According to participants, neuroimaging would come between the child and the teacher and the child's social experience with his or her peers, at the cost of their relationship.

Freedom and determinism

Some participants argued that adapting learning exactly to the needs of students would not be a good development because students need to develop a way to deal with uncertainty in finding their own way. For example, some participants mentioned that children should not know which path they should take to meet the least resistance, the path of "optimized education". Taking a nosedive, learning from mistakes and "going on adventures" is not only the way that individuals learn the most, it is an essential part of human life. The questions below, asked by a participant in reaction to the presentation of the scenario in which a teacher uses the results of a personalized brain scan, illustrate this:

> What is important in this life? How important is it that someone achieves optimally and gets all there is out of life? Why would we want that? Why do we want to know everything, why do we need a strategy for everything? [teacher]

There is a relation with determinism to this way of framing when participants argue that there is an optimized path, or the possibility to know everything in advance. Some participants were afraid that the scanner would be used to predict the future development of a child's brain. These participants did not want to know what their brain looks like or what their talents or weaknesses are. One parent argued:

> You yourself also don't want to know when you are going to die. No, let's be honest. I get goose bumps from the idea. (…) because of research into the brain it is determined how your life is going to be. I think enough is determined already. [parent]

Another example of this is the following quote, which stems from one of the focus groups with parents. At the start of the second phase of the focus group, the facilitator asked one of the participants what he wanted to express with his drawing. He replied:

> Actually, I feel more like angry. (…) When you start to do something like this… You are sentenced for life, you have a criminal record because your brain does or does not function well. Similar to what we have now, a white school and a black school and whatever. You get a stamp with 'small brain' or 'large brain'. [parent]

Participants were also afraid that this type of determinism would become institutionalized, as described here by one of the teachers:

> What I am afraid of is that [on the basis of a brain scan] different streams will be created. You belong there, you there, you there and you there. […] There is a set route and you almost cannot deviate from the route because you are destined to become like that. [teacher]

According to some participants, both the loss of a social environment and the emphasis put on achievement would make schools treat children less as "humans". Many robots and conveyor belts were drawn during the drawing exercise. The moral anti-reductionist argument lies at the root of the fear that a lot of participants have of reducing children into achievement machines. For example:

Parent 1	"Achievement, achievement, achievement. It's really a big reason why many children leave school early, because of that oppression. And if they then also get a stamp with their brain on it, well then…"
Parent 2	"Children that commit suicide."
Parent 1	"Yes, that is what I mean."

The anti-reductionist argument has such significant value for these participants that they find it difficult to see any possibility for the application of neuroimaging for personalized learning.

8.3.3 What Do Brain Scans Measure

The second question central to participants' appreciation of neuroimaging was what could be made visible with a brain scan, and what a brain scan could offer that other types of assessments could not. On the one hand, some participants argued that neuroimaging was a very objective way of gaining insight into a child's needs. On the other hand, some participants did not think that neuroimaging could do justice to the child as a human being and some were afraid that this development would limit the definition of intelligence, inflicting injustice on children who did not measure up to that definition.

Objective measurement

Some participants thought that the use of neuroimaging would improve personalized learning because it could be used as an objective instrument, often referring to the common Dutch expression "meten is weten" (to measure is to know). Many participants considered a brain scan to be more "objective" than a written test or the teacher's perspective. It is a more direct way of measurement, cutting out the possibility of children giving socially acceptable or "lazy" answers. Teachers interpret why children behave like they do, simply because they are human beings that have relationships with the students which can obscure what is "really" going on. Neuroimaging would make it possible to "see" things that remained hidden before, whether or not such a possibility would be a good development. As one teacher put it:

> I think an important condition is that you cannot force people to do this. Because then I get the image of a school doctor that says: take off your clothes. And then it turns out you have a small one. [other participants laugh]. No, but you are just naked, you are simply turned inside out. [teacher]

Different applications of neuroimaging relating to this objectivity were mentioned. For example, some participants thought that it would be useful to use neuroimaging to see when a child was underachieving in relation to his potential. According to these participants, that would make the education system fairer to students who had less opportunity due to their home environment. Similarly, children who had struggled with a particular subject for a long time would have objective proof to show that was the case. This could enable informed decision making about for example the type of education or job a student would have the most aptitude for. Another argument made was that, more importantly than scientifically proving what the capacity of a child was, the scan could give objective insight into why children were doing what they were doing, creating the possibility of intervening knowledgeably. Especially when a child was experiencing problems, a brain scan would make it possible for teachers to react on the basis of objective information, whereas they now often have to work on the basis of an instinctive feeling. According to one of the parents, it would also help in cases where there was a disagreement between for example the teacher and the parents as to what the cause of a learning problem was:

> I don't see how, as a parent, you can argue with a teacher about the results of a scan. Because a scan is a scan and that's the way it is. [parent]

Reduction of a child to a scan

Some other participants, in contrast, thought a brain scan would not do justice to the child. This way of framing comprised three elements: (1) a brain scan only captures one moment in time and (2) does not do justice to the complexity of human beings and (3) their environment. During the second phase of the focus group, a number of participants argued that a brain scan is only a snapshot, a "picture" or "map" that cannot give any insight into the further development of the child. One of the parents reasoned in the following way:

> I get the feeling that you make a map of someone. Instead of letting a child be a child, you have to make him into a map. And you don't let a child be determined by his development, everything changes. In the first year, he is like this… It's all a snapshot. And when you use this [a brain scan], I'm afraid that you only take the first two years with you, for the rest of your life. [parent]

Aside from the temporal argument, some participants reasoned that a scan could not capture the essence of what it is to be human because that is something that a machine can only approximate, simply because machines are fundamentally different from humans. Factors that participants are afraid to lose are, for example, the child's personality, motivation and emotions. One of the parents put it as follows:

> I wrote down the 'sweet child'. The scan must know if a child can learn, can talk, can think, can love, can believe, can do, can be creative, unconscious reactions, can DO something

> and can WANT something. A scan cannot do that. I don't believe a word of it. (…) All these things, are the essence of a human. A machine cannot do that, in my opinion.

This was echoed by one of the students:

> But maths, physics, geography and biology, these are different interests for different people. We need FAR more technology before we can read all this from the brain, because there are thousands of factors that influence the brain. [student]

In addition, some participants argued that a brain scan reduces a child to his brain, ignoring all other factors in the environment that influence the development of the child. The brain image does not say anything about for example the circumstances at home, or the quality of the teacher.

According to these participants, due to the technical reductionism of the brain scan, it has little predictive value. By allowing the scan to determine what kind of education a child needs, participants are afraid to lose the complexity of a child's development, doing injustice to the human being. This line of argument leads to the conviction that brain scans do not have a part to play in the classroom unless a child is experiencing problems or is not developing according to expectations. Participants that were afraid of technological reductionism only saw the benefit of brain scans in problematic situations that could not be dealt with by other means.

Reduction of child performance to the brain

A second concern some participants had was that the brain scan would more narrowly define intelligence. Some participants explained that a brain scan would change intelligence into a measure determined by the size, structure and function of the brain, not by how a child faced problems or what a person achieved in life. Intelligence would thus be reduced to an image. Furthermore, some participants were afraid that scientists would use neuroimaging to look for a very specific kind of intelligence in the brain. One of the students argued:

> How can you measure who is smarter? Is someone smarter who understands economy or someone who understands biology very well? It's not possible to measure this. [student]

This would go against their idea of there being different kinds of intelligence. This is also demonstrated by the following discussion, which took place during one of the focus groups with parents:

[Parent 1] "You have very intelligent people who are just not that smart and you simply have less intelligent people who are streetwise."[1]

[Parent 2] "A doctor's reading brain. What is a doctor? He knows, he reads a lot and remembers that, but sometimes they are not able to think creatively."

[Parent 1] "Take a surgeon, he can cut really well, but he doesn't speak one friendly word, you know."

[1]Translation of the Dutch word "boerenslim".

According to the participants, this development would be immoral for several reasons. First, having the method of education adapted to a narrow definition of normal development could lead to a loss of diversity. This would have negative consequences for society, which needs people with different skills.

A second fear expressed by some participants was that children would be evaluated by their brain scan, not by their actions. Here, the issue of privacy becomes especially important. All parents and teachers agreed that this information should not become common knowledge. One student argued:

> Well it's the ultimate breach of privacy, because, well it's possible to map out your whole… being. If that information becomes public… That must never happen. [student]

Some participants were afraid that children would be labelled by the results of their brain scans, influencing how they were seen and treated by others. For example, participants argued that this information about the brain could be used against them by insurance companies or by potential employers. In addition, they thought a brain scan might lead to self-stigmatization when a child had insight into his or her capacities.

According to some parents and teachers, this fixed view of the self might lead to children being less inclined to try, as they would have an excuse for why they were not doing well in school. At the same time, students that had been told they had certain talents also might not live up to their potential because they would feel that they did not have to work for it. Participants using this frame often argued that neuroimaging should only be used "in a positive way": only to discover opportunities, like talents that could be stimulated. The technology should not be used to identify problems, since the identification of problems could lead to labelling and stigmatization.

8.3.4 Differences Between Parents, Teachers and Students

The different arguments presented above were all visible in the focus groups with parents, students and teachers. When comparing the general trends in the focus groups, the main differences discerned between the different groups was the emphasis laid on some elements compared with others, and the effect that application of personalized learning was thought to have on the stakeholders' own role. Students put most emphasis on the importance of social learning and the freedom to choose their own path, arguing for example that:

> I don't think others should be allowed to decide what you are going to do. We all have freedom of choice. [student]

Many participating parents were worried that their child would be stigmatized if it were possible to see the capacities of their children on a brain scan. Teachers were most engaged with the practical aspects of the implementation of this application. Many were concerned that personalized learning would change their role in education considerably, from a teacher to a learning supervisor. This was visible in, for example, the following excerpt, taken from a discussion with teachers about what they would be afraid to lose with the introduction of individualized neuroimaging in the classroom.

Teacher 1 "Passion, telling stories."
Teacher 2 "No, we mustn't lose that."
Teacher 3 "Because that is the power of teachers."
Teacher 4 "And the professional knowledge, because that has not been included yet. Because if you are a coach… aside from the professional knowledge everyone can be a coach. A teacher has knowledge that is unique for a particular subject."

In addition, parents and teachers differed on the issue of privacy and how data should be handled. Some students did not see privacy with regard to learning as that much of an issue; one student explained:

> With everything that relates to learning I don't really care, but with other things that have nothing to do with learning I do. (…) A lot of things at home they [the school or the teacher] really don't have to know about. But… what I think of certain subjects or what I can do, I don't mind if they know [student]

Teachers mostly thought they should be in charge of the information, which should be a part of the private files on students they manage now. Parents wanted the information to be property of the child or the family and students agreed they themselves should be the owners of the information about them.

8.4 Conclusion and Discussion

In our analysis, a number of tensions have been identified between different values that appear to be conflictive, like the tension between equality and equity, and between freedom and achievement. As discussed, value conflicts are difficult to resolve. However, when we looked more closely at the argumentation patterns around these values, it became clear that in some cases it is not simply the values that are at odds with each other. Sometimes, differences were visible in the way that participants framed the application: which elements were central to their stories, and how were they understood? We have summarized this in Table 8.2. Here, we show different areas of tension and the underlying differences in how particular elements are framed.

These different framings also lead to different ideas about potential applications of neuroimaging for personalized learning. The basis of the arguments against neuroimaging for personalized learning was often a rejection of reductionism. Participants argued against the reduction of different elements in their stories. We have distinguished three different kind of anti-reductionism. The first and second kind both relate to the reduction of a child, the first focusing on what a scan can do, and the second on what a scan should do. The third form of anti-reductionist argument relates to narrowly defining intelligence. Although reductionism is the main issue in all three cases, the arguments are in fact very different, and lead to different ideas about application possibilities.

Table 8.2 Areas of tension and elements framed differently when answering the questions '*How to approach a learning child?' and 'What can a scan measure?'*

	Framework	Learning	Child	Brain	Scan	Neuroimaging for Personalized Learning
What should be the approach to the learning child?	Equality versus	Focus on equal process	–	–	–	Not at the cost of equal treatment
	Equity	Focus on equal outcome	–	Plastic	–	Yes
	Social learning versus	Focus on social process	Needs social interaction	–	–	Not at the cost of social learning
	Individualized learning	Focus on outcome	Should make most of potential	–		Yes
	Freedom versus	Focus on freedom to make mistakes	Does not need to be perfect	Fixed	Can make predictions	No application because freedom is impaired
	Achievement	Focus on outcome	Should make most of potential	–	–	Yes
What can a scan measure?	Objective measurement versus	–	Can be captured by scan	Is the centre of learning	Objective instrument	Yes
	Reduction of child	–	Is very complex, cannot be captured by scan	Is not the only place where learning occurs	Can capture only one moment in time	Only in case of learning problems
	Narrow definition intelligence	–	Is intelligent in different ways	Intelligence is more than what is learned in school	Will narrowly define intelligence	Only to discover opportunities

The first anti-reductionist argument is that neuroimaging is not able to capture the essence of a child. This tension is part of a larger philosophical discussion about the relation between brain and mind. Participants that came with this argument mostly argued for a brain scan to be done only in case of learning disorders. The second argument is that a scan limits children in their freedom. This discussion is a paradox in the sense that here, images of societal determinism tend to creep in, while at the same time scientific insights are providing more and more evidence for plasticity of the brain. The participants who were of this opinion did not see any application possibilities for neuroimaging in the classroom at all. Finally, the third argument is that a scan would define intelligence too narrowly. This relates to the ethical discussion about how intelligence should be constructed, and what role researchers should play in this. Participants using this argument often thought that a scan should be used to create more opportunities for a child, by looking at what a child could do well instead of at problem areas. These three forms of reductionism all need to be addressed and discussed separately in a (future) science-society dialogue setting.

Other issues that need to be further discussed relate to the role that the teacher, parents and students play in this development and the issue of privacy. Sheridan et al. (2006) argue that value conflicts often occur when technological or cultural developments force people to take on new or too many roles in their profession. Increased individualization will create further demands on educators, and in the short term it might become more difficult to align the desires of students, parents and teachers with, for example, governmental performance standards and teachers' personal value systems. Regarding privacy, participants argued that it was important that the information would be protected to prevent social or economic harm. Considerations that have to be taken into account with regard to privacy and the medicalization of intellectual ability are discussed in detail by for example Heinrichs (2011).

Some ethical issues addressed in the literature did not arise in the focus groups. According to Gray and Thompson (2004) it is considered "mildly impolite" in the United States to attend to the differences between people, because it undermines the higher ethical principle of social equality. The suggestion that individual differences in ability have a biological basis would be considered "distinctly impolite" because evidence about this could be misused as evidence about an individual's or group's social or moral value. This type of argument was not recognized during the focus groups conducted in the Netherlands. Instead, participants often said that everyone was unique in their mental abilities, and that there were many different types of intelligence, all having their own value. They did not express the fear that this type of application would lead to stigmatization of groups based on race or class.

We have seen that participants hold different views on the approach to a learning child and on what a brain scan can and should measure. This leads to difference in attitude towards the application of neuroimaging for personalized learning. Different views indeed have to do with a difference in 'problem setting', the defining of the elements that are important in relation to the application. Although the arguments encountered in the focus groups were not surprising or novel to 'new and emerging science and technology' (NEST)-ethics (Swierstra and Rip 2007), we have found that detailed analysis of the societal stakeholders' different types of framing can be a valuable way of doing a pragmatic ethics and can contribute to an RRI process. Below, we reflect on the value of this approach and on the different choices we made in the focus groups.

First of all, we decided to include teachers, students and their parents in this study. There are two issues regarding this selection that we would like to discuss. The first relates to the inclusion of participants that are only involved in one aspect of education, the actual teaching and learning in the classroom. There are a number of other societal stakeholders that have not been included in this study. For example, as mentioned in the introduction, educational designers are already developing and implementing digital technologies for personalized learning, for which neuroimaging could possibly have great potential. However, we would argue that the end-users participating in this study are the ones that have to cope with the most significant changes in the future. In the Netherlands, the past has shown how important it is to include the values of end-users in education innovation, as is shown by the Dutch parliamentary inquiry on educational innovation in 2007–2008 (Commissie Dijsselbloem 2008). Second, as Howard Jones and Fenton (2012) mention, it may be argued that the public is not well informed enough to contribute to an informed ethical discussion. However, in complex issues like this, uncertainty and uninformedness regarding applications are not confined to one group, as none of the stakeholders of society or science have knowledge of all aspects of the application. We feel that with this project, we have shown that it is possible to have valuable ethical discussions even with (older) children about the value this application would have for them. Some values, such as the fear of determinism, are based on assumptions about neuroimaging that are not informed by the results of neuroscience. However, this was only the case for some arguments, and frame analysis makes it possible to separate these arguments from others.

The construction of frames was a way to work with both the value plurality of participants visible within and between stakeholder groups, and the inherent uncertainty of the application. By disentangling the frames in this way, we hope to have shown that the values at stake for stakeholders are to a large extent dependent on the way the problem is set, or the situation is framed. The frame construction was a way of unraveling the mesh of arguments that were put forward during the focus group discussions. More insight was gained into the features of neuroimaging and personalized learning that are problematic for societal stakeholders and more insight was gained into the beliefs, assumptions and appreciations that underlie these features. At the same time, the uncertainty of the application was taken into account to a large extent because the discussion was more about the idea of the application than about concrete technical matters.

In order to contribute to neuroimaging research that is "anticipatory, reflective, inclusive, deliberative and responsive", as called for by Owen et al. (2013, p. 29), further and continued stakeholder engagement with regard to the application of neuroimaging for personalized learning is necessary. As argued by Correljé et al. in Chap. 10 of this volume, value specification with stakeholders takes place in a dynamic and emergent social process, as perceptions and expectations with regard to an innovation and the domain of application can change over time. The approach focusing on framing and the specific frames that have been described in this chapter could be utilized in different ways during future dialogue, as indicated by Kupper (2009). First, frames can be reflected upon in order to reveal implicit assumptions and exclusion effects of current habits, practices and institutions. Second, frames can be used to identify key areas of conflict and misunderstanding: multi-perspectival problem construction. For example, a departure point for dialogue could be the ways in which different stakeholders from science and society frame what a child should achieve at school. From there, potential applications of neuroimaging can be discussed in a way that takes these different framings into account. Third, it is possible to organize sessions in which the different frames are used as input for the construction of new, integrated frames.

Finally, we wish to reflect on some limitations of this study. First, we reflect on our own role in this process. In this form of pragmatic ethics, the ethicist becomes facilitator of moral argumentation, rather than the provider of arguments or critiques. Nevertheless, the facilitator still plays an important and active role in the construction of framing, both in the design and conduct of the focus groups. The application of individualized neuroimaging for personalized learning was a guiding vision constructed during the expert session, and was reflected in the literature. This complex and blurry vision was then taken by our research team and introduced to the stakeholders, possibly taking on new forms during the focus group. It can be argued that our role has been constitutive to the very forming of the future, and therefore not a good example of "managing expectations" as advised by Ansari et al. (2011). At the same time, if we wish to anticipate future developments, we need to gain insight into the expectations that exist among neuroscientists. These expectations shape the future in the sense that they contribute to the building of research agendas and the attraction of actors that could play a role in this future (Borup et al. 2006). We therefore need to find ways to gain insight into these expectations without creating new hypes. A first step we take here is being explicit and reflexive about our role as facilitators.

Second, the question arises to what extent the results of this study are generalizable, both with respect to the location of the study, and with regard to the innovation of neuroimaging. It needs to be said that the Netherlands has a particular education system with different streams of different levels. Which level a child can attend is determined at the end of primary school, to a large extent by a national test that all children take. In addition, the Dutch education system places much emphasis on the freedom and autonomy of students, with the goal of stimulating independent learning.[2] The local values embedded in and outside of educational

[2]For more information the secondary school system in the Netherlands, see the website of the Dutch Ministry of Education, Culture and Science at http://www.government.nl/ministries/ocw. For a comparison of the Netherlands with other European countries, see http://www.trendsinbeeld.minocw.nl/grafieken/2_1_2.php (last accessed August 17, 2014).

practice undoubtedly are reflected in the framings of participants, and it is possible that the participants in this study were more fearful of determinism because of this cultural background. However, it is expected that most areas of frame conflict identified in this study also feature in most other welfare states. With regard to the generalizability of the technology studied, it can be argued that the opportunities and concerns identified by participants also have relevance for Responsible Research and Innovation processes involving other technologies that have predictive, diagnostic value and emphasize the differences between individuals, such as (neuro)genomics. Although genomics can be argued to be a step further away from education and research into stakeholder perceptions of the application of genomics to education have not yet been undertaken, we recognize the fear of determinism, labelling and stigma, discrimination and privacy encountered during our focus groups in the literature about the social, legal and ethical aspects of genomics.

Irrespective of whether or not the individual brain scan at school lies ahead, it is valuable to understand what this future would mean to participants and why the framings that have been described in this chapter also give insight into the way people perceive more general developments in the neurosciences and the boundaries between applications in health care and education. The technology of neuroimaging can alter the view we have of our minds and ourselves. The question therefore is to what extent this technology should 'make visible', and what should be done with this information. It is important to somehow anticipate the potential applications of new technologies when it is still possible to influence the course of innovation (Collingridge 1981). The next step in our deliberation process is the reflection on these different framings with stakeholders in a science-society dialogue. We believe that insight into these framings will make it possible to have a more constructive interaction in which ethical issues are not reduced to scientific ones and an environment is created in which it is possible to discuss what is at stake.

Acknowledgments This chapter is the result of the research project Neurosciences in Dialogue, which is part of MVI and the Centre for Society and the Life Sciences, funded by the Netherlands Genomics Initiative. The authors would like to thank Sanne Koot for her contribution to the collection of data, and the anonymous reviewer for his/her insightful comments.

References

Ansari, Daniel, Bert Smedt, and Roland H. Grabner. 2011. Neuroeducation—a critical overview of an emerging field. *Neuroethics* 5(2): 105–117.

Bohman, James. 1996. *Public deliberation: pluralism, complexity, and democracy*. Cambridge, MA: MIT Press.

Borup, Mads, Nik Brown, Kornelia Konrad, and Harro van Lente. 2006. The sociology of expectations in science and technology. *Technology Analysis & Strategic Management* 18(3): 1–15.

Brammer, Michael. 2009. The role of neuroimaging in diagnosis and personalized medicine-current position and likely future directions. *Dialogues in Clinical Neuroscience* 11(4): 389.

Braun, Virginia, and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3(2): 77–101.

Collingridge, David. 1981. *The social control of technology*. Milton Keynes: Taylor & Francis Group.

Commissie Dijsselbloem. 2008. *Parliamentary inquiry educational innovation: 'Tijd voor Onderwijs'*. Summary available at http://www.parlement.com/9291000/d/svrapportonderwijs.pdf (retrieved June 29, 2013).

Cutler, Tony, Barbara Waine, and Kevin Brehony. 2007. A new epoch of individualization? problems with the 'personalization'of public sector services. *Public Administration* 85(3): 847–855.

Edelenbosch, Rosanne, Frank Kupper and Jacqueline E.W. Broerse. 2014. Evidence based learning and neuroimaging: reflections with potential end-users. *in preparation*.

Gray, Jeremy R., and Paul M. Thompson. 2004. Neurobiology of intelligence: science and ethics. *Nature Reviews Neuroscience* 5(6): 471–482.

Grin, John, and Armin Grunwald. 2000. *Vision assessment: shaping technology in 21st century society*. Berlin: Springer.

Hardiman, Mariale, Luke Rinne, Emma Gregory, and Julia Yarmolinskaya. 2011. Neuroethics, neuroeducation, and classroom teaching: where the brain sciences meet pedagogy. *Neuroethics* 5(2): 135–143.

Heinrichs, Jan-Hendrik. 2011. The sensitivity of neuroimaging data. *Neuroethics* 5(2): 185–195.

Howard-Jones, Paul A., and Kate D. Fenton. 2011. The need for interdisciplinary dialogue in developing ethical approaches to neuroeducational research. *Neuroethics* 5(2): 119–134. doi:10.1007/s12152-011-9101-0.

Keulartz, Jozef, Maartje Schermer, Michiel Korthals, and Tsjalling Swierstra. 2004. Ethics in technological culture: a programmatic proposal for a pragmatist approach. *Science, Technology and Human Values* 29(1): 3–29.

Kupper, Frank. 2009. *Democratizing animal biotechnology*. Oisterwijk: Box Press.

Kupper, Frank, and Tjard Cock Buning. 2010. Deliberating animal values: a pragmatic—pluralistic approach to animal ethics. *Journal of Agricultural and Environmental Ethics* 24(5): 431–450.

Kupper, Frank, Linda Krijgsman, Henriette Bout, and Tjard Cock De Buning. 2007. The value lab: exploring moral frameworks in the deliberation of values in the animal biotechnology debate. *Science and Public Policy* 34(9): 657–670.

MacQueen, Glenda. 2010. Will there be a role for neuroimaging in clinical psychiatry? *Journal of Psychiatry and Neuroscience* 35(5): 291–293.

Maxwell, Bruce, and Eric Racine. 2012. The ethics of neuroeducation: research: research, practice and policy. *Neuroethics* 5(2): 101–103.

Oudshoorn, Nelly, and T.J. Pinch. 2005. *How users matter: the co-construction of users and technology*. Cambridge, MA: MIT Press.

Owen, Richard, Phil Macnaghten, and Jack Stilgoe. 2012. Responsible research and innovation: from science in society to science for society, with society. *Science and Public Policy* 39(6): 751–760.

Owen, Richard, John Bessant, and Maggie Heintz, eds. 2013. Responsible innovation. Chichester: Wiley.

Dana Press. 2010. Cerebrum 2010: emerging ideas in brain science. Washington: Dana Press.

Raschle, N.M., Maria Chang, and Nadine Gaab. 2011. Structural brain alterations associated with dyslexia predate reading onset. *NeuroImage* 57(3): 742–749.

Rip, A. 2012. Futures of technology assessment. In *Der systemblick auf innovation—technikfolgenabschatzung in der technikgestaltung*, ed. Michael Dekker, Armin Grunwald, and Martin Knapp, 29–39. Berlin: Edition Sigma Verlag.

Roelofsen, A., Roy R. Kloet, Jacqueline E.W. Broerse, Tjard de Cock Buning, and Joske F.G. Bunders. 2010. Guiding visions in ecological genomics: a first step to exploring the future. *New Genetics and Society* 29(1): 19–36.

Rohracher, Harald. 2003. The role of users in the social shaping of environmental technologies. *Innovation: The European Journal of Social Science Research* 16(2): 177–192.

Schön, Donald A., and Martin Rein. 1995. *Frame reflection*. New York: Basic Books.

Sheridan, Kimberly, Elena Zinchenko, and Howard Gardner. 2006. Neuroethics in education. In *Neuroethics: defining the issues in theory, practice, and policy*, ed. Judy Illes, 265–275. New York: Oxford University Press.

STT Netherlands Study Centre for Technology Trends. 2008. *Brain Visions*, ed. Ira van Keulen. The Hague: STT.

Swierstra, Tjalling, and Arie Rip. 2007. Nano-ethics as NEST-ethics: patterns of moral argumentation about new and emerging science and technology. *Nanoethics* 1(1): 3–20.

The Royal Society (ed.). 2011. *Brain waves module 2: neuroscience*. London: The Royal Society.

Wynne, Brian. 1996. May the sheep safely graze? a reflexive view of the expert-lay knowledge divide. In *Risk environment and modernity towards a new ecology*, ed. Scott Lash, Bronislaw Szerszynski, and Brian Wynne, 44–83. London: Sage Publications Limited.

Yoo, Julie J., Oliver Hinds, Noa Ofen, Todd W. Thompson, Susan Whitfield-Gabrieli, Christina Triantafyllou, and John D.E. Gabrieli. 2012. When the brain is prepared to learn: enhancing human learning using real-time fMRI. *NeuroImage* 59(1): 846–852.

Chapter 9
Natural Food: Organizing 'Responsiveness' in Responsible Innovation of Food Technology

Dirk Haen, Petra Sneijder, Hedwig te Molder and Tsjalling Swierstra

Abstract Responsible innovation requires mutual responsiveness between various stakeholders around technological innovation. But in public engagement exercises, concerns about ethical, cultural and political impacts are too easily set aside, so that no one is actually encouraged to discuss responsibilities for these impacts. A typical example in the field of food innovation is the consumer's recurring concern for natural food. In discussions, both consumers and engineers tend to consider the meaning of naturalness as subjective and private. In this chapter, we present an interdisciplinary design tool for public engagement that is more hospitable to such concerns, based on the Discursive Action Method and Techno-Ethical Imagination. We describe the advancements we made and the obstacles we faced when applying this tool in two dialogue workshops on novel foods and naturalness.

Keywords Responsiveness · Food technology · Naturalness · Soft concerns · Stakeholder dialogue · Conversation analysis · Techno-ethical imagination

D. Haen (✉) · T. Swierstra
Maastricht University, Maastricht, The Netherlands
e-mail: dirk@dirkhaen.nl

T. Swierstra
e-mail: t.swierstra@maastrichtuniversity.nl

P. Sneijder · H. te Molder
Wageningen University and Research Centre, Wageningen, The Netherlands
e-mail: petra.sneijder@wur.nl

H. te Molder
University of Twente, Enschede, The Netherlands
e-mail: hedwig.temolder@wur.nl

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_9

9.1 Introduction

Over the last twenty years, professionals in food technology and food industry have become fully aware that successful technological applications and innovations in food production and processing require some degree of societal acceptance, and that public trust in processed food is far from a given. The concept of responsible innovation (RI) offers a promising, if not necessary step towards more socially robust innovations that take into account the very needs and concerns that citizens may have. In the context of both morals and markets, this makes sense.

Von Schomberg (2011) offers a useful definition of the concept: “Responsible Research and Innovation is a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to 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).”

One of the key features of this approach is mutual responsiveness. Stilgoe et al. (2013) point out that responsiveness refers to both reacting and responding; it involves “responding to new knowledge as this emerges and to emerging perspectives, views and norms.” (cf. Pellizzoni 2004). One responds, in a modest sense, by answering to and taking into account emerging concerns: if not by concrete actions then at least by words. Responsiveness thus implies the readiness to provisionally acknowledge the legitimacy of raised concerns, and the willingness to think and speak about them. In this sense it somewhat overlaps with other dimensions of responsible innovation that Stilgoe et al. (2013) have identified, such as inclusive dialogue and reflexivity of technology actors.

Although mere “talk” may not be sufficient for true responsiveness, it is certainly required as a start. This kind of responsiveness suggests that various stakeholders explore and assess the opportunities and threats of a new or emerging technology through ongoing dialogue, and discuss potential distributions of responsibilities. In the cited paper, Von Schomberg specifies a range of commonly accepted values (at least, within the EU) that, apart from technological and economic considerations, serve as topical anchor points for such engagement: the promotion of social justice, equality, solidarity, fundamental rights, quality of life, protection of human health and the environment. “We can’t make an appeal to concepts of the good life, but we can make an appeal to the normative targets which we can find in the Treaty on the European Union”, Von Schomberg explains.

But what does the ideal of responsiveness really mean if certain citizen concerns are structurally marginalized and barely recognized as legitimate public issues in the process of public deliberation? Yet, this is what we see happening in the context of dialogues on food and food technology. As Swierstra and te Molder (2012) have argued, the agendas of policy and technology actors here show a strong bias against cultural, moral and political concerns surrounding technologies, such as those concerning implications for identities and meanings, world views and life styles, the (global) distribution of power and control, and—as Von Schomberg already sets aside—conceptions of the good life.

Collective reflection and attribution of responsibilities for any of these issues remain out of the picture. Such concerns have indeed been raised in public controversies on GMO's in Europe, but have been frequently dismissed by technology actors and decision makers as rather emotional, non-scientific and uninformed responses. In an in-depth study on public attitudes and evaluations of biotechnology in agriculture and food in Europe, Marris et al. (2002) conclude that such perception of the public (i.e. one relying on a "deficit-model") failed to recognize the social, cultural and institutional nature of such concerns. Arguably, if such concerns influenced the debate and resulting regulation, it was not because these were taken as substantial, legitimate and serious issues on their own terms.

This is also true of less prominent discussions of food technology. In the Netherlands, for example, certain food additives (such as colorings, preservatives, and flavor enhancers known as E-numbers) have suffered from a bad reputation for years and are still subject to public debate. While experts primarily focus their contributions on the health risks (or lack thereof) of additives, they tend to overlook the broader concerns on what the use of additives mean for our food culture, good taste, and the ideal of naturalness (Haen 2014).

Another example concerns the discussion of functional foods—products that are designed for achieving health related goals such as losing weight or reducing stress. While experts tend to foreground the health claims of functional foods, Sneijder and te Molder (2014) show that citizen-consumers may treat these claims as a threat to their identities as autonomous and critical consumers. In a discursive psychological study based on group interviews, Sneijder and te Molder demonstrate how participants display themselves to be immune for these health claims by referring merely to the taste of innovative products as grounds for choosing them. Such practices show that consumers treat the health claims that are communicated to them as if these would demand their blind trust and absolute confidence in the food producers. In response, they (re)establish their identities as autonomous consumers who simply know what is best for them.

Many of these concerns have in common that they are instantly privatized—sometimes by food engineers or policy makers, and sometimes by citizen-consumers themselves (Haen 2014). By privatization we mean that in terms of what people know, or what people do, no room is offered or claimed for collective accountability practices. Yet, such room is a necessary condition for mutual responsiveness. After all, responsiveness starts with the readiness and willingness to think and speak about raised concerns. One cannot be responsive without engaging in the game of giving and asking for reasons, arguments, explanations and so on.

At face value, it is no surprise that the relevance of cultural, moral and political concerns about technologies is easily overlooked in public dialogue. They are not the classic issues, so to say. In contrast to issues like food security, safety, health and environmental risks, it seems that these concerns fail to meet common standards for becoming legitimate topics on the innovation agenda. That is, most of them are supposedly hard to quantify and measure; they do not seem to indicate any harm or serious public repercussions; and no direct, unambiguous causal links are established between the technology itself and what citizens are concerned about. For the purpose of our argument, we dub these concerns 'soft', only to indicate that influential actors, such as policy makers and food engineers, rarely consider them on a par with 'hard impacts'.

The crux of our approach is that this dichotomy of soft concerns and hard impacts is detrimental to meaningful dialogue on food innovations. Soft concerns should not be dismissed from the agenda all too soon, for several reasons. First, they are not as subjective, irrational and private as they seem at face value. The implicit standards that they would fail to meet are questionable and multi-interpretable in themselves; what is truly of public relevance, for example, can be subject to public deliberation itself and needs public attention before it becomes visible and public. Second, soft concerns simply express what is dear to people as citizens, consumers, clients, users. As the public controversy on GM food in Europe has shown, successful innovation fails 'downstream' if those concerns that we may not classify as scientific or rational are not taken seriously in the first place. Third, a more inclusive dialogue gives way to a broader range of voices and repertoires, which is generally more in line with democratic principles. Apart from considerations of efficacy, these should also play their part in justifying the ideal of responsiveness. Last but not least, a certain degree of trust between producers and consumers is both a precondition for, and a potential result of public engagement exercises. Trustworthiness involves the recognition of other people's interests and concerns, whether hard or soft.

In sum, science-society dialogues on technological innovation may seem responsive, but they do not automatically result in a serious and critical exploration of Von Schomberg's envisioned "(ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products". Blok and Lemmens (this volume, Chap. 2) are right in raising doubts about the viability of the notion of responsible innovation as is currently envisioned: if only the "grand challenges" are to be discussed, and a "symmetry between moral agents and moral addressees" is considered a given, responsiveness is a rather naive ideal. Rather than giving into skepticism, however, we like to believe that responsiveness is something that can be organized—albeit imperfectly. In the typology of RI approaches drawn by Koops (this volume, Introduction), we thus consider our approach to be process-oriented—keeping in mind that procedures, by nature, require continuous improvement. Responsiveness in this sense is a regulative ideal and far from naive. Although substantial results from engagement exercises may not be entirely decisive in assessing the quality of responsiveness, at least the process requires, and allows for, further attention and specification. The current chapter describes how public engagement can be made more inclusive, content wise, in organizing mutual responsiveness.

9.2 Challenges for Public Engagement

Broadly speaking, current public engagement exercises are beset by a twofold problematic. On the one hand, there is a lack of representativeness. It is well known that public engagement activities mostly draw on a limited number and diversity of citizen participants (te Molder and Gutteling 2003; Einsiedel 2008; Rowe and Frewer 2005). A less tangible but perhaps more influential aspect of the representativeness problem are the ways in which 'the public' is defined in and through public participation, i.e. what and whom it is designed to represent or not to represent (Felt et al. 2009; Wynne 2006). This is not only a matter of following rules and roles: citizen participants—and experts for that matter (Padmos et al. 2006)—actively contribute to what being 'a citizen' or 'lay person' consists of by making relevant some identity features and resisting others. Felt and Fochler (2010) show how citizens in a number of public engagements events in Austria draw on images of the general public as disinterested and reluctant to accept new technologies. More generally, public engagement activities seem to reproduce rather than challenge traditional hierarchical relations between experts and citizens. For example, technical expertise is often privileged over experiential knowledge, by experts as well as by 'lay persons' (Kerr et al. 2007). Citizen participants also rarely construct themselves as holding counter expertise (Felt and Fochler 2010).

While expertise may seem democratized at first sight, this is much less the case when looked at from a more empirical point of view. This brings us to a second, related problem: that of the restricted topical agenda (Veen et al. 2012). Apart from the exclusion of citizens or particular roles for citizens, the participatory agenda is managed in such a way as to deal with certain topics and not with others. Wynne (2001, 2006; cf. Mcnagthen et al. 2005) argues that there are relics from the 'deficit model' in most participatory procedures. According to this model, the public understanding of science is central to the interaction between science and public, and citizens would be ignorant in that respect. Alleged acceptance problems are managed on the basis of requiring more knowledge or trust from the publics—thereby drawing attention away from wider social and political questions these publics pose about for example ownership, control, and responsibility.

It is the second problem in particular that this chapter wants to follow up on. Where Swierstra and te Molder (2012) focus on the question how soft impacts of technology get dismissed or ignored, for what purposes (consciously or not) and with what consequences, this chapter looks at how stakeholders can be made aware of emerging soft concerns and how, on the basis of this recognition, such concerns can possibly be taken into account. While Wynne and others recognize the dismissal of 'other than risk' issues in public engagement exercises, they tend to focus on one side of the soft concern spectrum, that is, political concerns. These are political in the sense that they question how "existing power structures and institutional cultures around (…) innovation, development, regulation and exploitation" handle risk and uncertainty (Wynne 2001).

Moreover, there is little attention for how these concerns are put to use in real-life situations, especially with regard to people's claimed epistemic and action-oriented responsibilities (see also below). This is not only important in the light of improving our understanding of the actual dynamics of soft and hard impacts, but also because the different uses have different implications for how to achieve a more comprehensive public dialogue. That the demand for natural food is both constructed as a private consumer concern and black-boxed as requiring no further investigation, makes it a different problem than, for example, the issue of taste (Swierstra and te Molder 2012). Food experts tend to assign full responsibility to consumers for telling them what tastes good, and only claim epistemic access to the technicalities of how to achieve that flavor. Since anything can be made—if people want it to be made—there would be no such thing as a lack of good taste. Yet, even when we commonly refer to (good) taste, we may not always refer to the technical property of a thing, or to sense perception as such; we may also hint at a larger web of social and material relations. Taste can be seen as the expression of a cognitive and cultivating process (Teil and Hennion 2004).

Both naturalness and good taste are swept from the public discussion table but it is done in different ways, and therefore the starting point for a more comprehensive dialogue would be different. The latter is often constructed as a private but legitimate concern whereas naturalness first needs to be understood and taken seriously before it can be unpacked. So knowing what the soft concerns are is not enough. It is crucial to understand how they are put to use by the different stakeholders.

Against the background of these considerations we developed what we will label as DATEI—Discursive Awareness and Techno-Ethical Imagination: a design tool for facilitators who seek to organize stakeholder dialogues that are more hospitable to moral, political and cultural concerns around technological innovations. We explain the purpose and rationale of this interdisciplinary tool, and describe what it can do for stakeholder dialogues on innovations in food technology. While it proves to be a useful tool in challenging participants to explore common ground and understanding of soft concerns, and explicitly addresses questions of responsibility with regard to those concerns, it also reveals specific obstacles to fruitful public discussion of moral, political and cultural concerns in the context of technological innovation. In this sense, this chapter is not another success story but rather a critical evaluation of a promising tool that requires further development and improvement.

9.3 Discursive Awareness and Techno-Ethical Imagination

DATEI is an interdisciplinary tool that advances the inclusion of a wider range of ethical, cultural and political concerns in stakeholder dialogues on technological innovation in order to promote the discussion of responsibilities regarding those concerns. Dialogue facilitators may indeed recognize the importance of allowing room for soft concerns, but simply putting some of these concerns on the table may not be that effective when participants are reluctant to perceive these as legitimate concerns in the first place.

DATEI is built on the premise that, once certain concerns are raised, participants need to openly discuss rights and responsibilities for what they (claim to) know about such concerns before they can meaningfully express any expectations to each other for what to do about them. Without some agreed sense of knowledge

Table 9.1 Attribution of several kinds of rights and responsibilities (examples)

	Epistemic	Action-oriented
Rights	"You are entitled to claim this knowledge"	"You are entitled to act in this way"
Responsibilities	"You are expected to know about this"	"You are supposed to act in this way"

distribution among participants (e.g. about who is justified in claiming to know what naturalness is about), certain concerns will fail to become a legitimate issue on the agenda, as some participants will not recognize those concerns as topics to be discussed in a cogent and rational way. People can hold each other accountable for actions, but also for ideas; if no agreement exists on who is to be held accountable for giving meaning and interpretation to the concern at hand, it becomes a dialogue of the deaf.

For purposes of clarification, we distinguish two main senses of rights and responsibilities: epistemic and action-oriented (see Table 9.1). Epistemic rights and responsibilities refer to the kind that participants attribute to themselves and each other regarding what they and others can or should know. We are interested in how these rights and responsibilities are implicated in participants' organized practices of speaking, that is, how they are directly and indirectly attributed, refuted or agreed upon (cf. Heritage and Raymond 2005). For the concept of action-oriented responsibility, we follow van de Poel et al. (2012) in distinguishing backward-looking from forward-looking responsibilities. Roughly put, whereas the first covers things like blame, accountability, and liability for past wrongs, the latter refers to efficacy, obligation and due care with regard to what remains to be done or taken care of in the future.

For dialogues on new and emerging innovations in food technology, we believe that responsiveness requires the discussion of epistemic responsibilities and action-oriented, forward-looking responsibilities, since both touch upon the question what parties can reasonably expect of each other. Claiming not to have access to some knowledge or experience, for example, is not the same as refusing the responsibility to act but it allows for doing so, just as epistemic access to a particular issue is needed as to make oneself responsible for action regarding that issue. Superior epistemic access, on the other hand, may preclude others from (also) becoming responsible. Epistemic rights and responsibilities are bound up with the rights or responsibilities to act, and in a more general manner, with one's identity or entitlement to speak (Heritage and Raymond 2005; te Molder 2012).

DATEI is an interdisciplinary tool, in the sense that it integrates insights from conversation analysis in the field of discursive psychology (Edwards and Potter 1992; Potter 1996; Lamerichs and te Molder 2011) on the one hand, and critical theory and pragmatism in the field of philosophy (Foucault 1970; Swierstra 2002) on the other. Where the latter identifies and interprets issues that are structurally marginalized from public discourse, the first stimulates actors themselves in recognizing interactional patterns that obstruct the discussion of such issues in actual conversations. DATEI is based on the idea that the perceived legitimacy of agenda items is partially shaped by these interactional patterns.

9.3.1 The First Element of DATEI: The Discursive Action Method

DATEI consists of two elements. The first element is the Discursive Action Method or DAM (Lamerichs and te Molder 2011). The Discursive Action Method has its roots in Discursive Psychology; an approach to discourse that explores both the sequential and rhetorical features of talk and text (Edwards and Potter 1992; Potter 1996; Edwards 1997). Participants in interaction are understood as performing actions with their talk, such as managing responsibility or blame. Descriptions, here, are not treated as reflections of reality, but as tools for achieving particular interactional goals. For instance, a description such as ‘it is cold in here’ functions as a request when the recipient shows he treats it as such—for example by closing the window. Discourse is approached as being both constructed and constructive of reality, rather than a reflection of reality. In addition to the sequential analysis of talk (how is the talk treated by other participants), analyses are made of how versions of reality are produced to counter rhetorical alternatives (what alternative description is undermined by producing this one?). The point of departure here is that participants have a range of potential descriptions at their disposal and (consciously or not) select the description that is suitable for achieving certain goals in comparison to other formulations.

The Discursive Action Method was developed by discursive psychologists (see Lamerichs and te Molder 2011) as a tool for professionals to analyze the everyday talk of their target groups in a number of steps, preferably done together with these groups. The first step is to take on a non-normative view so as to discourage participants from criticizing the content of the talk, and first behave like distant observers of their own and other people’s conversations. The second step is to make participants aware of the interactional effects of people’s talk and the practices that are used to achieve these effects (consciously or not). In doing so, the focus is not on individual cognitions (intentions, motives, attitudes) but on understanding how the talk is treated by others (as blame, compliment et cetera). Once the effects and strategies have been established, the non-normative style of looking at interaction is left and one is asked to judge the desirability of the used strategies and their effects. Depending on the goals and motivations of DAM practitioners, additional steps exist of designing and executing interventions.

In our meetings, DAM familiarizes participants with recurring interactional patterns that hinder open communication between stakeholders. Our aim is to make participants critically aware of these interactional patterns that normally remain implicit, and to improve their own practices in the course of further conversation. To achieve this, we present them with selected fragments (transcripts) of expert and citizen-consumer interactions. These fragments highlight the various epistemic claims made by discussants on food innovation. As pointed out earlier, epistemic claims describe the rights and responsibilities regarding what one can(not) know and vouch for, and such claims can be made implicitly or explicitly. Also, it encourages participants to spell out the unfruitful consequences of such implicated claims for the development of dialogue: e.g. participants identify how some discussants, unknowingly, disqualify others as not being able to make sound judgments on the topic, or how other discussants implicitly discharge themselves from having to account for whatever they claim. Such critical analysis of speech, commonly carried out together with the participants, is a necessary step if one wants to break with such interactional patterns in future discussions.

9.3.2 The Second Element of DATEI: Techno-Ethical Imagination

The second element, Techno-Ethical Imagination, invites participants to consider potential ethical, cultural or political impacts that may occur once new or emerging technologies become embedded in society. Participants are presented with a scenario in which a future application of the technology in case is met by resistance: it describes what could happen to technological innovations when certain norms, values or meanings that users hold dear (i.e. ethical, cultural and political impacts) are not properly taken into account in the process of innovation. In the case of food technology, the ultimate consequence is a lack of public trust, even when health risks are not the issue. In the end, Techno-Ethical Imagination challenges the participants to reflect and possibly agree on a common distribution of forward-looking responsibilities.

This scenario has a threefold function. First, it helps those dismissing soft concerns to realize that these concerns can be entangled with what they already recognize as serious consequences; e.g. food technologies that promise to benefit public health fail to do so if critical consumers do not see them fit in with their conception of good food, or the good life, for that matter.

Second, it invites participants to collectively identify what exactly could be the problem (if they see any) in the scenario's storyline, thereby encouraging them to articulate the various norms, values and meanings that could be promoted, countered, or reshaped by the technology. Here's where the importance of DAM comes in: having analyzed earlier conversations between experts and citizen-consumers in terms of interactional patterns, participants should now be able to critically and openly assess implicit claims to epistemic authority in the course of their own conversation. It allows them to discuss and explore the meaning of e.g. naturalness, or the practice of buying, preparing and eating food as part of the good life. Rather than 'parking' such concerns, for example by claiming superior definitions without creating grounds for exploring these, participants are now encouraged to explore and articulate them through dialogue.

Third, the technoethical scenario invites participants to imagine a concrete future in which a technological innovation finds its way to social practices and public discourse. The scenario invokes the discussion of alternative innovation paths, and enables participants to mutually express expectations of each other and, ideally, to arrive at a temporary agreement on the distribution of action-oriented responsibilities.

9.4 The Workshop Setting

In order to evaluate our experiences with DATEI in actual stakeholder dialogues, we will discuss two dialogues on new and emerging food technologies that we organized in the Netherlands, in Spring 2012. In each of these dialogues, we explicitly thematized natural food as a topic for further discussion. Naturalness is an exemplary issue in recent food discussions for investigating obstacles, since the citizen-consumer's wish for natural food is a strong and ongoing trend (Rozin et al. 2004; Siegrist 2008; Kampffmeyer 2012; Dornblaser 2013), while many food experts and policy makers seem to consider it a puzzling if not annoying issue. Though rich of moral, cultural and political meanings, naturalness seems to be a term that one hardly gets any grip on. How to discuss naturalness? For current and future generations of food professionals, as well as groups of consumers and citizens, this may well be a vital question.

In the two meetings, we brought together a total of 13 food experts, professionals and citizen-consumers, who take an interest in new and emerging food technologies. For a list of the participants' affiliations, see Table 9.2.

We formulated three targets we aimed to meet with the participants in these meetings. In Table 9.3, the targets are presented together with the DATEI element that served to meet the specific target.

In each of the meetings, the participants were challenged to collectively identify the effects of specific utterances by providing them with two fragments of interaction about novel foods, derived from group discussions with consumers and stakeholders. In the first, an expert implicitly claims a superior definition of

Table 9.2 Overview of participating stakeholders

Stakeholder group	Organisation
Citizen-consumers	Dutch Association for Housewives
	Youth Food Movement Netherlands
Government	Dutch Ministry of Health, Welfare and Sport
	Netherlands National Institute for Public Health and the Environment
Semi-government	Netherland Nutrition Centre
Industry	Danone Netherlands
	CSM Bakery Supplies Europe
	Unilever R&D Netherlands
Academia	NIZO food research
	Wageningen University, Food Chemistry

Table 9.3 Meeting targets

Meeting target	DATEI element
1. Gaining insight in the present distribution of epistemic rights and responsibilities in relation to food innovation	DAM: analysis of interaction on naturalness by expert and consumers
2. Discussing the consequences of the present distribution of epistemic rights and responsibilities for action-oriented rights and responsibilities	DAM and Techno-Ethical Imagination: application of DAM insights to scenario discussion
3. Creating shared visions for what would be a desirable distribution of action-oriented rights and responsibilities	Techno-ethical imagination: rewriting the scenario

naturalness, thereby disqualifying the knowledge of consumers of this concept (see also Swierstra and te Molder 2012). In the second fragment, consumers claim objective knowledge of naturalness, again in such a way that it left no room nor created the urgency for further dialogue. The fragments are presented in Table 9.4.

Table 9.4 Fragments from earlier group discussions, as presented to the participants

Fragment *industrial expert on naturalness*	Fragment *consumers on naturalness*
Facilitator: 1. but the picture that emerges now is of uh 2. as it were 3. an uhh (0.4) somewhat powerless industry 4. that has to dance 5. to the contradictory whims 6. of the consumer (0.6) 7. uh is that the current feeling 8. or are there also ideas 9. about naturalness 10. with the industry itself? 11. (*2 lines omitted*) **Expert:** 12. yes I think the industry **Facilitator:** 13. okay [name expert] **Expert:** 14. I think the industry 15. views it a little bit - 16. a little bit differently (0.4) 17. uhh there are indeed (0.8) 18. consumers who indeed 19. want *natural* 20. without probably 21. many consumers 22. uhh understanding what that then means 23. and what it entails	**Consumer 1:** 1. fish oil is another type of substance of course uh 2. fish oil is something of which you know it is 3. healthy, say and **Consumer 2:** 4. that is proven **Consumer 1:** 5. that reducing stress, you can do that in a different way **Consumer 2:** 6. yes **Consumer 1:** 7. you can do that in a much healthier way, say, 8. reducing stress **Consumer 3:** 9. yes **Consumer 1:** 10. by indeed you say yoga, relaxation exercises, 11. and if if you have too much stress **Consumer 2:** 12. hmm **Consumer 1:** 13. there are other ways indeed 14. that do go in a natural way indeed and **Consumer 2:** 15. yes **Consumer 1:** 16. and and that idea is not with fish oil, say, 17. that can actually be a substance 18. that is something natural of course, 19. fish oil

We guided the group discussions about the fragments by providing the participants with a number of questions, such as: what does this person suggest or claim to know about naturalness? What effect could this have on the course of the conversation? Is an alternative way of speaking possible and/or desirable?

Then, the participants discussed a techno-ethical scenario in which the future of probiotic food products is imagined. The narrative starts from the assumption that the European Food Safety Authority has admitted an increasing number of health claims on the reduction of several health risks by probiotic foods. The innovation, however, fails to become a success and yields distrust as the proclaimed naturalness of the products is contested by critical consumers and citizens. The scenario envisioned how public controversy can escalate when concerns about naturalness are not given a fair hearing. The participants were asked to discuss the nature of the problem in this narrative, the possible reasons why to take these concerns seriously and why not to. Also, they were invited to explore the meaning and value of natural food. In the end, we asked participants to divide in teams and rewrite the scenario by proposing a new distribution of action-oriented, forward-looking responsibilities with regard to natural food.

9.5 Results

9.5.1 Achievements

The group discussions of the interaction fragments (DAM), the techno-ethical scenario and the rewriting assignment demonstrated that DATEI generated many results in the construction of a stakeholder dialogue on naturalness. Below we have listed some of the most relevant achievements, illustrated by some of the statements and phrases that participants made during these discussions.

Gaining Insights into Interactional Effects
As described, in the first part of the workshop we applied the Discursive Action Method to analyze two exemplary fragments of food expert talk (Swierstra and te Molder 2012) and citizen-consumer interactions (Sneijder and te Molder 2014) on food innovations, together with the participants. The observations the participants made when analyzing these fragments show that they did gain insight into the interactional effects of both types of talk.

They made clear statements on how the food expert from the industry 'shelved' the issue of naturalness by (indirectly) claiming a superior definition and how consumers did the same by talking about naturalness in absolute terms, thereby inviting confirmation rather than exploration of what naturalness possibly entails. The next examples show how participants interpreted the statements of the quoted expert and their effects.

In the first fragment, we see how a food engineer displays his interpretation of the effect of the experts' strategy, namely that the expert disqualifies the consumer as a conversational partner by suggesting that he knows better.

Fragment 1, Workshop 2

Food engineer, Industry:

I think he also disqualifies one of the parties to the conversation, namely that of the consumer. [quoting the industrial expert:] "They don't know what they are talking about."

I think that's one of the reasons that the conversation bleeds to death, actually.

In the second fragment a policy advisor describes the arrogant impression the expert makes and extends this to the actual practice of industry to put 'natural' on packages just because the consumer is assumed to want this.

Fragment 2, Workshop 2

Policy advisor, Government:

That is a bit of "well yes okay then we put 'natural' on the package we make", say, and the idea is actually "yes, we all know that they do not know,

or have no idea of what natural is. But they want that and okay we put it on." That's the impression I get.

9.5.2 Problem Recognition

When discussing the techno-ethical scenario, most participants recognized that there will be problems for multiple stakeholders if naturalness is turned into a private problem and consequently does not enter the public agenda.

Fragment 3, Workshop 1

Food engineer, Academia:

Well I do think there is a problem but the problem is that people draw a definition to themselves without communicating it.

Fragment 4, Workshop 2

Policy advisor, Government:

It looks like the food industry has taken ownership of concepts you cannot take ownership of. And if they had just stayed with "this is probiotics and that is the claim", the problem would not have existed.

The problem occurs because they use concepts that are dear to people, in the wrong way, say, and people stand up against that.

Participants gradually agreed on the observation that if the industry does not handle the idea of naturalness with care, a health-enhancing innovation such as probiotics may die in the end (see Fragment 5). A promising innovation, then, is blocked for what most of the participants formulated as the "wrong" reasons. While some claimed that it would be hard to discuss naturalness in a meaningful way, they all agreed that a total absence of naturalness on the public agenda could not be a solution either.

Fragment 5, Workshop 2

Food engineer, Industry:

Imagine it says that product is really capable of reducing the risk of colon cancer. I think then a lot of consumers would have liked to know that, for sure. They want it that badly. And now that has disappeared on other grounds. And that is a shame in the end, for the producers as well as for the consumer.

Exploring the meaning of naturalness

In the discussion participants tentatively proposed their views on naturalness and its meanings. Although they soon agreed that no exact border between natural and unnatural can be drawn, they mentioned several conditions that could play a role in the assessment of naturalness. Participants mentioned that natural food is familiar, recognizable, transparent, and as little processed as possible, while they also acknowledged that processed food can be labeled as natural in some other respects as well. As a food engineer explained, food can be processed and to some extent natural if the processing does not differ too much from preparing meals in your average household kitchen. He noted that this was actually one of the challenges he faced as a food engineer.

Furthermore, the preference for natural food was framed as part of one's identity and lifestyle. Several citizen-consumers explained, for example, that home cooking, do-it-yourself, and selecting fresh vegetables at the farmer's market actually makes them happy: that's what natural food is about. It is part of what they consider to be the good life.

Also, one participant mentioned that naturalness is a value typically cherished in urbanized societies in which the distance between food production and consumption is rather long. While the participants did not reach at a definite conclusion on the meaning of naturalness, at least they started to explore and reflect on the various meanings of the concept, implicitly acknowledging that such exercise is worthwhile.

Talking responsibilities
In the third part of the workshop participants rewrote the scenario. The aim in this part of the workshop was not that they provided the solution to the problems that were unfolded in the scenario, but rather that the participants were willing to take naturalness—and the distribution of action-oriented responsibilities for this issue—seriously.

Participants ascribed responsibility to the industry and suggested that this stakeholder should be more pro-active in the future, that is, listen and, particularly, talk to consumers. Another suggestion was that the national government should set an ultimatum for stakeholders in society: if they do not come to an agreement on naturalness, a commission will be constructed to investigate if and how natural-ness can be regulated as a label. One may question the desirability and viability of these suggestions, but they clearly show that academic and industrial food engineers, citizen-consumers and policy advisors started to deliberate and negotiate about a common distribution of action-oriented responsibilities on future food innovations.

9.5.3 Obstacles

Aside from achievements, we also found a number of obstacles that stood in the way of a fruitful stakeholder dialogue on naturalness. This is an important result, as it shows us what issues should be tackled for further enhancing this dialogue. In future applications of DATEI we will adjust the method so as to prevent or overcome these obstacles.

Sit and wait for objectivity
Participants repeatedly expressed the desire for a clear and unequivocal definition of naturalness, and consensus on such definition, before they were willing to start the discussion of options and responsibilities. At the same time, they showed a deep skepticism on the possibility of reaching any consensus on a definition. Also, participants pointed out that naturalness cannot be measured or quantified, and implied that, for this reason, naturalness cannot be meaningfully discussed any further. Arguably, a perceived lack of "objectivity" in the concept of naturalness seemed to hold back the discussants to freely engage in deliberation. As described above, eventually they did explore several meanings of the concept, and even seemed to reach at least some mutual understanding about it, but recurring skepticism formed a serious obstacle to that effort.

Delegation to the market

Applications of food technology are not only the fruits of science and technology, but also commodities on the market place. Participants expressed their worries about the significance of such dialogue if, after all, anyone is free to buy or not to buy the discussed food products (Fragment 6). Also, experts displayed some degree of indifference as naturalness could simply appear as a meaningless but harmless marketing catch phrase.

Fragment 6, Workshop 2

Food engineer, Industry:

The beautiful thing of the consumer is: then he won't buy it.

Citizen-consumers seemed reluctant to further elaborate on the concept of naturalness. They privatized their views and preferences as just another individual and personal opinion (Fragment 7). Interestingly, this resembles the way the subjects in the DAM fragments discussed naturalness, as it privatizes the concept and does not stimulate exploration or discussion.

Fragment 7, Workshop 1

Citizen-consumer:

I rather think that you should just look at "what do I find natural?", "what do I find important with respect to that?" If one person says "for me natural is the least processed, so that I can see who the producer is and that it is biological", then you look at the words or logos on that product and that is natural for you then.

Thus, the free market economy was repeatedly brought up as an effective mechanism that makes further conversation no longer necessary.

Naturalness as an exchangeable argument

Some of the participating engineers raised the idea that naturalness is just another tool for the consumer to 'argue' with the industry. They suggested that if consumers cannot raise the concern of naturalness, eventually they will search for other issues to claim grounds for complaining (Fragment 8).

Fragment 8, Workshop 1

Food engineer, Industry:

So if you don't use [naturalness as a concept] anymore, it is clear. But then they will come up with another concept within half a year.

Food engineer, Industry:
Sure, then they’ll come up with something else.
Food engineer, Academia:
For which everyone has his own picture in mind.

Comments like these make clear that in matters of food technology, not only certain products or producers are distrusted but also citizen-consumers can suffer from a bad image. As we noted before, some degree of mutual trust can be the result of public engagement exercises but is, at the same time, a precondition for meaningful dialogue. Regardless whether this rejection of naturalness as an exchangeable argument is justified, the fact that the integrity and veracity of such concerns were openly questioned, hindered actual discussion of the substance of naturalness.

9.6 Discussion and Conclusion

Responsible Innovation is about enhancing the mutual responsiveness between stakeholders. This implies that all stakeholders’ concerns are taken seriously and are given a fair hearing. But the discussed stakeholder dialogues show that to reach this goal it is not sufficient to simply arrange meetings between stakeholders. Even if everyone gathered round the table desires to engage in an open and productive dialogue, certain discursive and psychological mechanisms constitute obstacles to a productive and truly responsive engagement with each other’s concerns. Some concerns are taken seriously; others get dismissed as ‘private’ and ‘subjective’, and thus not meriting public deliberation.

Our interactive tool DATEI is designed to make these obstacles visible and to enable dialogue partners to work around them. First the stakeholders are made aware, using examples taken from similar focus group exercises, of discursive mechanisms that frustrate a key presupposition of mutual responsiveness: the willingness to share responsibility for knowledge claims, that is: to accept such epistemic responsibility one self, and to grant it to other participants. In a second step, the action-oriented (rather than the epistemic) dimension of this sharing of responsibility is foregrounded. Where the first step seeks to enhance stakeholder’s willingness to share the epistemic responsibility of ‘defining what is the case’, the second step asks them to solve a practical problem together—that is: to share the moral responsibility for what has to be done.

Epistemic responsibility constitutes a precondition for action-oriented responsibility, as practical solutions can only be devised on the basis of shared definitions of what is the case. To enhance the constructive problem-solving mode of this type of dialogue, DATEI steers away from any blame-game by focusing on the future, thus appealing to everyone's action-oriented responsibility. It achieves this by offering future scenarios. Because of the narrative form of the scenario, it is left to the participants to collectively identify what the problem is. Furthermore, because of the way the scenario is structured, they are more or less forced to discuss together whether some specific soft concerns (in this case about 'naturalness') are part of that collective problem or not.

So, how successful was DATEI in creating mutual responsiveness between the stakeholders by inviting them to accept both their epistemic and action-oriented responsibility? Here we have to differentiate between DATEI's two constitutive moments: highlighting the participants' epistemic responsibilities, and their action-oriented responsibilities. The outcome of the DAM part gave us cause for hoping that DATEI would actually succeed in enhancing mutual responsiveness. Confronted with examples of 'negotiations' of epistemic responsibility, participants proved to be able to recognize these mechanisms and acknowledged that they were obstacles to an open dialogue. More specifically, they were both able to point out that privatization of issues was not conducive to mutual responsiveness, and that pausing to ask what the other meant by 'naturalness' was rather constructive.

Unfortunately, the results from the scenario exercise, focused on sharing first epistemic and then subsequently action-oriented responsibility, were less promising. The good news is that the participants refrained from using some of the discursive strategies they had just been made aware of: that is to say, no one monopolized epistemic authority. Unfortunately, rather than accepting the burden of sharing epistemic responsibility of defining naturalness, the participants opted to shift this responsibility to vague, distant and future 'others'. In stark contrast to their earlier resolution not to dismiss the opponent's use of 'naturalness' beforehand but to open up that black box by inquiring into what the other meant by the concept, now they withdrew to a position that such an inquiry was doomed to failure because naturalness had no fixed meaning anyway. As the meaning was subjective and changeable anyway, everyone was entitled to their own opinion. With regard to sharing action-oriented responsibility for creating good food, even though they had previously identified privatization as a strategy incompatible with mutual responsiveness, now—faced with a problem that called for collective action—all agreed that the decision whether or not our food should be more natural or not, should be privatized. To our surprise, all participants agreed that anonymous market mechanisms rather than collective deliberation, consumers rather citizens, should decide what 'good' food was, and whether that included 'naturalness' or not.

So, the result of our focus groups was mixed. On the one hand we can say that it is possible to make dialogue partners aware of the discursive patterns that frustrate mutual responsiveness through the DAM method. On the other hand, it proves to be much more difficult to make them transport these insights to the more practical arena of discussing action-oriented responsibilities, i.e.: agreeing on shared definitions on which collective decisions are based about what course of action to follow when faced with soft concerns. There we find the participants quickly redrawing on the privatisation strategy, thus avoiding the need to be mutually responsive to each other.

How to explain this? And how can our method be improved so as to really enhance mutual responsiveness in the case of soft concerns like 'naturalness'? Part of the explanation probably has to do with the artificiality of the focus group setting, where there is a large bonus on being polite to the other participants, and little necessity to really come to terms with one another, as there is no necessity to decide on a shared course of action. However, we also see the focus group reproducing the coping strategy that society at large has developed to deal with soft concerns: privatization and delegation to the market.

Framing the issue of 'natural' food as a personal consumer preference may seem a benefit to dialogue participants; they do not have to come up with moral judgements of the behaviour of others and consequently cannot be accused of moralism or paternalism. The market is a coordination mechanism that needs little or no deliberate coordination, so it seems there is no need to justify one's preferences and choices to other stakeholders. However, a cost connected to this framing is that some values that may be part of one's conception of naturalness, e.g. a less instrumental relation to animals, or a more sustainable food system, are only imperfectly realized, as groups of consumers will simply follow their purses when purchasing food. A similar cost is involved in privatizing 'good taste' when it actually entails the promotion of a rich and diverse food culture.

In other words, the a priori privatization of concerns about good food, i.e. before they have been subject to serious public deliberation at all, is unproductive when doing something about it actually requires some degree of social coordination. Granting each and any individual the right to know for himself what naturalness is about and act upon it, still leaves everything open with regard to responsibilities, epistemic or action oriented. In our liberal society, it seems, the innovation of good food is up to the private consumer—probably based on the assumption that in tackling soft concerns, paternal state interference would be the only alternative.

How to improve this? What are the next methodological steps of DATEI? The key is to change the setup of the discussion in such a way that the privatization strategy becomes part of the discussion, rather than substituting for it. This can be done in two ways. First, the pros and cons of privatization, including the delegation to individual consumer choice and the free market, could be more thoroughly explored by the participants at forehand. Second, the scenario should invite the participants to come up with a shared course of action that can be justified to all stakeholders. This can be done by creating a scenario that envisions the stakeholders as members of a 'community of fate' (van Gunsteren 1998). Such a community exists when the choices of others affect my situation, and vice versa. In such a situation we cannot simply allow individual freedom to reign supreme, but we then have to coordinate our choices and actions to produce a course of action that is acceptable to all. We hope that by introducing these two new elements into DATEI, the method will be more successful in enhancing the mutual responsiveness between stakeholders.

References

Dornblaser, L. 2013. The changing face of natural foods. *Food Technology*, March 2013.

Edwards, D. 1997. *Discourse and cognition*. London: Sage.

Edwards, D., and J. Potter. 1992. *Discursive psychology*. London: Sage.

Einsiedel, E. 2008. Public participation and dialogue. In *Handbook of public communication of science and technology*, ed. M. Bucchi, and B. Trench, 173–184. Abingdon: Routledge.

Felt, U., and M. Fochler. 2010. Machineries for making publics: inscribing and describing publics in public engagement. *Minerva* 48(3): 219–238.

Felt, U., M. Fochler, A. Müller, and M. Strassnig. 2009. Unruly ethics: on the difficulties of a bottom-up approach to ethics in the field of genomics. *Public Understanding of Science* 18(3): 354–371.

Foucault, M. 1970. The order of discourse. Inaugural lecture at the College de France. In *Untying the text: A Post-Structuralist Reader*, ed. R. Young (ed.) (1981), London: Routledge, 51–78.

Haen, D. 2014. The paradox of E-numbers: ethical, aesthetic, and cultural concerns in the Dutch discourse on food additives. *Journal of Agricultural and Environmental Ethics* 27: 27–42.

Heritage, J., and G. Raymond. 2005. The terms of agreement: indexing epistemic authority and subordination in talk-in-interaction. *SocPsychol Q* 68(1): 15–38.

Kampffmeyer Food Innovation. 2012. Report 02, How to make clean label. http://kfi.kampffmeyer.com/cleanlabelreport/en/index.html. Accessed 28 January 2013.

Kerr, A., S. Cunningham-Burley, and R. Tutton. 2007. Shifting subject positions. Expert and lay people in public dialogue. *Social Studies of Science* 37(3): 385–411.

Lamerichs, J. and te Molder, H. 2011. Reflecting on your own talk: the Discursive Action Method at work. In *Applied conversation analysis. Intervention and Change in Institutional Talk*, ed. C. Antaki, 184–206. Basingstoke: Palgrave Macmillan.

Macnaghten, P., M. Kearnes, and B. Wynne. 2005. Nanotechnology, governance, and public deliberation: what role for the social sciences? *Science Communication* 27(2): 268–291.

Marris, C., Wynne, B., Simmons, P. and Weldon, S. 2002. *Public perceptions of agricultural biotechnologies in Europe*. Final Report of the PABE research project funded by the Commission of European Communities.

Padmos, H., H. Mazeland, and H. te Molder. 2006. On doing being personal: citizen talk as an identity-suspending device in public debates on GMOs. In *Analysing citizenship talk: Social positioning in political and legal decision-making procedures*, ed. H. Hausendorf, and A. Bora, 276–295. Amsterdam: John Benjamins.

Pellizzoni, L. 2004. Responsibility and environmental governance. *Environmental Politics* 13: 541–565.

Potter, J. 1996. *Representing reality. Discourse, rhetoric and social construction*. London: Sage.

Rowe, G., and L. Frewer. 2005. A typology of public engagement mechanisms. *Science, Technology and Human Values* 30(2): 251–290.

Rozin, P., M. Spranca, Z. Krieger, R. Neuhaus, D. Surillo, A. Swerdlin, and K. Wood. 2004. Preference for natural: instrumental and ideational/moral motivations, and the contrast between foods and medicines. *Appetite* 43: 147–154.

Siegrist, M. 2008. Factors influencing public acceptance of innovative food technologies and products. *Trends in Food Science and Technology* 19(11): 603–608.

Sneijder, P. and te Molder, H. 2014. 'I don't believe in light mayonnaise': Epistemics-in-action in consumer/citizen talk on health claims. (Manuscript submitted for publication).

Stilgoe, J., R. Owen, and P. Macnaghten. 2013. Developing a framework for responsible innovation. *Research Policy* 42(9): 1568–1580.

Swierstra, T. 2002. Moral vocabularies and public debate: the cases of cloning and new reproductive technologies. In *Pragmatist ethics for a technological culture*, ed. T. Swierstra, J. Keulartz, M. Korthals, and M. Schermer, 223–240. Deventer: Kluwer Academic Publishers.

Swierstra, T., and te Molder, H. 2012. Risk and soft impacts. In *Handbook of risk theory. Epistemology, decision theory, ethics, and social implications of risk*, ed. S. Roeser, R. Hillerbrand, M. Peterson, and P. Sandin, 1050–1066. Dordrecht: Springer.

te Molder, H. 2012. Discourse communities as catalysts for science and technology communication. In *Performing public participation in science and environment communication*, ed. L. Phillips, A. Carvalho, and J. Doyle, 97–118. Bristol/Chicago: Intellect/The University of Chicago Press.

te Molder, H., and Gutteling, J. 2003. The issue of food genomics: about uncaring citizens and united experts. In *Genes for your food—Food for your genes. Societal issues and dilemmas in food genomics*, ed. R. van Est, L. Hanssen, and O. Crapel, 117–136. Working document 92. The Hague: Rathenau Institute.

Teil, G., and A. Hennion. 2004. Discovering quality or performing taste? A sociology of the amateur. In *Qualities of food*, ed. M. Harvey, A. McMeekin, and A. Warde. Manchester/New York: Manchester University Press.

van de Poel, I., J. Nihlén Fahlquist, N. Doorn, S. Zwart, and L. Royakkers. 2012. The problem of many hands: climate change as an example. *Science and Engineering Ethics* 18(1): 49–67.

van Gunsteren, H. 1998. *A theory of citizenship. Organising plurality in contemporary democracies*. Boulder, CO: Westview Press.

Veen, M., H. te Molder, B. Gremmen, and C. Van Woerkum. 2012. Competing agendas in upstream engagement meetings between celiac disease experts and patients. *Science Communication* 34(4): 460–486.

Von Schomberg, R. 2011. Prospects for Technology Assessment in a framework of responsible research and innovation. In *Technikfolgenabschätzen lehren: Bildungspotenziale transdisziplinärer Methoden*, ed. M. Dusseldorp, and R. Beecroft. Wiesbaden: VsVerlag.

Wynne, B. 2001. Creating public alienation: expert cultures of risk and ethics on GMOs. *Science as Culture* 10(4): 446–481.

Wynne, B. 2006. Public engagement as a means of restoring public trust in science—hitting the notes, but missing the music? *Community Genetics* 9: 211–220.

Chapter 10
Responsible Innovation in Energy Projects: Values in the Design of Technologies, Institutions and Stakeholder Interactions

Aad Correljé, Eefje Cuppen, Marloes Dignum, Udo Pesch and Behnam Taebi

Abstract Projects that deal with unconventional ways to produce, store, or transport energy often give rise to resistance by local communities. The value-laden basis of such resistance is often ignored by decision makers. This chapter operationalizes the concept of Responsible Innovation by using and adapting the approach of value sensitive design. This approach holds that the variety of stakeholders' values might be taken as a point of departure for the (re)design of a technological system in such a way that divergent values can be accommodated. The scope of value sensitive design can be extended beyond the technology, however. Values are also embedded in the institutional context and in the processes of interaction between stakeholders. Hence, the prevention of controversies over conflicting values may be pursued by redesigning the institutional context, and by taking the dynamics of stakeholder interaction explicitly into account.

Keywords Value sensitive design · Energy projects · Stakeholder engagement · Values · Institutional design

A. Correljé (✉) · E. Cuppen · M. Dignum · U. Pesch · B. Taebi
Faculty of Technology Policy and Management,
University of Technology Delft, Delft, The Netherlands
e-mail: A.F.Correlje@tudelft.nl

E. Cuppen
e-mail: E.H.W.J.Cuppen@tudelft.nl

M. Dignum
e-mail: M.Dignum@tudelft.nl

U. Pesch
e-mail: U.Pesch@tudelft.nl

B. Taebi
Belfer Center for Science and International Affairs, Kennedy School of Government,
Harvard University, Cambridge, USA
e-mail: B.Taebi@tudelft.nl

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_10

10.1 Introduction

The supply of energy is a fundamental prerequisite for the functioning of society. Yet, traditionally, the supply of energy is associated with many problems. Today, three principal problems can be identified. Firstly, the use of fossil fuel causes air pollution which directly jeopardizes human health while the CO_2 emitted affects the earth's climate. Secondly, there is the perception that the resources that can be exploited easily and at low cost are being depleted rapidly, driving up the price of energy. The third problem is that the uneven regional distribution of energy resources is causing international geopolitical and economic frictions.

Such problems foster new initiatives and technological developments to produce, store or transport energy in currently unconventional ways. Schemes to solve these problems include the development of wind, solar, thermal and other renewable forms of energy. Furthermore, there are new developments of advanced methods for exploring and producing hydrocarbon fuels from the deep sea, tar sands, and geological layers that, until recently, were considered too expensive and risky, or for which technologies simply were lacking. A completely new phenomenon is the underground disposal of CO_2, in order to reduce the emissions into the atmosphere.

New energy initiatives repeatedly give rise to problems of societal acceptance because their implementation and operation have national or local repercussions. Often, in the development of new energy supply facilities new inconveniences are created, for instance by the siting of extraction, conversion and transport facilities, or by the economic, social and environmental impact of these activities upon the local population, or even society at large. As a consequence, controversies between local populations, governments, and industry are always imminent.

In the Netherlands, a small country in Europe, new ways to dispose of CO_2 and to store and produce natural gas caused huge controversy (Wolsink 2000; Persson 2012). Citizens in the municipality of Barendrecht revolted against the plans of the national government and Shell to feed CO_2 into a depleted gas field underneath this town (Feenstra et al. 2012). Apparently, it was not only the perceived risk, but also the lack of a serious consideration of people's concerns, both by the government and industry, that fuelled their resistance. This brought about a (temporary) cancellation of all onshore carbon capture and storage-projects in the Netherlands.

Another example concerns a partly depleted gas field near the municipality of Bergen, which was sold by BP to TAQA from Abu Dhabi to be used as an underground gas storage facility. Natural gas will be injected into the subsurface in the summer season and distributed among users in the Netherlands and North West Europe in the winter. Although not with the same vigor as in Barendrecht, citizens from Bergen together with local authorities and environmental organizations protested against these plans and, in the end unsuccessfully, challenged the legitimacy of the project in the Dutch Supreme Court.

A third example from the Netherlands involved the construction of an exploration facility to test the potential production of shale gas in the municipality of Boxtel. The British company Cuadrilla had obtained an exploration license from the state and a municipal permit to conduct drillings, but this permit was successfully challenged in court by local inhabitants, together with the Dutch Rabobank, which has one of its data centers in the vicinity.

These examples demonstrate that the interaction between citizens, businesses, local authorities and environmental organizations may be, or become, problematic, turning energy projects into difficult and risky enterprises. The government and the energy industry therefore consider the antagonists in such cases as showstoppers. Thus, policies and communication are organized accordingly and the value-laden basis of controversies is ignored. Hence national authorities and energy companies complain that either the public is ill-informed (Wynne 2001), resistant to scientific information, or only concerned with its own short-term interests (Bell et al. 2005; Wolsink 2006). This is referred to as the "technocratic pitfall" (Roeser 2011). Nevertheless, we do not suggest that simply adhering to the desire of local communities is the preferred solution. It would be myopic to assume that the complexity of values and interests related to such projects could be resolved this way; bringing in the "populist pitfall" (Roeser 2011).

The challenge is to avoid both pitfalls, by creating strategies and solutions that bridge the diversity in stakeholders' values. Different actors make different assessments of the costs and benefits of these projects. Gaining public acceptance not only requires more or better dissemination of information, or a more elaborate risk assessment. It also, as we argue, requires the acknowledgment of different (moral) viewpoints of stakeholders, which should be taken as a point of departure to identify and to construct shared solutions. A societally responsible development of energy projects requires the accommodation of the variety of stakeholders' values.

In this chapter, we will sketch the contours of a methodology that will open up the black box of this variety of values. In contrast to Blok and Lemmens (Chap. 2, this volume) and in line with Robaey and Simons (Chap. 5, this volume), we argue that this meta-insight is a fruitful entrance for creating a shared solution. Both historic examples and conceptual developments strengthen this viewpoint. We propose the application of *value sensitive design* (VSD) to such energy projects, in line with a product-based approach to responsible innovation as defined by Koops (Chap. 1, this volume). The VSD approach was originally developed to target the incorporation of a diverse range of values in information technology (Friedman and Kahn 2000). Our aim is to extend the scope of VSD, not only by relating it to other technologies, but also by applying it to the institutional context in which such new technologies are implemented and/or used. In this regard we are aligned with Blok and Lemmens' views on the need to take a broad view of innovation when issues of responsibility are at stake (Chap. 2, this volume).

Our focus is on projects that individually have a local impact, such as a wind park or a gas storage or production facility. However, in aggregation these projects have the potential to contribute greatly to the overall energy supply. Often, this potential is clearly voiced at the initiation of a single project. VSD aims to incorporate the values of all relevant stakeholders in the design process. For example, it includes the values that are articulated by local stakeholders regarding a specific project as well as possible large societal benefits or concerns.

Most approaches of VSD specifically focus on the design of technological artifacts or systems. However, in respect to energy projects, it is not only technology design that affects the divergent sets of values. It is also the design of the *institutional context* and of the *interactions between stakeholders* that may lead to a deepening, or conversely, a resolution of value conflicts. The institutional context, constituted by formal institutions, such as laws, standards, regulations, contracts, and informal institutions such as customs, traditions, and routines, embodies values that have important ramifications for the distribution of (perceived) burdens and gains of a specific project. Many of these institutions, especially the formal ones, can be redesigned in order to accommodate divergent values. Also the way in which the project itself is arranged and executed embodies values that may be of the utmost importance, especially regarding procedural justice. Indeed, we consider the interaction and communication between project 'owners' and stakeholders essential for achieving public support of projects. Controversies around value sets may be prevented by considering the variety of values in the design processes.

In the next section, we will explore the way values are embedded in technological artifacts and systems. In Sect. 10.3, we will examine how economists specify values. We highlight the different perspectives of mainstream economics, most commonly applied in economic valuation, and institutional economics, which allows for a much wider conception of the role of values in social life. In Sect. 10.4, we will focus on processes of value specification, emerging in concrete interactions between the stakeholders in a specific project. To describe these processes, we will make use of insights from science and technology studies (STS) and participatory theory. Section 10.5 will present the approach of value sensitive design, providing an analytical framework for a "value hierarchy" that helps clarify the values that (should) underlie particular decisions or characteristics of a design. Based on our theoretical explorations, we will suggest that processes of value specification can be extended to not only include the design of technologies but can also include processes and institutions. This is argued in Sect. 10.6.

10.2 Values in Technology

It is tempting to see technologies as value-neutral, identifying them as simply instrumental practical objects. However, research on the relation between technology and society reveals that this does not correspond to reality. In many ways, technologies are strongly value-laden (Winner 1980), as they incorporate certain (often dominant) values while failing to represent other values. Furthermore, they may also give rise to new types of behavior, and with that they also lead to new expectations and new sets of values. In other words, technologies *mediate* perceptions, experiences, practices, and norms (Verbeek 2006).

As a first manifestation of values in technologies, we may think of an artifact or system that invites or discourages a certain kind of normative behaviour. A clear example is that of a speed bump ("the dead policeman") which urges a driver to take up a cautious driving style. But there are also less obvious applications of value-laden functionalities in the design of technological artifacts and system, as was illustrated in Winner's article "Do artifacts have politics?" (1980). In this article, Winner claimed that the urban planner Robert Moses designed low overpasses over the parkways on Long Island, New York so that buses from New York City could not access the beaches. As a result, the urban poor, primarily African–American population, dependent on the buses for transport, could not reach the shore. The beaches were therefore de facto only accessible for the white, upper and middle classes. Winner concludes that, values (or in his words "politics") are sometimes deliberately designed into technological artifacts. However, the value-laden aspects of technology do not always have to be the result of explicit design. Often these aspects are the outcome of *implicit* design: designers and producers have an implicit world view that drives their technological design. Oudshoorn et al. (2002) showed how different artifacts, such as electric shavers, bicycles, and microwave ovens, were specifically designed with a definite idea about how male and female users relate to such technological artifacts. Moreover, in the world of infrastructures and large socio-technical systems, much of the 'design process' constitutes a path dependent continuation of technical and socio-institutional habits, institutions and practices, which are adapted and expanded according to new insights where possible.

Such an explicit or implicit inclusion of values in technologies may be hard to identify in technologies related to the production and processing of energy. This is because these technologies are part of the wider energy system, which is characterized by a long-term dynamic interplay between technology and societal behavior. Values are not designed as such into a socio-technical system; instead they emerge as an outcome of the expansion and adaptation to heterogeneous activities and technical and social developments in and around the system. Hence, ideas about the 'right' behavior of users on the one hand, or the expectations of the 'right' functionality of the technology on the other hand, are values that *co-evolve* with the development of the system itself (Friedman and Kahn 2002).

For instance, the increasing spatial coverage, density and quality of energy infrastructures have given rise to specific public values. In industrialized regions of the world, citizens not only expect but also take for granted an unlimited, safe, inexpensive energy supply available 24 hours per day, thereby constituting the value of security of supply. Gradually, increasingly stringent conditions have been formulated regarding the impact of energy supply on the environment and on energy resources we pass on to future generations, constituting the value of sustainability. It is clear that these values—so far—have not been fully embedded in technological and institutional design of these infrastructures. The least expensive energy technologies compromise the value of sustainability (e.g. coal power plants) and the technology that maximizes security of supply, may also jeopardize safety (e.g. nuclear power plants). Trade-offs between these values may be inevitable.

The explicit acknowledgment of values in technological design creates the opportunity to include deliberately a variety of public values in the design itself; this is referred to as "front-loaded" ethics (Van den Hoven 2005). We see, for example, that an oil drilling installation or a gas storage facility are often considered by their functionality, whereby questions regarding the public's acceptance of the facility are often ignored. Our proposal is to consider the whole of the trajectory that has led to the realization of the drilling or storage facility, from the start of the project decision-making, its siting and its implementation, to identify the different values that are held by different stakeholders. Only by focusing on these diverse values and their background can we understand and assess the acceptability of the project.

Subsequently, by proactively including values in the design process of the technological artifacts and system, and also in the development of formal and informal institutions surrounding these systems, we aim to support the development of ethically acceptable technological projects that accommodate a plurality of stakeholder values and attenuate potential conflicts and contestation.

An example of such a project that considered various stakeholder values in the design process is the Storm Surge Barrier in the Eastern Scheldt in the Dutch province of Zeeland. After the huge flood disaster of 1953, which killed over 1800 people, it was decided to close off the Eastern Scheldt from the North Sea. However, a conventional dam would have caused the destruction of the unique eco-system of the Eastern Scheldt. Environmental groups and local fisherman protested against the original plan, thus inspiring a new design of the flexible storm surge barrier. In this new design, a flexible barrier was developed. This barrier allowed water to pass through, only if the hinterland is threatened by a flood the barrier would be closed. This new design accommodated both safety and ecological values (Van de Poel 2009a) (see also; Correljé and Broekhans 2013).

10.3 Institutional Economics and Values

Despite the interference of a divergent range of stakeholders, energy projects are in principle economic activities in the market domain. Hence, the lens of economic theory is important for our analysis. Our lens focuses on how economic theories address private and public values in the wide societal context of energy systems. Schools of economic theory have different positions regarding the definition and incorporation of public values in their theories and methodologies. They also approach the safeguarding of public values by markets and governments differently (Correlje and Groenewegen 2009). Here we focus on two mainstream schools: *neoclassical economics* (NCE) and *new institutional economics* (NIE). NCE is founded on the premise of an ideal market in which private stakeholders[1] exchange goods and services. Assuming a set of preconditions, the accumulation

[1]Although economists use the word 'actor' rather than 'stakeholder' in this context, we use the word 'stakeholder' throughout this chapter for reasons of consistency. A stakeholder is defined as any person or party who is affected by, or can affect, the technology and/or its institutional and societal context.

of all private benefits and costs are equal to the social benefits and costs. All values are of an individual nature. However, as is acknowledged by mainstream economists, markets rarely function ideally. Classic cases of market failure, involving natural monopolies, public goods and externalities, justify government intervention. Environmental and safety aspects of (unconventional forms of) energy development are clear examples of negative external effects. Yet, at the same time such projects also may make a potentially positive contribution to security of energy supply. In case of market failures through such externalities, NCE theory suggests that governments should intervene and internalize the external costs and benefits by correcting the prices. Unwanted effects, such as harmful emissions, can be taxed while desired developments can be supported by subsidies. This would imply that authorities are able to value external societal costs and benefits unequivocally, attributing them to the appropriate actors to achieve the optimal outcome for society. This is a fairly abstract perspective on the role of economic science in policy-making far away from the complex social reality.

Also NIE recognizes that actions of stakeholders can have positive or negative consequences for stakeholders who are indirectly involved in the activities. Energy projects that, for instance, cause health risks to the people living nearby may cause a negative externality. Essential in the NIE approach are property rights. It is argued that if there is a setting of clear and complete property rights, private stakeholders will negotiate with each other about using those rights. Victims harmed by a project may, for example, be offered compensation by the facility's owner to accept the risk. Alternatively, the operator may take appropriate measures to minimize the potential impact of its activities. NIE suggests that private stakeholders will negotiate changes in their behavior and should make adequate contracts and agreements. When formal institutions are "right", then the conditions are set for efficient contracting, as postulated by Coase (1960). In principle, the stakeholders will then internalize all consequences of their transactions and the externalities will disappear. Yet, as Coase explained, full internalization of all externalities will only be the case in a world with complete information, in the absence of transaction costs, and with fully rational actor behavior.

Negotiating, drafting, finalizing and monitoring contracts can be an extremely costly affair, taking place in uncertain conditions by actors who are boundedly rational. When transaction costs prevent private stakeholders from internalizing complex externalities, government intervention in (re)allocating rights and arranging compensation is legitimized providing the social benefits of the intervention are higher than the social costs. Both NCE and NIE are based on an individualistic, utilitarian conceptualization of the stakeholders, in which *all* values are eventually aggregated and added up into one price, whether or not they are part of specific transactional arrangements imposed by authorities. However, in our introduction, we proposed to open up this black box of aggregated values, focusing on the variety of stakeholders' values and the mechanisms to incorporate these values in processes of decision-making. This suggests that we have to look for an alternative economic approach.

An alternative conceptualization of economic stakeholders results in another view on the selection through markets and in another role for government with respect to public values. The *Original Institutional Economics* (OIE) is founded on a view of heterogeneous stakeholders who have mental maps and preferences that are influenced by formal (e.g. laws, property rights, bureaucracy) and informal institutions (e.g. customs, traditions, religion). OIE argues that public values reflect preferences of different, sometimes competing, groups in society regarding welfare, well-being, safety, equity, etc. in a given society and at a given time. So, welfare is not considered to be a simple neutral aggregation of the outcome of all individual interests aiming at maximization of utility and profits, as in mainstream economics. Societal welfare is a phenomenon the constituent components and values of which are identified, articulated, developed and operationalized in the socio-political process. Hence, public values do not result from aggregation and neutral selection mechanisms in markets, but are the product of selection processes in a politicized and institutionalized society.

The two mainstream perspectives in the economic framing of energy issues ignore time and location specific micro and macro-relationships between energy production and use in a particular society. Many positive and negative effects are external to today's markets and they will remain so, unless they are explicitly recognized by societies as important public values and deliberately internalized and institutionalized in their energy market. The production and the consumption of energy are inextricably linked to positive and negative environmental, economic and social effects, with local, regional and global impacts. Generally, it can be questioned how and whether such effects and also the benefits of potential solutions are taken into consideration as public values, in the current practices of the evaluation of economic transactions, investments and innovation in energy systems. OIE is concerned with the way in which individual values, like care for the environment, safety, profit-making and security of supply issues become framed and institutionalized as "public" values. OIE examines how such values, that are at first expressed at the level of individuals, evolve collectively towards societal and political pressures in different societies, and how these pressures may drive political decision making and public and private strategy development, followed by subsequent processes of institutionalization as procedures, norms and incentives, guiding technical innovation (also see Veenman et al. 2009).

According to Commons (1936), institutions are the collective action in the control, liberation and expansion of the individual action, highlighting the dual nature of institutions in constraining and allowing or enabling (economic) activity. Achieving responsible innovation of the energy system via the market, thus requires a revision of the prevailing practices of project evaluation, particularly in respect to security of energy supply, safety, and social and environmental aspects. This is less far away from reality than it seems: labour relations, education, social security and external safety rules are all examples of direct and indirect economic, environmental, and social effects that were once fully external to market transactions. They are internalized as public values in today's economies and markets that embrace successfully seemingly divergent and conflicting values and interests.

OIE provides the foundation of a framework with which we can engage in value-based research in a predominantly market-centered institutional environment. It allows us to look for values in places that mainstream economic theory avoid. It also presents us with a wide institutional landscape, in which the activities of heterogeneous groups of stakeholders matter. In fact, we may see the establishment of an institutional environment as the outcome of a heterogeneous process that, in turn, allows the analytical connection with the value hierarchy, as presented below.

It will be the task of the researchers to identify the values that are embedded in the institutional environment that is comprised of formal and informal institutions as well as the (potential) conflicts between these values. Such an analysis implies the study of a broad empirical domain. Not only does it pertain to legal frameworks at different territorial levels (supranational, national, and regional), but it also pertains to strategies, cultures, and routines in a variety of segments of civil society such as businesses, and realms of policy-making.

10.4 The Specification of Values in a Project: Stakeholder Interactions

Above, we have focused on the technology and the institutional environment as value-laden domains. Value conflicts among stakeholders may possibly intensify or decline. Still, one important ingredient is missing, namely if we want to give full consideration to value-laden elements that might contribute to a controversy on the implementation of an energy project, we also have to address the interaction patterns of the stakeholders involved. The rich body of literature on how the general public and local stakeholders respond to and interact with science and technology (Wynne 1992; Eden 1996; Wynne and Irwin 1996; Beierle and Konisky 2000; Devine-Wright 2012) demonstrates how responses to new technologies are largely determined by the *process* through which publics are informed and engaged (Ellis et al. 2007; Walker et al. 2011). This means that the acceptability of a new energy project is determined not only by the characteristics of the technology and the institutional environment but also by characteristics of the decision-making procedure, such as fairness (e.g. procedural justice) (Wüstenhagen et al. 2007).

Public responses to technology are produced in an interaction process between stakeholders with different backgrounds, interests, expectations and attitudes towards the technology (Devine-Wright 2012). Walker et al. (2011) developed a descriptive conceptual framework (see Fig. 10.1) based on multiple European case-studies on public engagement with renewable energy projects. This framework schematically shows how public engagement with renewable energy projects results from the interaction between project developers and public stakeholders who have varying expectations of the technology, of each other, as well as of the process through which the project will be developed.

Four characteristics of this framework are critical to understanding the process of values specification in the interaction between actors and need to be highlighted.

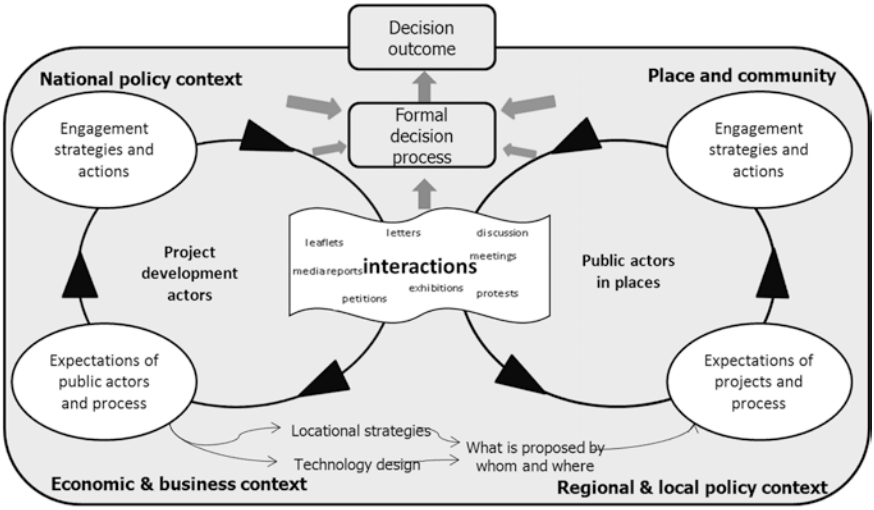

Fig. 10.1 Public engagement with energy technologies (*Source* Walker et al. 2011)

Firstly, the framework is dynamic "in recognizing that, over time, anticipations and expectations evolve and that both the details of proposed projects and the currents of local debates can shift considerably" (Walker et al. 2011). Values in relation to technology are specified in an emergent societal process in which a technology is developed and implemented, and in which multiple stakeholders act and interact in a specific context. For instance, in the Netherlands, the value of flood safety is being reformulated as a reaction to both changes in the perceived threat and in the degree of acceptance of high dikes as the primary means of protection (Broekhans et al. 2010). However, the value of flood safety could also be reshaped if it conflicted with another value; see the example of the Eastern Scheldt Estuary in section two. This demonstrates that the perception of what is acceptable may change over time. Aspects that may influence acceptance include the interaction between stakeholders, stakeholders' values, how conflicting values are addressed, the experiences of the stakeholders, and how these aspects are incorporated in the design of technology and surrounding institutions. Value specification thus takes place in an emergent and dynamic societal process, and the value specification itself can be dynamic. The dynamic nature of value specification in stakeholder interaction points to an essential difference between the way in which values are specified in the interaction processes between stakeholders, on the one hand, and the way in which values are specified in technologies and institutions, on the other hand. A characteristic of technologies and institutions is their relatively fixed nature, that embeds and solidifies certain values in their design. The dynamic character of value specification between stakeholders is an important addition to the theoretical reflections that have been presented in the previous two sections. It emphasizes that values cannot be taken for granted but that that these may pop up and transform during the implementation of the process itself. Ignoring this emergent nature of values might lead to a deepening of the antagonism between different stakeholders, potentially leading to an escalation of the conflict. Such a conflict may be avoidable, however. If the process is managed well, the stakeholder interaction can increase understanding and trust. This implies that it is not possible to fully specify *ex ante* which values need to be taken into account. This aspect has to be acknowledged in the application of value-sensitive design.

Secondly, the framework describes public engagement as a symmetrical process (Walker et al. 2011) that gives equal attention to the stakeholders involved in promoting the project—who can be seen as the project stakeholders—and to the local community, i.e. the 'public'. This symmetry is considered crucial for value sensitive design of energy projects. Most research on public acceptance has focused on the way local communities form their opinions, how they process information, and how they can be involved in participatory processes. It ignores the values, interests, and expectations that project stakeholders have despite the significant impact of how these stakeholders engage with local communities.

Thirdly, the framework identifies expectations and anticipations as shaping local acceptance of projects (Walker et al. 2011). These expectations help to unravel the complex social dynamics in controversial energy projects. Walker et al. (2011) identify four types of expectations that public stakeholders may have: (1) expectations about the form and impact of a project: (2) expectations about the project developer; (3) expectations about the process; and (4) expectations about the proper and appropriate distribution of costs and benefits of a project. The project stakeholders have expectations about the public stakeholders and their responses to the project, on the one hand, and about the development and decision-making process on the other.

These expectations articulate specific values and therefore need consideration in the value-sensitive design of energy projects. For instance, imagine that a project developer announces his plan in a local newspaper, and citizens may respond to that plan based on their expectations of what the project entails and the consequences it may have to their living environment (*distributional* justice issues), whether they would have a voice in the decision-making processes, and how the project will be realized in their community (*procedural* justice issues). They may attend a public hearing and voice their ideas and concerns, to which, in turn, the project developer will respond. Stakeholders' expectations regarding each other, the technology and the decision-making procedure shape their specific values. This specification is a dynamic process; neither the specified values nor the operationalization of the values are fixed in the design process. The project developer may expect the public to be ill-informed and risk-averse; a deeply rooted belief (see Wynne 1993, 2001). Interaction is then likely to be geared towards providing technical facts that underscore the safety of a project. Yet, actually, the public may be more concerned about procedural issues, such as fairness and transparency (Walker et al. 2010), or the distribution of costs and benefits. These concerns are not addressed by providing more information on technology and risk. This mismatch may frustrate the process, leading to the paradox that preventive efforts, by providing "the hard facts", may actually provoke public opposition.

The Not In My Backyard (NIMBY) label is another well-known example of framing that might increase the gap experienced between project stakeholders and local communities. This label is often used to explain public resistance to (local) technology projects. The NIMBY notion implies a social dilemma: it suggests that citizens have a favorable attitude towards a specific technology (e.g. wind energy) that becomes a negative attitude when the siting of that technology is "in their backyards". This is a simplistic understanding of public response that also proves to be invalid in many cases (Wolsink 2000). The NIMBY label frames the public as concerned only with its own short-term interest. As such, the NIMBY label influences the dynamics of the debate by steering it towards a conflict between the public good versus individual interests.

Fourthly, the framework acknowledges the influence of contextual factors on public engagement (Walker et al. 2011). Four types of context are distinguished: (1) characteristics of place and community, (2) regional and local policy, (3) national policy, and (4) business. Tapping into the meanings that are assigned within these contexts suggests that different values may be at stake in different contexts. An example (from Walker et al. 2011) from the context of place and community illustrates this. An offshore wind farm was planned in Llandudno, a Welsh village. The more people felt attached to this village, the more they opposed: "Tapping into place meanings provided contextual information as to why this was the case—Llandudno was a unique place that was attractive to tourists, characterized by its scenic, natural beauty also because of the view on the sea and its Victorian heritage. These meaningful characteristics were widely perceived to be threatened by a wind farm that would 'industrialize' the area and 'fence in the bay'".

This example illustrates the dynamic nature of value specification through stakeholder interaction. This is an important notion for value sensitive design. The emergent and dynamic nature of value specification demonstrates that a full *ex ante* assessment of relevant and conflicting values is not possible. Designing for values requires a continuous and flexible participatory approach. If managed well, such a participatory approach can increase trust and mutual understanding between the stakeholders which is necessary to facilitate the shaping of widely supported technological and institutional designs. This approach should be symmetrical in its consideration of the values of the project stakeholders and those of the local communities affected by the project. It should also include the interaction between these groups of actors and the dynamics of value specification that results from this interaction process. The vales that are articulated by stakeholders shall incorporate expectations about the project and its impacts, about the decision-making procedure, and about the other stakeholders involved in the process. Finally, the approach should be sensitive to specific contextual factors which may render some values more salient in context A than in context B. Value sensitive design will never be a blueprint but it must target a specific context in which specific cultural, political and economic factors shape the process of value specification.

10.5 Value Sensitive Design

Value sensitive design aims at systematically incorporating diverse human values in the design of new technologies. This method has been primarily introduced and developed in information technology and for designing human computer interaction (Friedman and Kahn 2000, 2002; Van den Hoven 2007), but later it has been elaborated to address the inclusion of moral values in other domains of technological design (Nissenbaum 2005; Van de Poel 2009b; Van den Hoven et al. 2014; Taebi and Kloosterman 2014). Scholars of VSD argue that the design process has value implications because new technology can shape our practice and hence promote or undermine certain values (Van Den Hoven 2008). Friedman and Kahn (2002) present VSD as a tripartite iterative method that integrates conceptual, empirical and technical investigations. Conceptual investigations involve a philosophical questioning of the values. Which values are affected in what way by technological design? Who is affected? How to engage in trade-offs among values? Empirical investigations are aimed at social-scientific understanding of experiences of people affected by technological design. Technical investigations analyze the technical artifact or system to assess how they support or undermine certain values and inspire the development of alternative technical solutions.

There are many challenges and difficulties in following the tripartite methodology of VSD. Manders-Huits (2011) discusses two key issues. Firstly, whereas VSD emphasizes the need to consider all stakeholders (i.e. those who use the technology and those who could be affected by the use of technology), it lacks a clear methodology for identifying these stakeholders and for assessing and systematically including stakeholders' values. So, VSD will always be in need of social-scientific empirical methods. Secondly, once we are familiar with the conflicting values, it is not clear how trade-offs should be dealt with. In other words, VSD will always be in need of moral analysis and ethical theory. In the remainder of this section, we will present our approach and discuss how our adjustments to VSD could help overcome some of these methodological problems.

We propose to apply VSD not only in the case of the technological design of energy systems but also in institutional design and in designing public participation. As shown above, values are specified in dynamic social processes and are embedded in the formal and informal institutions that surround a major technological system. Hence, in the iterative approach of VSD, we will analyze this dynamic and the *design* of the surrounding institutions in such projects. Following Van de Poel's approach for translating values into design requirements, we will distinguish between three different levels in a "value hierarchy" (Van de Poel 2014). At the highest—most abstract—level, there are fundamental *values* someone may hold paramount such as safety, environmental friendliness, economic efficiency and so forth. Contestations do not (often) arise from what constitutes a value. Everybody will supposedly endorse abstract values like safety, equity, efficiency, etc. Rather, controversy arises from how the value is specified into *norms*. Norms are located at the second level of hierarchy and form 'prescription for or restriction on' actions (Van de Poel 2014). Such norms may include *objectives* (like "maximize safety", "safeguard environment" or "minimize costs" without a specific target), *goals* that specify a more tangible target, and *constraints* that set boundary or minimum conditions. The bottom level of a value hierarchy, which is also the most concrete one, indicates the technical and institutional *design requirements* that are derived from the norms. Figure 10.2 illustrates this hierarchy.

The value hierarchy can be used both as an analytical tool and as a design tool. As an analytical tool, it can help to analyze *why*, or *for the sake of what* (Van de Poel 2014), something is being done or preferred by someone. It can help to explicate the values that underlie certain decisions or characteristics of a design (analyze *why*) and it can help to illuminate controversies when values and/or norms were specified in the design process but not incorporated in the design (analyze *for the sake of what*). As a design tool, the value hierarchy can be used to come up with a design that is robust in the sense that it can bring together divergent values and norms into a coherent set of *ex ante* design requirements, as regards process and substance. VSD investigations should start with applying this value hierarchy as an *analytical* tool to unpack the emergent societal process of value specification in energy projects. With this analytical tool, the following could be achieved:

- Insight in values, norms and design requirements can be identified in the *technological* and *institutional* design;
- Insight in the way that the interaction between stakeholders and their expectations (about the technology, a specific project, other stakeholders and decision-making procedures) has shaped the process of value specification;
- Insight in the way that values that are held by stakeholders are specified and institutionalized as procedures, norms providing the (dis)incentives for technical innovation.

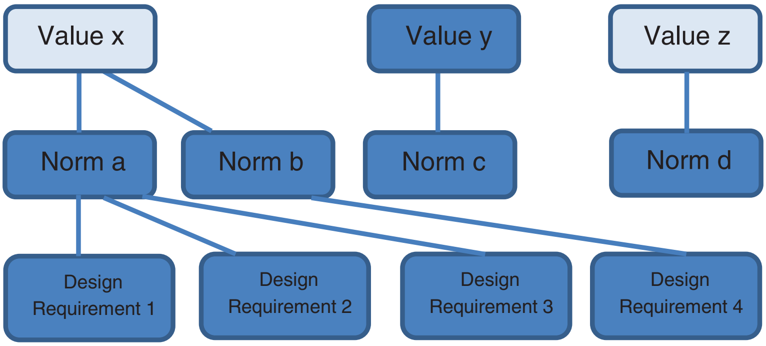

Fig. 10.2 The three levels of value hierarchy: i.e. values, norms and design requirements

Our main goal in applying VSD as an analytical tool in future work is to identify and understand the conflicting values and controversies (see also Taebi et al. 2014). However, VSD can also be applied to the value hierarchy in a design tool fashion. In the design fashion, VSD will explore the potential for changing technological features or institutional design characteristics in a way that conflicts can be resolved. Indeed, controversies and conflicting values may fuel technological innovation as they did in the Eastern Scheldt Estuary. Moreover, controversies may also give rise to substantial changes in the institutional context including the rules for decision-making.

10.6 Discussion and Conclusion: Designing Technology, Institutions, and Stakeholder Interactions

Energy projects are strongly driven by market incentives and regulation. However, for their successful implementation, a focus on market and regulatory incentives alone is insufficient. Even when externalities and transaction costs are effectively incorporated in the project design, economic evaluation ignores the dynamic in stakeholders' values. These omissions may cause trouble in the implementation of projects even though they may seem rational from an economic perspective. Stakeholders' values are also influenced by the specific process through which an energy project is initiated and licensed. So, for the successful initiation of an energy project, it is important to pay attention to the process through which the project becomes established and to the stakeholders' values that are addressed in this process.

Each design process embeds latent values in technologies and in its institutional environment. By making the potentially conflicting values explicit, they can be accommodated in the design of technologies and institutions. Anticipatory actions can be taken, so that the expensive and cumbersome management of potential public controversy can be avoided. The approach presented in this paper opens up the black-box of values and provides an analytical framework with which value conflicts around energy projects can be managed.

A challenge for VSD is how to use the identified value conflicts to (re)design technologies and institutions. The value hierarchy suggests that we first have to find out which values and norms are embedded in technologies and institutions and which are specified by the various stakeholders. Then the sets of values and norms have to be translated into design requirements. To do so, we will have to address the following questions:

- Are there any values and/or norms missing in the current *technological* design? How can these be specified into design requirements?
- Are there any values and/or norms missing in the current *institutional* context? How can these be specified into design requirements?
- Do the processes in which the different groups of stakeholders interact allow for the specification of all stakeholders' values?

Until now, the framework proposed in this paper has been based on theoretical notions. Empirical and contextual detail will shed light upon a number of important questions that as yet remain unanswered. This requires a thorough and context-specific operationalization of the framework, which will be the first step to take. One of the most important questions then, involves the relation between the technological design and the institutional design. Ideally, both institutions and technology could be subject to value-sensitive design. However, it may not always be possible to have a satisfactory (re)design of both elements, which prompts questions such as: are people willing to accept a technological artifact in their backyard, if their values are taken seriously in a well-designed institutional context? Under what conditions would they do so? In other words, what determines the flexibility of public acceptance? This question cannot be answered theoretically; it clearly urges the operationalization and empirical testing of our framework. The framework we propose, promises an integral approach to manage value-laden conflicts in the development of new energy projects.

Acknowledgements The work for this article has been funded by NWO (the Netherlands Organization for Scientific Research) [grant number 313 99 007].

References

Beierle, T.C., and D.M. Konisky. 2000. Values, conflict, and trust in participatory environmental planning. *Journal of Policy Analysis and Management* 19: 587–602.

Bell, D., T. Gray, and C. Haggett. 2005. The 'social gap' in wind farm siting decisions: Explanations and policy responses. *Environmental Politics* 14: 460–477.

Broekhans, B., A. Correljé, and J. Van Ast. 2010. Allemaal op de bok. *Kijk op waterveiligheid*, 123.

Coase, R.H. 1960. The problem of social cost. *Journal of Law and Economics* 3: 1.

Commons, J.R. 1936. Institutional economics. *The American Economic Review* 26: 237–249.

Correljé, A., and B. Broekhans. 2013. Floodrisk management in the Netherlands after the 1953 flood: A competition between the public value(s) of water. *Journal of Flood Risk Management*.

Correlje, A.F., and J.P. Groenewegen. 2009. Public values in the energy sector: Economic perspectives. *International Journal of Public Policy* 4: 395–413.

Devine-Wright, P. 2012. *Renewable energy and the public: From NIMBY to Participation*: London: Routledge.

Eden, C. 1996. The stakeholder/collaborator strategy workshop. In *Creating collaborative advantage*, 44–56. London: Sage.

Ellis, G., J. Barry, and C. Robinson. 2007. Many ways to say 'no', different ways to say 'yes': Applying Q-methodology to understand public acceptance of wind farm proposals. *Journal of Environmental Planning and Management* 50: 517–551.

Feenstra, C., T. Mikunda, and Brunsting, S. 2012. What happened in Barendrecht?! Case study on the planned onshore carbon dioxide storage in Barendrecht, the Netherlands. *Policy Studies*, 2011, 2010.

Friedman, B. and P.H. Kahn Jr. 2000. New directions: A value-sensitive design approach to augmented reality. In *Proceedings of DARE 2000 on designing augmented reality environments*, 163–164. New York: ACM.

Friedman, B. and Peter H. Kahn Jr. 2000. New directions: A value-sensitive design approach to augmented reality. In *Proceedings of DARE 2000 on designing augmented reality environments*, 163–164. Elsinore, Denmark: ACM.
Friedman, B. and P.H. Kahn Jr. 2002. Human values, ethics, and designed. *The human-computer interaction handbook,* 1177–1201. New Jersey: L. Erlbaum Associates Inc.
Manders-Huits, N. 2011. What values in design? The challenge of incorporating moral values into design. *Science and Engineering Ethics* 17: 271–287.
Nissenbaum, H. 2005. Values in technical design. In *Encyclopedia of science, technology and society*, ed. C. Mitcham. MacMillan: New York.
Oudshoorn, N., A.R. Saetnan, and M. Lie. 2002. On gender and things: Reflections on an exhibition on gendered artifacts. *Women's Studies International Forum* 25: 471–483.
Persson, M. 2012. Bodem-oorlog *De Volkskrant*, 7 January 2012.
Roeser, S. 2011. Nuclear energy, risk, and emotions. *Philosophy and Technology* 24: 197–201.
Taebi, B., and J.L. Kloosterman. 2014. Design for values in nuclear technology. In *Handbook of ethics, values, and technological design: Sources, Theory, Values and Application Domains*, ed. J. Van den Hoven, P. Vermaas and I. Van de Poel. Dordrecht: Springer.
Taebi, B., A. Correljé, E. Cuppen, M. Dignum, and U. Pesch. 2014. Responsible innovation as an endorsement of public values: The need for interdisciplinary research. *Journal of Responsible Innovation* 1: 118–124.
Van De Poel, I.R. 2009a. De Oosterscheldekering. Een voorbeeld van waardenbewust ontwerpen. In *De politiek der dingen*, ed. F. Bolkestein, J.M. Van Den Hoven, I.R. Van Den Poel, and I. Oosterlaken, 63–78. Budel: Damon.
Van De Poel, I.R. 2009b. Values in engineering design. In *Philosophy of technology and engineering sciences*, ed. A. Meijer, 973–1006. Amsterdam: Elsevier.
Van De Poel, I. 2014. Translating values into design requirements. In *Philosophy and Engineering: Reflections on practice, principles and process*, ed. D. Mitchfelder, N. Mccarty and D.E. Goldberg, 253–266. Dordrecht: Springer.
Van Den Hoven, J. 2005. Design for values and values for design. *Information Age* 4: 4–7.
Van Den Hoven, J. 2007. ICT and value sensitive design. In *The information society: Innovation, legitimacy, ethics and democracy in honor of Professor Jacques Berleur sj*, ed. P. Goujon, S. Lavelle, P. Duquenoy, K. Kimppa and V. Laurent, 67–72. Boston: Springer.
Van Den Hoven, J. 2008. Moral methodology and information technology. In *The handbook of information and computer ethics*, ed. K.E. Himma and H.T. Tavani, 49–68. New York: Wiley.
Van Den Hoven, J., P. Vermaas, and I. Van de Poel. 2014. *Handbook of ethics and values in technological design*. Dordrecht: Springer.
Veenman, S., D. Liefferink, and B. Arts. 2009. A short history of Dutch forest policy: The 'de-institutionalisation' of a policy arrangement. *Forest Policy and Economics* 11: 202–208.
Verbeek, P.P. 2006. Materializing morality design ethics and technological mediation. *Science, Technology and Human Values* 31: 361–380.
Walker, G., N. Cass, K. Burningham, and J. Barnett. 2010. Renewable energy and sociotechnical change: Imagined subjectivities of 'the public' and their implications. *Environment and Planning A* 42: 931–947.
Walker, G., Devine-Wright, P., Barnett, J., Burningham, K., Cass, N., Devine-Wright, H., Speller, G., Barton, J., Evans, B. and Heath, Y. 2011. Symmetries, expectations, dynamics, and contexts: A framework for understanding public engagement with renewable energy projects. *Renewable Energy and the Public. From NIMBY to Participation*, 1–14.
Winner, L. 1980. Do artifacts have politics? *Daedalus* 109: 121–136.
Wolsink, M. 2000. Wind power and the NIMBY-myth: institutional capacity and the limited significance of public support. *Renewable Energy* 21: 49–64.
Wolsink, M. 2006. Invalid theory impedes our understanding: A critique on the persistence of the language of NIMBY. *Transactions of the Institute of British Geographers* 31: 85–91.
Wüstenhagen, R., M. Wolsink, and M.J. Bürer. 2007. Social acceptance of renewable energy innovation: An introduction to the concept. *Energy Policy* 35: 2683–2691.

Wynne, B. 1992. Misunderstood misunderstanding: Social identities and public uptake of science. *Public Understanding of Science* 1: 281–304.
Wynne, B. 1993. Public uptake of science: A case for institutional reflexivity. *Public Understanding of Science* 2: 321–337.
Wynne, B. 2001. Creating public alienation: Expert cultures of risk and ethics on GMOs. *Science as Culture* 10: 445–481.
Wynne, B. and Irwin, A. 1996. Misunderstanding science. *The Public Reconstruction of Science*. Cambridge: Cambridge University Press.

Chapter 11
Capacities, Development and Responsible Innovation

J. Otto Kroesen, Rudi Darson and David J. Ndegwah

Abstract Moral aspects of capacity building as part of responsible innovation deserve a more central place in the development debate and research. This comes to the fore in the capacity to deal with project implementation and the attitudes and values involved in it. The authors first provide some clarifications on the concept of capacity, emphasizing the value laden meaning of the concept and relating it to responsible innovation. In development programs different value sets and attitudes are confronted with each other, leading to different capacities, often competing for priority. Two cases involving capacity building in technology are analyzed, one on the introduction of tropical greenhouses in Kenya and one on the renovation of a vocational school in Surinam. This shows the central meaning of learning processes and careful and intense cooperation with the owners of the change processes. These are important aspects of innovating in a responsible way. The authors propose a path dependent way forward, prioritizing different values and cultural traits (and related capacities) according to time, situation and need. With that objective in mind—in order to innovate responsibly—there is an important role for the (meta-)capacity to alternate between different values and capacities, finding the right rhythm and equilibrium between different modes of behavior and cooperation, thereby integrating different values into a comprehensive strategy for development.

J.O. Kroesen (✉)
Faculty of Technology, Policy and Management,
Technical University Delft, Delft, The Netherlands
e-mail: J.O.Kroesen@TuDelft.nl

R. Darson
Rotterdam, The Netherlands
e-mail: R.Darson@gmail.com

D.J. Ndegwah
School of Humanities and Social Sciences,
Jaramogi Oginga Odinga University of Science and Technology,
Bondo, Kenya
e-mail: NdegwahDavid@gmail.com

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_11

Keywords Capacity · Capacity building · Economic culture · Knowledge · Innovative learning processes · Stakeholder involvement · Responsibility · Sustainable development

11.1 Introduction

This contribution is born out of the experience that capacity building is a crucial factor for development. Capacity can in first instance be understood as the ability to get things done (Balassanian 2006). But in order to do so, the people involved need the competence, the character, and mental orientation required (UNDP 1997). For that reason capacity is a value laden concept and as such intrinsically related to responsible innovation. Responsible innovation can be approached from a process angle and from a product angle (Koops, Chap. 1, this volume). The same counts for capacity training. The product approach primarily looks at the objective of responsible innovation and as such it mobilizes the language of moral values that should be incorporated in innovative products. The process approach focuses comparatively more on the way leading to this objective—it should be participatory and responsive. Although this contribution involves values, setting a standard for social and technological innovation, it focuses more specifically on the process aspects: learning, dialogue, responsiveness, the same as with Edelenbosch et al. (Chap. 8, this volume). The authors treat capacity building as a learning process involving knowledge, skills and (changing) attitudes and values. The ultimate goal is sustainable development, which is considered by Blok and Lemmens (Chap. 2, this volume) to be a wicked problem, making it quite difficult to satisfy the goal of responsible innovation. That is most certainly also the case for responsible innovation in development contexts.

Achieving sustainable development and taking responsible innovation into consideration is indeed very hard, but not impossible. Optimism in itself will not do, as also comes to the fore in Balkema and Pols (Chap. 15, this volume) where jathropa, a bio-fuel plant, which first was considered a miracle solution for the local farmers in Tanzania to achieve sustainable development, turned out to be a problem due to poor results and negative consequences for the farmers. What this experience shows, as do the cases described in the present chapter, is that change cannot be imposed from the outside. That would also run against the grain of what responsible innovation tries to achieve (Sutcliffe 2013; Owen et al. 2013). Real durable positive change must come from within. Continuity is guaranteed if capacity is ingrained in the people and shared in the social network involved, and ultimately realized in technology and institutions. That is the reason for the authors of this chapter to focus more on the process side of responsible innovation.

Technology has been generally accepted as a major driving force behind development (Bell 2009; Kim 1997; Lall 1992; Nelson and Pack 1999) and the capacity to deal with technology is often mentioned as an important factor. Nevertheless, its building blocks have not been properly analyzed. For instance Arond et al. (2011), Nelson and Phelps (1966) as well as the UNDP (2001) mention it, but without proper analysis of its components. An important component consists, so this chapter proposes, of an underlying set of cultural values. Technical solutions should be designed in constant dialogue with these underlying values in a mutually reinforcing process of growth, both in terms of economic output and in terms of capacity for responsible innovation. In that regard it should be taken into account that growth in contexts where a strong state and a well-developed civil society are lacking, like mostly in Africa, takes a different shape than in Asia or Latin America, where state institutions are much more powerful (inter alia, Katz 2001; Bell 2006).

The authors will put the concept of capacity within the broader context of the development debate and theories. Then they will define the concept of capacity, focusing on its value laden aspects. They will then present case material derived from the internship program of the minor "International Entrepreneurship and Development" from Delft University of Technology. Finally, they will set a research agenda, which should lead to a more encompassing understanding of these capacities. Since capacities are closely related to inherited cultural values, solutions need to be found for tensions between old and new values, between traditional and modern sets of value, as a matter of responsible innovation. Of course short-term economic success is not the only criterion for deciding such trade-offs. If long-term productive capacity is not increased, short-term economic successes will not be sustainable in the first place. A responsible equilibrium between old and new values should at least entail an increase in human capabilities (Nussbaum 2006). For all those reasons the cases will focus on the participatory component of responsible innovation emphasizing the importance for the learning process and the building of capacity.

11.2 Capacities as Crucial for Development

In the course of the last 50 years, the development debate has witnessed a remarkable shift in attention from economic and technical factors to human actors. After the independence of the former colonies the idea was predominant that financial and technical inputs would make it possible to catch up soon (Moyo 2009). However, no matter how much money and effort was put into it, development often did not really take off. Nevertheless the emphasis on technical and economic hardware keeps popping up now and then, like in the work of Sachs (2005) in relation to the Millennium Development Goals, when he pleads for big investments and large-scale plans in order to get out of the poverty trap.

However, on the whole, slowly but steadily society became more central in development debates and discourses. During the 90s the word capacity entered the development debate with two streams of thought. The first one, primarily from the perspective of economics, emphasized technological capability, in terms of skills, knowledge to choose, to assimilate and absorb, to adapt, modify and ultimately create new technology (e.g. Katz 2001; Bell 2006). The second stream introduced the notion of institution building, community management, civil society and governance. As such and also related to the term governance the emphasis on capacity building was more or less the answer from the World Bank (1989) and other institutions to the criticism, that the structural adjustment programs of the 1980s in developing countries were doing more harm than good (Chang 2007). The expression “governance issues” was diplomatically used to increase the awareness of the underlying dynamics within the societies in question. It is also related to a poorly developed civil society and the related cultural outlooks and values (Ndegwah 2011). One would be inclined to ask, why the developmental institutions in those countries were functioning so poorly in the first place? The answer became increasingly clear: behind the institutional framework we have to take a look at the mentality of the people, characterized by a particular set of traditional values, which function within and in turn support and maintain these institutional frameworks—or not.

Speaking of a particular mentality undoubtedly implies generalizing and may sound a bit pre-judgmental. But if we realize that the typology we propose is indeed a generalization, which does not cover all individual cases, we come to the following picture. The reader who is familiar with the intercultural work of Hofstede (1991) and Trompenaars (1999) will easily recognize the conceptual distinctions they make in the description below. Different cultural value sets may enhance or stand in the way of economic development (Porter 2000; Grondona 2000). In this regard Porter coined the term “economic culture”: “Economic culture is defined as the beliefs, attitudes, and values that bear on the economic activities of individuals, organizations, and other institutions” (Porter 2000, 14).

In sub-Saharan Africa—according to personal experience in project monitoring—the following economic culture is predominant. The community and (extended) family one belongs to are very important. Actually, there is a clear dividing line between the family and the rest of society. The inside community, the family, is treated as the primary goal and value—although this can be more or less extended towards neighborhood, clan or tribe, even political party and company—and beyond that it is allowed to use people and relationships merely in an instrumental way (Ndegwah 2006). The position in the social hierarchy is important. People respected for their age or their standing in the community, have a higher status and are prone to have better access to the state bureaucracy. People of important and rich families will also have better jobs even though they may lack competence or expertise. The state bureaucracy in turn grants privileges to people of high status and creates obstacles for people of low status. The rule of law is often bent under the pressure of collective interest groups, important relationships, and privileges. The state does not create a level playing field. The group that occupies the state power brings its constituency to rule with no role for the opposition and no eye for other people to be served. Groups, and being part of the group, are important means to create access to state services and group membership is instrumental to better treatment and more privileges. There is little or no opportunity for multiple memberships and regrouping of individuals or organizations. Instead there is strong compartmentalization of society. Great as the solidarity of the African families and clans may be, and praiseworthy as the concept of Ubuntu (being human with and through others) is—and indeed in the slums of

Africa nobody dies from immediate hunger because there is always some form of family or neighborhood solidarity for anybody in need—nevertheless, lack of anonymous trust and lack of confidence stands in the way of cooperation between different parties, families, clans, tribes, companies, organizations. In addition, the general attitude towards labor as a value in itself is not very positive. People at the top grant themselves privileges and leave the hard work to others—the employees at the bottom. This mentality reinforces the lack of growth of productivity, because the smarter people at the top do not invest in innovation and in improving the labor conditions of the workers. In the development and invention of technology an important condition for progress from the Middle Ages onwards has been that intellectuals did not eschew labor in order to make progress. Many inventions originated in that attitude (White 1962; Rosenstock-Huessy (1938) 1993; Cardwell 1995). This low value put on labor adds up to the strong traditionalism in the African society, doing things as they always were. This is the main reason that people do not consider themselves and their initiatives as turning points of change. Instead uncertainty avoidance is strong and people do not like to get taken out of their comfort zone. The future, which is unknown after all, is not looked upon as a resource of potential change or improvement. The present is more important than the future. At any time a person's work can be stopped to take up the cell phone or attend to a person coming in. Only the drive to change the present and to reach a future goal can lead to a voluntaristic attitude of initiative and change. Only then planning is important in order to make a difference.

This is not to deny that traditional values are meaningful. African communalism is friendly and warm and the fact that the present is more important than the future means that everybody knocking at the door is welcome at any moment. The important question, however, is how much and in what mix traditional values are counterbalanced by more entrepreneurial values in order to promote a viable and responsibly innovative economic culture. Those values which are indispensable for running a modern economy are not well developed. They were not required in a traditional agrarian or nomadic society like Africa was, until recently. "Is this culture compatible with the demands faced by individuals and nations at the beginning of the 21st century?"—asks Daniel Etounga-Manguelle (Etounga-Manguelle 2000, 66) and rightly so. He continues: "Does Africa need a cultural adjustment program?" If he is right, the question becomes all-important: how to innovate responsibly or in other words, how to keep the good things of traditional African ethics and at the same time introduce a more modern entrepreneurial and innovative attitude? Here a better understanding of the concept of capacity may help the discussion forward.

11.3 Capacity and Responsible Innovation

In the literature the meaning of capacity appears to be rather context dependent (Eade 1997). It is often mentioned along with knowledge, skills and competences, but it also has a ring of basic beliefs, values and culture (James and Hailey 2007;

James and Wrigley 2007). Often capacity training or capacity building is called for, but the question remains what actually is trained, if capacity is trained. The UNDP proposes the following definition: "Capacity development is the process by which individuals, organisations, institutions and societies develop abilities (individually and collectively) to perform functions, solve problems and set and achieve objectives" (UNDP Report Capacity Development 1997). The document distinguishes between four levels, the individual (skills, knowledge, salary, values etc.), the organizational level (a person can only work effectively as part of a larger entity), the level of interrelationships and arrangements between organizations and finally an enabling environment (state, civil society and private sector). Elsewhere the UNDP stresses the fact that capacity should include change (UNDP 2006), which, taking into account the traditional cultural attitude of Africa is not a value free statement. Based on the components randomly mentioned in the literature and on our own experience with the capacity problem in so many projects we propose to distinguish three layers of meaning. The first layer involves knowledge (information, insight, judgment) (1), and experience (2) as well as skills (competences) (3) and expedience (4). These can be summarized in the term IESA: Information, the Experience, the Skills and Attitude—an acronym borrowed from Weggeman (1997, 2000). In our approach we would like to replace his word "attitude" by the word algorithm in the sense of expedience. Attitude, in our conceptualization, belongs to the second layer of capacity.

The combination of capacity with "training" or "building" suggests that also the inner core of the self is targeted. Thus capacity also entails values such as taking responsibility, internal discipline, openness, loyalty etc. (5). These values are expressed in attitudes towards other people and towards work (6). Actually, values always imply social relationships and interactions (Joas 1999). They are based on meanings and beliefs (7). For this 2nd layer of capacity we propose the term VAB, Values, Attitudes and Beliefs regulating ways of interaction and cooperation. Changing values, attitudes and beliefs is a slow process, because the unfamiliar takes time to be integrated (Maturana en Varela 1980). For that reason long-term learning processes and processes of internal growth are important. In the teacher–student relationship for instance so many times the real lesson is only learned when the student after extensive supervision has to act on his own for the first time. Internal capabilities cannot survive without external support from the social environment (8) and finally the institutional framework in terms of laws and regulations (9) should also be drawn into the orbit of a definition of capacity. For instance, the fact that individual persons express their judgment is an important asset for freedom of speech. However, the institution of a free press in turn highly stimulates the human quality of having a judgment of one's own. Political freedom of speech is dependent on a parliament and an open society in which the opposition also has a say. But it is also dependent on the courage of individuals to speak out in public. Therefore capacities and institutions support each other in a reciprocal way. This 3rd layer of capacity can be captured in the formula SI (an enabling social environment and institutions).

Table 1 Components of capacity

IESA—knowledge	Information, insight, judgment experience, skills, competences, (algorithm) expedience
VAB—values and attitudes, beliefs	Values (such as taking responsibility, internal discipline, openness, loyalty etc.), meanings, beliefs, expressed in attitudes and codes of behavior
SI—enabling environment	Social institutions, policies and regulations supporting particular values and attitudes (as well as the other way around), organization for capacity building

These three layers need support from each other. Let us take entrepreneurship as an example. Entrepreneurship for development is not only a matter of business knowledge and organizational skills (IESA) but it is also value laden (Samli 2009). Entrepreneurship requires a nontraditional attitude of personal initiative, and being able to cope with uncertainty (VAB). But without an enabling environment consisting of adequate regulations and policy arrangements, the many initiatives from individual entrepreneurs will fail to bring about lasting change (SI). Capacity as described above can therefore be captured in the formula $C = f[(IESA) + (VAB) + (SI)]$ (Table 11.1).

The failure of many projects often has its root cause in the deeper layer of attitudes and values whereby human experience and judgment is shaped and framed. One cannot expect automatic acceptance of cultural traits, which go against the basic attitudes and assumptions of a culture (Nyasani 2010).

Now how is capacity building related to responsible innovation? Innovation can be of a disruptive (Christensen 2003) and a destructive nature (Schumpeter 1942) and for it not to harm people and their environment (ethical aspect) and for it to take hold it always needs to be carried out in a responsible way. In her report *Responsible Research and Innovation*, Sutcliffe (2013) mentions the following aspects, of which many are also mentioned by Owen et al. (2013): (1) a deliberate focus of research and innovation to achieve social and environmental benefits, (2) consistent, ongoing involvement of society from beginning to end in the innovation process, (3) assessing and effectively prioritising social, ethical and environmental impacts, risks and opportunities now and in the future. It also consists in (4) the possibility of using oversight mechanisms to anticipate and manage problems and opportunities and respond and adapt quickly to changing circumstances and knowledge; and (5) finally openness and transparency are an integral part of the research and innovation process. Going along with this framework the authors would like to emphasize, that innovation also implies the adoption of new values. Technological innovation and capacity training should therefore dovetail into each other. Again the decisive question is how such innovation can be implemented responsibly, finding a middle way between respect for tradition and the necessity of change. If it is not done responsibly, well intended interventions may cause a lot of damage. Therefore we present two cases, one in which the process is not conducted very well, although not much damage is done apart from the spilling of resources and efforts. In the second case the process has been monitored and coached more successfully, despite many failures and setbacks over time. It may not be a coincidence that this more successful case shows a long-term commitment over seven years, whereas the first case only shows an involvement of three months. This already shows that the second point of Sutcliffe (2013) is all-important for the success or failure of capacity building.

11.4 Case Material: Kenya and Surinam

The case material the authors present is derived from the internship program of Delft University of Technology especially in Kenya and Surinam. As part of a minor or as a master thesis, students often conduct an internship in developing countries. For the students these projects are as much training in capacity, the capacity to manage projects and reflect on their own role in a different cultural context, as they are for the partners overseas, who experiment with new ways of management. Also for the staff each new project adds to the intercultural expertise.

11.4.1 Amavo Greenhouses—Kenya

Like in developed countries also in developing countries SMEs have a large role to play in the economy. It is estimated that SMEs generate 40 % of the national income instead of the official 13 % (Daniels 1999). Many of these SMEs do not grow beyond the level of small initiatives. Among the causes generally mentioned are lack of education and work experience. There are also several institutional constraints. If they grow it becomes more difficult to bypass taxation, there are constraints on access to capital, on access to relevant networks and (lack of) cooperation between suppliers and firms and buyers, presence (or not) of supporting markets and the general level of skills (Field et al. 2000; Gibson et al. 2001; Lusby and Panlibuton 2002). Finally the management style makes a difference. If for instance the management style is too hierarchical and relying too much on command and control, the loyalty of the workforce will fade away quickly and turn into resentment with the result that many employees are secretly waiting for the opportunity to start a business of their own. This leads to many small businesses in the same market, competing for the same customers.

Amavo Greenhouses is such an SME that provides affordable, durable greenhouses for small-scale farmers in Kenya.[1] It is recently started up by private capital and trying to get a share of the market. It has a vision to reduce poverty and malnutrition by giving farmers the opportunity to grow more, different and healthier crops using greenhouses to produce all year round and stimulating to grow

[1]The names in this case have been anonymized. The authors thank the students, Emma Cherim, Lou van Heemst and Noortje Kallen for the work they did and for their contributions.

crops organically. The project started in cooperation with Dr. Makao, the owner of the company, and an American university. She has a background in psychology, but returned from America to Kenya. Students from the American university developed the first prototype of the greenhouse. From the Delft University of Technology three students went to Nairobi for a three-month internship at Amavo (Cherim et al. 2013). The students specialized in Industrial Design Engineering and Technology, Policy and Management.

The company was founded in May 2012, but as yet sales did not rise beyond the number of four greenhouses. Therefore the main goal for the interns was to help Amavo Greenhouses kick-start in this critical period. The biggest challenge was to help expand the sales, they were told. Immediately after they arrived, however, the students discovered that they needed to do a lot of ground work even before trying to expand. Together with Dr. Makao they carried out a SWOT analysis, from which some important insights were derived. A strong point of Amavo is the fact that it is the only company promoting organic farming and offering greenhouses made with sustainable materials and being based in Nairobi. In addition, it has access to a substantial group of potential customers that are interested in an urban greenhouse. On the basis of this customer support, poorer customers could be targeted in the future. But there were also a lot of weaker points. One greenhouse which was built already showed some failures, because the builders did not use reliable drawings and standardized building procedures. In part (for one type of greenhouse) these procedures which had been taught during a one-time training event in May 2012 had already been forgotten. Therefore standardizing the greenhouse and providing a building manual were needed first. The students concluded that their support was required in helping the CEO in managing all the aspects of her company well, keeping track of material costs, of incoming and outgoing bills, and setting the prices for the customers on the basis of those inputs. A fourth problem was the lack of agricultural knowledge among many of the smallholder farmers, who are the target group. This was not only necessary for the farmers themselves in order to be successful, but also to convince potential providers of microfinance that their investment would be secure. This problem was only partly solved when Dr. Makao came into contact with representatives of Kenya Agricultural Research Institute (KARI) and entered into a memorandum of understanding, in which they promised to help train the workers at Amavo and develop manuals, while Amavo would build a greenhouse at KARI for research purposes. An initial agreement on cooperation was established, but during the stay of the students no concrete arrangements were made by which the promises of KARI materialized in some or other way. Another threat to the company was the fact that employees of Amavo Greenhouses were not properly trained and that the company also depended on 'freelance' employees, who are called in as need arises (say, when there is an order of a new greenhouse). These were paid by the hour, which caused a very slow work pace.

In order to make the greenhouses affordable the CEO tries to involve microfinance institutions. All current customers were either family or close friends. This may have been one of the reasons that administration on materials was non-existent also. Although there was a small stock of greenhouse materials, there was neither a system for keeping track of these materials, nor was there a system for keeping track of tools, staff or anything else. There were no files, records or data kept about expenses except for one file with all the receipts for everything Dr. Makao had spent in the past 15 months (personally and professionally). Dr. Makao was the only one with access to the financial overviews and as the company is based on her personal finances, this seemed problematic in the future. Secondly, it was suspected that the prices set for the greenhouses (as calculated by the American students team) were too low to make a profit, but as all the expenses were not filed clearly there was no way of finding this out. Finally, it appeared that all the materials were bought on a day to day basis, so if materials were needed, they would only be purchased at the moment they were needed—there was no planning. This naturally resulted in much higher construction costs than necessary.

11.4.1.1 Case Evaluation

Actually, Amavo Greenhouses is not yet professionally run, in terms of finances, sales, marketing, R&D and staff management. This may be due to the personal management style of the CEO, but culture, values and capacities are also involved. The case shows that IESA (knowledge: information, experience, skills, algorithm), VAB (values, attitudes and beliefs) and finally SI (social environment and institutions) are in manifold ways intertwined. Actually due to the fact that the new procedure of building from standardized drawings was not accepted—of course it ran counter to traditional practices (VAB—values like voluntarism and sequential dealing with time are involved!)—knowledge and skills have been lost (IESA)! In this respect Amavo Greenhouses is not exceptional. From our experience we could add a number of examples dealing with the same problem. In spite of the fact that Dr. Makao is highly educated, she is also ‘a child of her community and culture’. She runs her company with a close introvert circle of family and friends. This is typical of a collectivist attitude. Relationships and belongingness decide whether a person is in or out. The clients too were only Kikuyu. There also has been no exploration of the needs of the customers. Actually Dr. Makao withheld the students from going into that, because she expected everything from microfinance. But the future farmers might also be provided with agricultural and operational options for making profitable use of the greenhouses. This type of marketing research had not taken place. The company identified the problem as one of lack of money. But some 10 % of the Kenyan population would be in a position to invest some €500–€1000, if only they would be sure of the revenue. Microcredit schemes in themselves do not yet serve capacity building, as it seems the company is doing something for them, rather than with them. Waiting for money often betrays an attitude of dependency and lack of initiative.

In Kenya there is no anonymous trust in public life. The concept of anonymous trust belongs to a well-developed civil society in which the default situation is to trust shopkeepers and businesspeople also if you don’t know them, until and unless they start cheating you. In Kenya it is the other way around. For that reason it is somewhat risky to hire people into the company, not because of their biological or social affinity with the CEO, but simply on account of merit and expertise. In this case the CEO had appointed nonfamily personnel but she trusted them too much as if they already were family. Some more modern Kenyan entrepreneurs do indeed like it more to hire nonfamily personnel to avoid the social claims traditionally put upon them by family members. But then again the risk is that either they are not reliable, as in this case. Or otherwise they might be bossed around to such an extent, that the employees after a while start looking for opportunities to start their own business. Talking about introducing the right set of capacities and values, it is difficult to get the right combination of hierarchy (the command structure) and egalitarianism (exchange, dialogue and loyalty). Here it appears that an enabling social environment and efficient regulation is also indispensable for success (SI). It is difficult for an individual company on its own to change the whole scene. Elsewhere we have emphasized that an entrepreneurial strategy consisting of many small initiatives from below should be supplemented by beneficial regulation from above in terms of equal access, consistent policies and coherent regulation (Kroesen and Ndegwah 2012).

Although the students did a good job in providing an organizational basis for the company, they too might have gained better results if they would have been more keen and alert in dealing with the capacity issue. They were quick in providing the company with an efficient Excel sheet for bookkeeping and quick in standardizing the production process. The question is: did the intervention of the students evolve on the basis of dialogue? There is a high risk that after a while the excel sheet is not anymore in use and the standardized building procedures are forgotten once more. The fact that change of values, attitudes and beliefs and change of the institutional environment takes a long breath should be a reason to opt for longer-term involvement. Actually this should not be the only student group involved in building up a company like this. In such a process they themselves learn how to insert themselves in the management of change, the partners abroad learn to integrate new approaches into their management style step-by-step, and the staff supervising the students builds up expertise in responsible innovation through intercultural dialogue and learning.

11.4.2 A Vocational School as a Capacity Institution—Surinam

The authors selected this case because of its duration of almost seven years (2004–2010) during which some beautiful instances of capacity building took place, although sometimes despite all the efforts even then it did not work; the case is also selected because of the changes the project brought about. And finally because of the learning opportunity the case presents concerning the insights in capacity building and responsible innovation.

11.4.2.1 Renovation of the Barronschool

From 1986 to 1992 an internal war was fought by the military and an indigenous group in the district of Marowijne (Suriname) with the village Moengo as the focal point of the fighting. Moengo is a village of about 10,000 inhabitants with roughly 10 schools, most of them at the primary level, and a vocational one, the Barronschool. During the war almost all the schools were either damaged or destroyed altogether. It has never been the general policy of the government to maintain schools or renovate them. When school buildings were too old they were just torn down and a new one would be built. Because of the dire conditions of the schools in Moengo an NGO was set up in the Netherlands to help renovate the school buildings. This type of support from the Surinam community from within the Netherlands is often considered foreign and met with distrust by the people in Surinam. This takes a tactical and careful attitude from the "Dutch" side. The concept of 'as good as new' was introduced to renovate schools properly to let them look like new buildings. In doing so the NGO set a very high standard in order to motivate the schools to go for renovation. The first school, a small one, with 8 classrooms, was renovated. The result was so positive, that the renovated primary school became the model for the other schools. The NGO then passed the need of the schools for renovation to the government in Paramaribo and the board of the schools in Moengo, all of them in Paramaribo. But there was no positive reaction, because allegedly there was no money and if they started with one school, they had to give all the other schools in all of Suriname the same treatment. But there was a pressing need for the renovation of the schools in Moengo and the schools kept repeating their request.

11.4.2.2 The Capacity Problem and Responsible Innovation

Local professionals were very scarce and would also cost a lot of money. Then the idea came up to use the students of the Barronschool, who had a technical background. The Barronschool had a student population of 300 with 100 of them studying at the technical department, which could form a good workforce. But the students lacked the (technical) knowledge: IESA, although the construction of the building was not complex, it was flat on the ground. The 27 classrooms were of the same size and built with wood, cement bricks and the roofs were of (zinc) corrugated iron. The activities to be mastered were the assessment of the condition of the building, writing a schedule of requirements, budgeting and managing the project, executing standard technical tasks and make a maintenance plan with a budget. It turned out that the vocational school had no vocational classes to create and transfer any knowledge about renovation, because the two workshops of the school were not in use anymore. There were no tools, there was no electricity and there were not enough teachers and hardly any teaching materials. There was no internship program for the students, so they lacked the experience and the skills (IESA). Students did not see any reason for attending classes, because they knew with their qualifications they could not find work anywhere. Moreover, the teachers lacked the attitude and knowledge to set up an internship program. And what lacked most of all was the belief and the conviction, that it was possible to carry out the renovation with teachers and students of the school (VAB). And since the government in Paramaribo gave all schools an equally bad treatment an enabling social environment was also lacking (SI). In short, the school did not have the capacity to carry out a renovation project. That is why from the very beginning the decision was made to create knowledge and transfer it to the teachers and the students as a key part of the project so that they could carry out the renovation themselves.

But it appeared that every teacher who could find a job somewhere else would leave the school, which happened very often. It was also common understanding by the teachers and the management of the school that their own students, were 'dumb' and wouldn't be able to learn a difficult task like renovation. Those same students would graduate and of course could not find a job, because they were unqualified. But it also became clear that the teachers themselves did not have any experience with renovation and with managing projects and that there was a lot of covering up of their own shortcomings by blaming the students. Apart from that it was their conviction that, it was the responsibility of the government in Paramaribo, hours away from Moengo, to do something about it, although it was the teachers themselves who had to work in such unfavorable conditions, living up to the saying in Suriname 'Winti wai lanti sa pai', which means 'the government always has to bear the cost for everything in the public domain'. The government agency in Paramaribo responsible for the school had to give permission to let students of the Barronschool work on the school as part of their education (internship). The management of the school needed to be convinced that this was a good learning practice for students in order to acquire practical skills. For the teachers this meant that they had to work on their own skills. The students needed to be convinced that doing internships by renovating their own school would make their education more valuable and would make it easier for them to find a job. Convincing the teachers was the most difficult challenge, because letting their students take extra classes in renovation or being supervised by students from Delft,—'who knew nothing of the construction or building in Suriname'—as some put it, was to them a sign that they themselves apparently were not qualified enough and, in addition, in the beginning they also took the attitude of 'not invented here'.

The NGO started setting up a communication structure to transfer the idea and hand over the ownership of the project to a group of key players. These were the teachers in architecture. In order to overcome their resistance they became the central project group. The second step was to get the government agency in Paramaribo so far as to give the NGO the assignment to put the idea on paper and then grant it the full authority of the project on behalf of the office of education, and also on behalf of the management of the Barronschool and the teachers. Due to lack of experience in renovation the NGO called for assistance first of the Haagse Hogeschool (department of architecture) to set up an internship office modeled after the well-established internship system of the Haagse Hogeschool and then of the Technical University of Delft to send interns to transfer knowledge to the students of the school and the teachers by carrying out and help manage the renovation of the school.

One of the continuing challenges for successful help was to walk the fine line between the transfer of knowledge, ownership, implementation and supervision. Too much supervision will undo ownership and let the project fail, but too little would also lead to failure because in that case the project would not be anchored. It is just like spices to a dish, it has got to be just enough.

11.4.2.3 Successes and Failures

The renovation program lasted seven years (2004–2010) and with almost 50 interns from the Haagse Hogeschool and mostly from the TU Delft, the school was transformed into a model of how to work on schools with students from the school itself. More than 100 students of the Barronschool did internships at renovating their own school and later students of the Barronschool were also involved in helping with the renovation of two elementary schools. And now they keep the renovation process going without outside intervention, while claiming it is their own idea, a sign of ownership. The government later started working on other schools and now most of the schools in Moengo have been renovated and renovation has become a common concept in Moengo and in Surinam.

The formal education in the technical department of the Barronschool takes three years. It took on average 1500 h per student of the Barronschool to become a technical worker on the elementary level and to be able to work on the renovation projects of the school. It took a lot of organizational effort and it was a process of dealing with tensions, misconceptions and mutual adjustment by the different stakeholders. Sometimes even bitter arguing. After the start of the internship office at the Barronschool by the NGO for instance a plan was written as an effort (Barronschool 2005) to create a common vision. It was the first time for the school to do so. Since 2005 they haven't written a new vision plan but till now they still refer to the one of 2005 as a positive experience.

After much effort from the NGO, an architect and teacher from Holland was sent on a mission for one month to help decide what needed to be done (Benneheij 2007). Then the manager of the internship office for some time organized a weekly meeting with the four teachers of the technical department where they exchanged ideas on what to do and they made plans. The initiative was good, but also soon suffered a setback. After a while they abandoned the meetings because the manager was much younger than the teachers and they felt because of his age he was not capable enough to take the lead. To keep records of the progress of the students working on different projects a portfolio system (Barronschool 2007, 2008) was introduced. In the process it was established that there were shortcomings in language and in mathematics as well. It was obvious because these were the two areas they failed in, which was why they had ended up attending the

Barronschool in the first place. Classes were organized to remedy this deficiency so they could keep the record in their own portfolio, communicate and carry out assignments. This program proved to be very successful.

There was an approach of 'outside-in and inside-out' which meant that the students had to do internships outside the school to gain experience in the outside world, but also that professionals from outside were brought inside the school to help to teach and prepare them for the outside world. One of these professionals was a playwright and stage artist who helped them improve their communication skills. A program was designed called 'Life Skills Theater' where communication concepts from theater were used. Some teachers and most of the students evaluated this as very useful, although it was a most scary thing for them, because they never did a presentation before and in yet the training it was compulsory.

One of the instruments that was particularly useful was a tracking system to deal with absenteeism and the drop-outs. This system was an important because it zoomed in on the individual students and all students knew that attention was focused on them. Another instrument was the 'senior-junior'-mentorship where older students mentored younger ones and both of them learned to deal with guidance and responsibility. Students internships supported this project. Things did not always go well, however, because sometimes certain students were overwhelmed by the culture in Surinam and went through a kind of culture shock and then closed themselves off, turning away from the difficult learning process they were in.

Also a new building, a design classroom, was built by students of the TU Delft (Gunst et al. 2007), on which the students of the Barronschool cooperated as much as the students of Delft. The students from Delft could not succeed in finishing the design classroom because they had to return to Delft, so that the school with the assistance of professionals finished it itself. The management of the school was so motivated by that new building that it gave an assignment for the design of a school library and documentation center. There was a change in attitude here. After that experience the school built a library with outside help but with full cooperation of students. Thanks to the renovation the student population of the Barronschool grew from 300 to almost 600 and the problem of drop-outs was nearly solved. Crime and vandalism virtually disappeared.

In the past nobody would work after 13.00 pm, but now it is common for certain teachers to stay longer. The director of the school made it a habit to return to school and work there in the afternoon till late. And more and more other staff members followed her example. One thing that was considered a miracle in the Barronschool was that students made use of the toilets of the school and sometimes even stayed there to have some chats and laughs, something that was unthinkable in the past because of the bad conditions and the odor of the toilets.

But still there were and are problems to solve, for example: jobs and more jobs. Employers need students with higher technical qualifications than offered by the Barronschool. Students of the TU Delft did a study with recommendations for higher education and the report was handed over to the authorities (Basak et al. 2012). They researched the extent to which the formal vocational training curriculum prepared the students for their future professions. It turned out that the learning the students received was mostly based on tradition, on the books. Although several times it was pointed out to the management of the school and the ministry of education that the school for most part was educating for unemployment if skills and experience acquired through internships were not part of the formal curriculum, nothing changed. The teachers of the Barronschool are not equipped to teach at a higher level, they do not have the capabilities and qualifications for that and they do not have the knowledge nor contact with the job market. Not wanting to wait for the government and knowing that the government often follows suit if there is a good example, an NGO has now designed a plan for technical education on agriculture on a small scale.

The school is changing over time. There are now new personal coming in and this causes discontinuity. The previous manager left the school, but took all knowledge and contacts with her, while the new manager is not so much involved in the capacity building aspect of renovation, although she had been part of the previous management team. Nevertheless, for the teachers the school is a success, but primarily because it looks much better than several years ago. All 27 classrooms have been renovated, a new library has been built and a new design classroom. There are now three workshops at the school and a computer laboratory with more than 30 computers with internet connection. From the standpoint of capacity building, however, it can easily turn into a failure if the learning process is not taken further and if there is no structural maintenance program.

11.4.2.4 Case Evaluation

As demonstrated above in the course of the project it appeared that the concept of IESA (Weggeman 1997) was not enough to cover all the aspects of capability training. The different layers of capacity as captured in the formula: IESA + VAB + SI are more suitable to explain and analyze these types of projects in terms of capacity building or the lack of it. Here we will evaluate them.

IESA: The involvement of the students of the TU Delft worked as magic because of the reputation of the university and because the students explicitly brought valuable technical knowledge from Delft. Apart from that books, readers and all sorts of practical information about renovation were sent to the school. In addition, the students brought their tacit knowledge about technology with them, their personalized knowledge (skills, experiences and algorithm, the way to process different parts of capacity or knowledge). Students of the Barronschool got the opportunity to work with the students from Delft and to do external internships. Because it was renovation in the tropics, the students of Delft also did not know everything and sometimes students from the Barronschool had to teach them a few tricks which in turn reinforced the self-esteem and thereby the learning of Barronschool students. As the students learned, so did some of the teachers, although hesitantly in the beginning.

VAB (values, attitudes, beliefs): In the creation, acquisition and transfer of the hard knowledge (IESA), culture always plays an important role. Values, attitudes and beliefs/assumptions as the soft parts of capacity were always difficult to deal with in the project, because one never sees them in space. One is confronted with them over time and sometimes they present themselves in the form of an implicit social contract (unspoken agreement), for instance the trust that an agreement is an agreement, whether to be on time or not. Interns from Delft tried changing the working hours, from 8.00 to 17.00 instead from 8.00 to 13.00, which were the regular school hours. They did so not by telling teachers and students to do so, but by doing it themselves. As a result it became common to work after school hours. They too learned from the situation, how to behave and what to say if you want to secure cooperation.

Social environment (S) and institutional framework (I) are the conditional elements of the broader society. Capacity building is taking place on the individual level but also at the organizational level. This was stimulated by the installation of a group of involved teachers, who formed a workgroup to exchange technical ideas and the students were encouraged to learn from each other. When at the end of the year the senior students left, other students were promoted to seniority so that the knowledge was secured. The newly created internship office played an important role as connector, organizer of knowledge creation and transfer station. The Barronschool was part of a larger institutional framework with its regulations, like the ministry of education with its sometimes competing departments. One department carried the overall responsibility for the school, but at the same time the vocational curriculum was supervised by another department and to get permission from both at the same time was not easy.

Building capacity on all three levels (IESA) + (VAB) + (SI) is very difficult. The crucial question is, whether internal growth is taking place in the organization and in the individuals. As stated already sometimes 'learning only takes place long after the teacher has left the scene'. The interns of Delft too had to prepare themselves in the minor International Entrepreneurship for Development for three months before they went to Moengo. Very often for students it is uncomfortable to discuss VAB-aspects, because their own VAB-aspects stand in the way, although these are the cornerstones to achieve capacity building and should be discussed openly. Nevertheless, the school organization changed over the years and this organizational change is more than the sum of individual changes. But innovation, even if it is responsible, comes with a price. Often the people involved had to go through a process of disruptive and destructive character. And one cannot always see what price the people involved had to pay for capacity building: demotivation, inferiority complex, false expectations, and disappointments.

Often lack of money is identified as the main problem. If only there would be investments, a business could take off. This can be a way to avoid the capacity problem. If the capacity is available even with little money a lot can be done. In this case it is clear that such capacity consists of knowledge and skills, but also attitudes. It is an attitude of dependency, traditionalism and uncertainty avoidance which makes people look at beneficial or less beneficial external conditions instead of relying on determination, labor and initiative.

For the NGO involved in the case of Surinam, but also for the students in the Amavo greenhouse project it is difficult to judge when and to what extent to intervene. A sense of timing is all important. The NGO in Surinam tried to convince all the stakeholders tactically and consistently to participate in new initiatives. And because this intervention was repeated over time and confirmed by the input of many students through the years, it brought gradual change. Although at the invitation of their hosts, the students in Kenya improved the internal processes of the business by means of Excel sheets, standardization of drawings, it still is highly questionable whether this effort will have a lasting effect. Students in Surinam did the same for a craft shop (Starre et al. 2013) also with stakeholder dialogue and participation of the crafters related to the craft shop. But who will tell, whether this leads to lasting change? How big is the risk, that these new procedures may be soon forgotten, only because the stakeholders fall back into old habits and the new procedures did not really become internalized? As already stated, it is like spices in a dish: intervening can easily turn out to be a matter of over-spicing. How much change is manageable? The underlying questions are: how to internalize modern values while keeping respect for old traditions? And how to do so not for, but together with the people involved as a common learning process? Again, these are the questions in which the issue of responsible innovation in many development contexts materializes.

11.5 Conclusion

The lessons learned from the conceptualization of capacity in combination with the practical experiences justify the conclusion that capacity development should receive much more attention in technology transfer and (re-)design for developing countries. Capacity training along with the introduction of technology really is a matter of responsible innovation. It is an ethical issue as well. The point is that an equilibrium between competing values needs to be found, which at the same time respects tradition and welcomes change.

Capacity building and training does not only involve skills and competencies (IESA), but also cultural values, attitudes and beliefs (VAB). These values, attitudes and beliefs in turn need to be rooted in the social environment and institutions, which they nurture and by which they are sustained (SI). To reach a comprehensive definition of capacity the authors proposed capacity as the function of these three: $C = f[(IESA) + (VAB) + (SI)]$.

The authors emphasized the value laden aspect of this understanding of capacity. It is the VAB part which is often not taken into account. Values, attitudes and beliefs are at the core of what we are able to do. The right order and alternation between different values is a matter of responsible choice, dialogue and deliberation. Little is as dear to people as the values and beliefs they adhere to.

Therefore it takes a tactical and careful process in order to find the right path forward, following the rhythm of the partners in the process abroad. Failure is more easily brought about than succes. If such sociocultural innovation is not conducted responsibly the negative effect may easily be that the interventions done in the name of development only leave behind a society in disruption, in which some presumably outdated modes of behavior have been broken down, but nothing new has been created. Or the other way around, in the name of presumably indigenous African solutions no innovation may take place at all, leaving Africa behind in traditionalism and poverty. The Amavo case shows that not to change is not an option, but at the same time that it is the most difficult thing to intervene timely and adapt to the rhythm of change and learning of the people involved. Timeliness is everything. It needs for instance to be decided not whether, but when the African time rhythm, prioritizing the present moment above future goals, is more expedient than Western-style planning. If it is necessary to reach a production target in order to be competitive, there is a need for planning, labor and prioritizing the future goal. But nervous Western managers could learn a lot from the African time rhythm at the moment they need to create a common support base with mutual understanding for actions to undertake. In the same vein African collectivism and hierarchy could learn a lot from the Western attitudes of egalitarianism and individualism. Too strict hierarchy causes irritations and makes people anxious to look for other jobs. But on the other hand Western companies as well cannot do without some level of hierarchy and efficient action to meet the production targets. They also need some sense of belongingness for their members, even if they will not reach the level of African communitarianism by far. Both in the West and in more traditional societies the real question is not either/or, but to what degree and in which alternation these different values and attitudes need to be exercised. And this question—when and how much?—should be dealt with by means of dialogue and by means of a path and context dependent approach. This dialogue, however, between tradition and innovation is not invented by our generation. The exchange is already going on for a long time and the priority struggle between different value sets is taking place as much in developing countries as the so-called developed countries. These as well are not just "arrived" but still fully in development.

Whereas in itself no particular value system can be claimed as dominant, this can only become true and effective, if all partners in the dialogue recognize and respect at least one value as decisive. This is the openness for change at the right time. The openness for change is the only meta-value, the power to intervene and to stop at the right moment, with an acute sense of timeliness. This value—or should we name it a capacity?—facilitates learning processes and growth. Interventions towards change should take shape within an ongoing dialogue, taking long-term involvement and commitment and exercising careful listening to the target group by community immersion (Ndegwah 2011). There is not one particular value set, which is more modern or justified than another, but the (meta-)capacity to alternate between different values and modes of behavior and cooperation depending on time and circumstances is the decisive kernel point of the process.

This value turns otherwise irreconcilable value priorities into a common repertoire for peace and justice.

These insights and conclusions open a new practice-based research agenda. Capacity training for development as a process of deep learning takes a long-term and a step-by-step approach of learning and doing by the trainer and the project staff. In this respect capacity building and technological innovation appear to be irreducibly intertwined. Learning does not take place in a vacuum, but by implementing new technology and dealing with it in the process of step by step growth. Solutions that work cannot be invented by design or behind a desk. Instead action research is required in a constant feedback loop between theory and practice and between innovation of technology and training of capacity. From our past experiences we can learn how to avoid the failures of the past, but this leaves a large field open of risk and opportunity. The road towards the future cannot be found without a true *experimentum vitae.* And this experiment is at the same time a human value oriented experiment and a technical innovation process. Further research should elaborate which capacities are crucial for development and in enhancing an "economic culture" (Porter), cultivating capacities for responsible growth. Further research should also explore how the technical innovation process can best be aligned to capacity growth. In different cultural settings, different (always temporary) solutions can be found. In this tortuous way a world society is emerging on the basis of respect and the capacity and the willingness to change and to innovate—responsibly. Can we call this tortuous way progress? We may hope so.

References

Arond, E., I. Rodriguez, V. Arza, F. Hewrrera, and M. Sanchez. 2011. *Innovation, sustainability, development and social inclusion: Lessons from Latin America*, STEPS Working paper 48, Brighton: STEPS Centre.

Balassanian, D. 2006. *Institutional reform and change management: Managing change in public-sector organizations*. A UNDP capacity development resource. www.capacity.undp.org (Retrieved May 10, 2013).

Barronschool. 2005. *Visieplan Barronschool 2005–2010*. Report about the vision of selectable a sunny and the school.

Barronschool. 2007. *Handleiding portfolio*. A report for students on how to keep a portfolio.

Barronschool. 2008. *Eindrapport pilotproject portfolio Barronschool 2008*. An internal report for the Barronschool on the results of the portfolio in a project.

Basak, A., A. Kasiem, and R. Rampersad. 2012. *Onderwijs Techniek Barronschool*. Student internship report TU Delft.

Bell, M. 2006. Time and technological learning in industrializing countries: How long does it take? How fast is it moving (if at all)? *International Journal of Technology Management* 36: 25–39.

Bell, M., 2009. *Innovation capabilities and directions of development*. Steps working paper 33. Brighton: STEPS Centre.

Benneheij, P. 2007. *Renovatieplan Barronschool.* Report for the Barronschool. Moengo.

Cardwell, D. 1995. *Wheels, clocks, and rockets: A history of technology*. New York: Norton & Co.

Chang, H.J. 2007. *Bad samaritans—The guilty secrets of rich nations and the threat to global prosperity*. London: Random House.
Cherim, E., L. van Reemst, and N. Kallen. 2013. *Affordable Greenhouses—Kenya*. Student internship report TU Delft.
Christensen, C.M. 2003. *The innovator's dilemma: The revolutionary book that will change the way you do business*. New York: Harper Collins.
Daniels, L. 1999. The role of small enterprises in the household and national economy in Kenya: A significant contribution or a last resort? *World Development* 27(1): 55–65.
Eade, D. 1997. *Capacity-building: An approach to people-centered development*. Oxfam Development Guidelines (available at Google books).
Etounga-Manguelle, M. 2000. Does Africa need a cultural adjustment program? In *Culture matters—How values shape human progress*, ed. L.E. Harrison, and S.P. Huntington, 65–77. New York: Basic Books.
Field, M., R. Hitchens, and M. Bear. 2000. Designing BDS interventions as if markets matter. In *Microenterprise best practices project.* Washington DC: USAID.
Gibson, A., R. Hitchens, and M. Bear. 2001. BDS market development: A guide for agencies on the emerging development approach to BDS. In *Microenterprise best practices project*. Washington DC: USAID.
Grondona, M. 2000. A cultural typology of economic development. In *Culture matters—How values shape human progress*, ed. L.E. Harrison, and S.P. Huntington, 44–56. New York: Basic Books.
Gunst, C., K. Meijer, and V. Nowee. 2007. *Realisatie tekenlokaal op de LBGO Barronschool te Moengo*. Student internship report Delft.
Hofstede, G. 1991. *Cultures and organizations, software of the mind*. London: McGraw-Hill.
James, R., and J. Hailey. 2007. *Capacity building for NGOs: Making it work, a survey of current northern NGO approaches, OPS 5*. Oxford: International NGO Training camp, Center (INTRAC).
James, R., and R. Wrigley. 2007. *Investigating the mystery of capacity building: Learning from the praxis program*. Praxis paper 18, Oxford: International NGO Training Research Center (INTRAC).
Joas, H. 1999. *Die Entstehung der Werte*. Frankfurt: Suhrkamp.
Katz, J. 2001. Structural reforms and technological behavior—The sources and nature of technological change in Latin America in the 1990s. *Research Policy* 30: 1–19.
Kim, L. 1997. *Imitation to innovation*. Cambridge Mass: Harvard Business School Press.
Kroesen, J.O., and D.J. Ndegwah. 2012. *Capacities and governance in Kenya, lessons in technology transfer.* CESUN conference, Delft, http://cesun2012.tudelft.nl/wiki/index.php/WB (Retrieved August 13, 2013).
Lall, S. 1992. Technological capabilities and industrialisation. In *World development* 20, no. 2, 165–186. Great Britain: Pergamon Press plc.
Lusby, F., and H. Panlibuton. 2002. *Subsector/business service approach to program design*. Washington DC: USAID.
Maturana, H.R., and F.G. Varela. 1980. *Autopoiesis and cognition; the realization of the living*. Dordrecht, Netherlands: Reidel & Co.
Moyo, D. 2009. *Dead aid, why aid is not working and how there is another way for Africa*. London: Penguin Books.
Ndegwah, D.J. 2006. Understanding the nomads: The role of culture in evangelisation. In *The pastoralists: A challenge to churches, state, civil society*, ed. Francesco Pierli, et al. Nairobi: Paulines Publications Africa.
Ndegwah, D.J. 2011. The relevance of philosophy to the development of Africa. In *Rethinking integral development in Africa*, ed. David W. Lutz, Paul M. Shimiyu, and George Ndemo Osengo. Nairobi: Consolata Institute of Philosophy Press.
Nelson, R., and E.S. Phelps. 1966. Investment in humans, technological diffusion and economic growth. In *The American Economic Review* 56(1/2): 69–75.

Nelson, R.R., and H. Pack. 1999. The Asian miracle and growth theory. In *The Economic Journal* 109(457): 416–436.
Nussbaum, M.C. 2006. *Frontiers of justice, disability, nationality, species membership*. Cambridge: Harvard University Press.
Nyasani, J.M. 2010. *Philosophy of development: An African perspective, reflections on why Africa may never develop on the western model*. Nairobi: Consolata Institute of Philosophy Press.
Owen, R., et al. (eds.). 2013. *Responsible innovation: Managing innovation in society*. Hoboken: Wiley.
Porter, M.E. 2000. Attitudes, values, beliefs, and the microeconomics of prosperity. In *Culture matters—How values shape human progress*, ed. L.E. Harrison, and S.P. Huntington, 14–28. New York: Basic Books.
Rosenstock-Huessy, E. 1993. *Out of revolution—Autobiography of Western Man*. New York: Argo Books (original 1938).
Sachs, J.D. 2005. *The end of poverty*. New York: Penguin Press.
Samli, A.C. 2009. *International entrepreneurship, innovative solutions for a fragile planet*. New York: Springer.
Schumpeter, J. 1942. *Capitalism, socialism and democracy*. New York: Harper & Bros.
Starre, S., S. Boute, and E. Leising. 2013. *Koni Mariwina Pikin*. Student internship report TU Delft.
Sutcliffe, H. 2013. *A report on responsible research and innovation.* http://ec.europa.eu/research/science-society/document_library/pdf_06/rri-report-hilary-sutcliffe_en.pdf. Consulted 16 May 2013.
Trompenaars, F., and C. Hampden-Turner. 1999. *Riding the waves of culture*. London: Brealey.
UNDP. 1997. *Capacity development*. New York: UNDP report, Management Development and Governance Division.
UNDP. 2001. *The human development report 2001: Making new technologies work for human development*. New York: Oxford University Press.
UNDP. 2006. *Institutional reform and change management: Managing change in public-sector organizations*. Geneva: Bureau of development policy, United Nations development program.
Weggeman, M.P. 1997. *Kennismanagement: inrichting en besturing van de kennisintensieve organisatie*. Scriptum: Schiedam.
Weggeman, M.P. 2000. *Kennismanagement: de praktijk*. Scriptum: Schiedam.
White Jr, L. 1962. *Medieval technology and social change*. London: Oxford University Press.
World Bank report. 1989. Cited in *Civil Society and Governance in Ethiopia*, Asnake Kefale and Dejene Aredo, p. 103 in *Good Governance and Civil Society Participation in Africa*, 2009, Organization for Social Science Research in Eastern and Southern Africa (OSSREA).

Part III
Applications—Case Studies in Responsible Innovation

Chapter 12
Responsible Innovation in Practice: The Adoption of Solar PV in Telecom Towers in Indonesia

Andri D. Setiawan and Rajbeer Singh

Abstract The adoption of solar PV in telecom towers is considered as a sustainable innovation in powering the towers. Such adoption involves different actors which hold different tasks and responsibilities. However, the inclination that comes up among the actors involved is that if and when a disaster occurs (e.g. telecom systems collapse) due to the failure in the operation and management of solar PV after its installation, they will most likely be prone to finger-pointing at each other. In such situation, the questions arise: how to locate the accountability, who should be accountable to whom and for what, and how to proportionally distribute the accountability? Through a case study in Indonesia, this paper discusses and analyzes how to address such issue by undertaking stakeholder and impact analysis and analysis of five dimensions of responsible innovation namely: anticipation, reflexivity, responsiveness, deliberation, and participation. Our analysis suggests that there is a need of an innovative way or technology that can proportionally distribute the accountability of actors involved. The paper also concludes that innovation without responsibility faces dangers of rejection or less acceptability of even greener and sustainable technology.

Keywords Accountability · Responsibility · Adoption · Dimensions of responsible innovation · Innovation · Solar PV · Telecom towers

A.D. Setiawan (✉)
Eindhoven Centre for Innovation Studies (ECIS), Department of Industrial Engineering & Innovation Sciences, TU Eindhoven, Den Dolech 2, 5612 AZ Eindhoven, The Netherlands
e-mail: a.d.setiawan@tue.nl

R. Singh
Department of Values and Technology,
Faculty of Technology, Policy and Management,
TU Delft, Jaffalaan 5, 2628 BX, Delft, The Netherlands

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_12

12.1 Introduction

Innovation often faces many challenges for its adoption. One major challenge that has increasingly gained attention is the issue of responsibility of innovation actors. The innovation adoption process can end up in failure when responsibility is neither located properly nor clearly defined. Moreover, in the adoption of innovation where various actors are involved, it is often difficult to locate responsibility when a disaster occurs (van de Poel et al. 2012). In addition, such failure can be caused by the inability of actors involved in the adoption process to foresee the unexpected impact of innovation. As such, the ability to foresee undesirable consequence is regarded as a condition for responsibility (Nihlén Fahlquist 2006; van de Poel et al. 2012).

Therefore, understanding the impact of innovation is essential for a responsible adoption of innovation. With regards to sustainability, the impact of innovation usually touches upon three aspects: environment, social, and economy. Considering these three aspects can lead not only to the successful adoption of innovation but also enable such innovation to become responsible and socially sustainable. In this paper we relate the meaning of responsibility with accountability. In particular it highlights the tasks and responsibilities of actors involved in the adoption of innovation whereas it is imperative that every actor should be held accountable for his tasks in such adoption. Hereafter, the question arises: how to ensure the accountability of innovation? To answer such question, the concept of responsible innovation and its application in practice are becoming increasingly relevant and important to gain further understanding. Therefore, in order to gain a deeper understanding of responsible innovation and to carry out our reflection on the concept, we take a combination of the following approaches: study of literature and case study. The case study in this paper particularly aims to show how responsible innovation as a process-based approach as defined by Koops (Chap. 1, this volume) could be operationalized in practice.

In this paper we will discuss and analyze a case in detail: the adoption of solar photovoltaic (PV) for telecom towers in Indonesia. The reasons for taking this case are twofold. The first is that solar PV is a good example of innovation in renewable energy technologies since it has been globally adopted as the alternative for the provision of off-grid electricity in the remote areas. The adoption of solar PV contributes to solving the power availability problem of telecom towers in the remote areas. Yet, such adoption might also contribute to the occurrence of a disaster in telecommunication system when the actors involved are unable to deliberately anticipate the cause of system failure such as the field related factors (i.e. local climate condition) from the beginning of the adoption process. The indication that arises is that if and when such disaster occurs, the actors involved will most likely be prone to finger-pointing at each other. In such situation, it would be difficult to locate the accountability of actors involved, and it is becoming unclear about who should be accountable to whom. Furthermore, it also reflects that such accountability is not proportionally distributed among actors involved.

In this regard, the paper will address the following questions: (1) How to locate the accountability? (2) Who should be accountable to whom and for what? And (3) How to proportionally distribute the accountability? To a greater extent, these questions are also relevant to be addressed in the (non)adoption of technology in different contexts.

The second reason is that the concept of responsible innovation is still relatively new for developing countries and the application of the concept needs a special study (Singh and Kroesen 2012). As suggested by de Jong et al. (Chap. 4, this volume) what is needed is the application of responsible innovation as a guiding concept within the specific context in which it is used. Developing countries like Indonesia are adopting many new technologies, like solar PV, and the adoption of such technology is considered as an important part of transitions to renewable energy. Taking into account the concept of responsibility in the adoption process will give opportunity to manage the consequences so that the adverse or unexpected impact of innovation can be minimized. And thus to a greater extent it might also help to foster a responsible energy transitions. Furthermore, it is expected that the paper would contribute to the enrichment of the concept of responsible innovation applicable to developing countries context. Since the concept of responsible innovation is still evolving and different meanings are at play, as shown by Koops (Chap. 1, this volume), in Sect. 12.2 we will start discussing its definition and propose our working definition as the frame for analyzing the case study. Section 12.3 will then discuss and analyze the case study in detail. This section describes the application of the concept in practice through problem identification, stakeholder analysis focusing on the tasks and responsibilities of actors involved in the adoption process, analysis of possible impact of innovation in terms of environment, social, and economy, and the finding of alternatives proposed solution or way-out for the problem. Finally Sect. 12.4 provides conclusions and outlines areas for future research.

12.2 Defining Responsible Innovation

The phrase responsible innovation has been properly defined in technology assessment and in science and technology policy field (Guston and Sarewitz 2002; Schot and Rip 1997). However, the definition of responsible innovation is still evolving. Instead of just focusing on pre-established definition of responsible innovation, we try to unfold this notion in order to frame our case study in this paper.

Responsible innovation reflects the connection between innovation and responsibility. The definition and interpretation of innovation have been widely discussed in literature (see e.g. Dosi 1988; Edquist 1997; Lundvall 1992; Rogers 2003; Schumpeter 1961). In economic terms, innovation is usually interpreted as the activity in which a new product and method of production are introduced, a new market is opened, and new organisation of any industry is carried out (Schumpeter 1961). Following this interpretation, Dosi (1988) described innovation as "…the search for, and the discovery, experimentation, development, imitation, and adoption of new products, new production processes and new organisational set-ups". The aforementioned definitions imply that what is meant by innovation is not limited to technological innovation, but rather constitutes the novelty of products, processes, organisations, and markets cf. the proposed broad definition in Blok and Lemmens (Chap. 4, this volume).

In a broader social context, innovation embraces any kind of changes that create certain kind of values for society while bringing any products, ideas, methods, and any other objects into the market, either radically or incrementally (Acosta and Hoolbrok 2012; Mytelka and Smith 2002; Rogers 2003). Innovation emerges not only new kinds of techniques, but ultimately it gives birth to new social practices that enable people improving their quality of life. By bringing new things into the market, innovation is seen as a future-creating activity that changes the market as well as the society itself (Grinbaum and Groves 2013). Crossan and Apaydin (2010) comprehended this by considering innovation both as a process and as an outcome of that process. They abridged innovation as "*production or adoption, assimilation, and exploitation of a value-added novelty in economic and social spheres; renewal and enlargement of products, services, and market; development of new methods of production; and establishment of new management systems. It is both a process and an outcome*."

Further, while innovation is expected to help creating a better future for the market, but the adoption of it might give unexpected outcomes. What emerges from an innovation is not always positive impact from the very start, but sometimes initial negative impact which is something unexpected by the market as well as by the society. And with regards to sustainability, in most cases, the impact of innovation touches upon three aspects: environment, social, and economy.

Concern on the (unexpected) impact of innovation is therefore relevant to be acknowledged as one of the reasons to associate the concept of responsibility with innovation. Such concern entails a forward-looking responsibility of innovation actors on what they create and its impact. This also reflects the conceptual understanding of responsible innovation. The term responsibility is usually referred as a duty, liability or obligation; to be responsible may mean to cause and reflecting one's accountability (Ubois 2010) and to honor those obligations when needed. One should act reliably in accordance with promises made. Therefore, as it is associated to innovation, responsibility should exist throughout the innovation process, from initial concept to application, from inventors to end users, as well as to engineers and vendors (Ubois 2010). This also implies that innovation actors should be accountable for the effects created by the roles and tasks they hold in the adoption of innovation. Further, since innovation is considered both as a process and as an outcome, the nature and quality of the processes involved in the adoption—especially processes to establish responsible behaviours among the actors involved at a very early stage in the adoption process—could determine in a major way the success of the ultimate adoption of the innovation. Also, as noted by Balkema and Pols (Chap. 15, this volume) a clear demarcation of responsibilities among actors involved is one of the pre-conditions of the success of innovation adoption.

As we have noted, innovation creates certain kind of values to society. Regardless of certain values are contextually universal, but since every society has a different culture, values can be perceived in many different ways. Western society perceives and appreciates values in a different way to the eastern society. The values that are appreciated by the western or developed part of the world may not be appreciated by other cultures, though those values are the same (Singh and Kroesen 2012). Differences are even present within the western and eastern societies. Therefore, the appreciation of such value differences in innovation can lead to better adoption and positive outcomes (Singh and Kroesen 2012). This is becoming more relevant to be discussed, especially when we talk about the adoption of technology in developing countries. Because, when the western or developed world invents or discovers something, there is still the question on how the developing world should adopt the innovation. And whether the values created from such innovation are culturally appreciated by or compatible with the developing world. In many events such adoption entails cross-cultural technology transfer, which is not a simple process. This means that the concept of responsible innovation should also well address the cross-cultural context to make it applicable to different cultural settings and contexts such as to developing countries context. It is therefore noteworthy to argue that an innovation is responsible when it has sensitivity to local socio-cultural values in which the innovation is adopted.

Furthermore, as explained by Singh and Kroesen (2012), responsible innovation means "*being caring or ensuring care for certain values for social, economic and environmental sustainability by engaging in anticipation, reflexivity, deliberation, responsiveness and participation for bringing up any change in any idea, product, process, method, way of doing business, technology, et cetera in order to bring them into a specific market or use them in a society*". Being caring means being sensitive or considerate, ensuring care means to make certain and assume responsibility, and certain values indicates a list of values including universal and culturally specific values. Meanwhile, anticipation, reflexivity, deliberation, responsiveness, and participation are considered as the five dimensions of responsible innovation. Anticipation denotes the act of looking forward and dealing with risks beforehand. Anticipation also indicates the act to foresee the plausible impact of innovation. Therefore, the knowledge about foresight and its methodologies helps in understanding innovation as well as the impact and possible applications of it. Furthermore, such understanding can help the decision making process of the adoption of innovation and other related activities. Reflexivity has a meaning of being reflexive or refers to cause-effect relationships. It indicates a circular or iterative process of creating and shaping innovations. Deliberation refers to a thorough exploration process and a careful consideration of different aspects and discussions in order to find a way forward. Responsiveness means being able to readily respond or address such circumstances due to different needs, requirement, views, issues, and values. Participation denotes the act of taking part or involve in something which indicates the involvement of different stakeholders as the condition of responsible innovation.

Based on above elucidation and in particular the definition by Singh and Kroesen (2012), we propose the following working definition of responsible innovation for the purpose of the case study: responsible innovation means ensuring the accountability of innovation actors (the actors involved in the adoption of innovation) through the engagement of anticipation, reflexivity, responsiveness, deliberation and participation in the adoption of innovation while looking the impact of innovation on three aspects: environment, social, and economy. This working definition will be used as the frame for analyzing our case study in this paper.

12.3 Case Study: The Adoption of Solar PV in Telecom Towers in Indonesia

For the purpose of this study, we obtained data from literature and interviews. Five telephone interviews were conducted with practitioners in the telecom and PV sectors in Indonesia especially those who involve in the adoption of solar PV in telecom towers.

12.3.1 Background

Telecom tower or Base Transceiver Station (BTS) tower is the main telecom infrastructure in Indonesia. Such infrastructure is highly critical for telecom operators in serving more than 240 million mobile telecommunication customers (MCIT 2011). The country has more than 95,000 telecom towers in 2011 and most of them are located in the rural and remote areas (Latif 2012). In order to ensure reliable data transmission for telecommunication services, the towers operate non-stop which makes them energy intensive. Therefore, in the remote areas with no connection to the main grid, power availability becomes the major challenge. To cope with this challenge, many telecom operators depend heavily on diesel generators (DG) to power telecom towers.

However, some issues arise from the use of DG, such as the uncertain supply of diesel in the remote areas. The difficulty in transporting diesel to the remote areas has made the price of diesel in these areas very expensive where the price could be three to five times to its normal price. Another issue is about the pollution and noise produced from DG operation. These issues pose some major hurdles in operating DG and create problem for the environment. Therefore, due to these issues, many telecom operators had been evaluating to gradually replace DG by adopting more sustainable and environment friendly technology such as solar PV for powering the towers (Kompas.com 2010; Latif 2012; Sriram 2009).

The adoption of solar PV in telecom towers in Indonesia started in 2005 with about 60 solar PV powered telecom towers (Sriram 2009). The number has been growing and was expected to considerably increase to more than 4000 by 2010

(Sriram 2009). Recently, more than 150 have been newly installed by some telecom operators in Indonesia (Mitratel 2010). The adoption of solar PV gives several advantages for the operation of telecom towers since it can provide supplementary power and solve other problems related to diesel. These are for instances: solar PV has lower operational cost compare to DG, less pollution in terms of CO_2 emissions, noiseless since it produces no sound in the operation (Maksudi et al. 2012; Mitratel 2010), and most importantly it reduces the risk of telecommunication systems collapse due to lack of supply of power. Such adoption is considered as an innovation for powering telecom towers in a sustainable way (Maksudi et al. 2012; Sriram 2009). In the long run it is expected that solar PV and other renewable energy technologies can replace fossil based generators for powering telecom towers in the country.

12.3.2 Problem Description

Telecom operators are already aware of the kind of benefits and advantages from the adoption of solar PV in telecom towers. Despite the advantages, some issues in the post-deployment are still remaining notably the field related factors such as the local climate conditions and the lack of good cleaning mechanism. PV modules need to be periodically cleaned, especially those in the remote areas. In the tropical country like Indonesia, the temperature and humidity are ideal for rapid accumulation of dust in PV modules. This could be worse with the presence of carbon and soil particles, acid components, birds dropping and other organic matters that are not properly maintained with a good cleaning mechanism (e.g. skills of cleaning, the way to transport water to remote areas, et cetera).

Due to above factors, the power production or output of PV modules is declining. The impact of such factors on solar PV performance has been studied in many researches (see e.g. El-Shobokshy and Hussein 1993; Elminir et al. 2006; Hassan et al. 2005; Hegazy 2001; Mani and Pillai 2010; Zorilla-Casanova et al. 2011). A recent study shows that the daily loss along a year caused by dust accumulation on the surface of a PV module varies, but the range could be between 4.4 and 20 % (Zorilla-Casanova et al. 2011). This could be higher when there is no rain, which to some extent acts as a natural cleaning mechanism (Zorilla-Casanova et al. 2011). In this study, we take an example of data drawn from a solar PV module adopted in a telecom tower located in an area in West Java. The tower consumes at about 26 kWh (kilo-Watt hour) of electricity per-day with a maximum power required of less than 1500 W for powering a microwave transceiver and two BTS equipment installed (3G and 2G units) including lighting. The tower is equipped with a polycrystalline PV module consist of 68 units of 200 Wp (Watt-peak) solar panels. So the maximum total power output of the system (TPOS) is 13,600 Wp. The tower is also backed up with a DG. Our findings indicate that the average power output of the PV module/system is declining as shown in Fig. 12.1 (Source: author's personal interview with a tower company. Due to privacy reason,

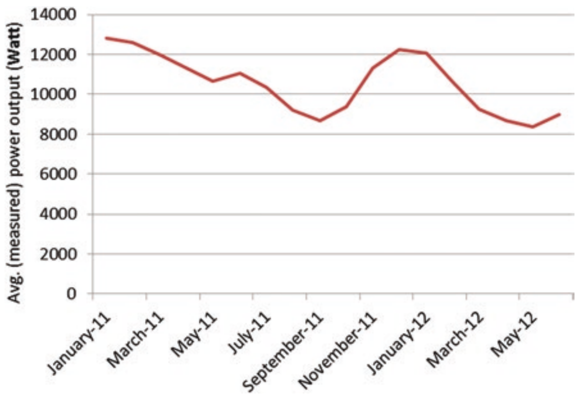

Fig. 12.1 Average power output of the PV module

name of the person and his company cannot be mentioned here). By comparing the average power output of the PV system between two periods in January–June 2011 (11,738 W) and January–June 2012 (9661 W), we can figure out that there is a declining at about 17 %. From the data, it is also apparent that the monthly average power output of the module in the rainy season (i.e. November–April) is about 80 % of TPOS, but in the dry season (i.e. May–October) it is less than 73 % of TPOS.

The problem arises: while the power consumption of the towers remains constant but the power production decreases, this will lead to the nonfunctioning of the towers. At the end it can lead to the collapse of telecommunication systems since the towers have insufficient power for operation. Surely such disaster is not expected to occur by telecom operators and the users of telecommunication services since the impact will be very severe.

However, in fact, so far the field related factors seem to have gotten only little attention. These factors are neither properly handled, nor anticipated by the actors involved from the beginning of the adoption process. This is as indicated from our findings that so far the cleaning of PV modules in most telecom towers in the remote areas is only performed once a year, especially in the dry season. As such, the adoption of solar PV in telecom towers involves several actors where each of them actually holds different tasks and responsibilities. But, seemingly the risk of telecommunication systems collapse due to field related factors have not yet been seriously deliberated by telecom operators and other actors involved in the adoption of solar PV. The indication that comes up among the actors involved is that if and when such disaster occurs because of these factors, they will most likely be prone to finger-pointing at each other regarding the liabilities in the adoption of solar PV.

While it is imperative for each of actors involved to be held accountable for their tasks, but when the incident occurs it would be difficult to pinpoint the accountability. Apparently, it would also become unclear about who should

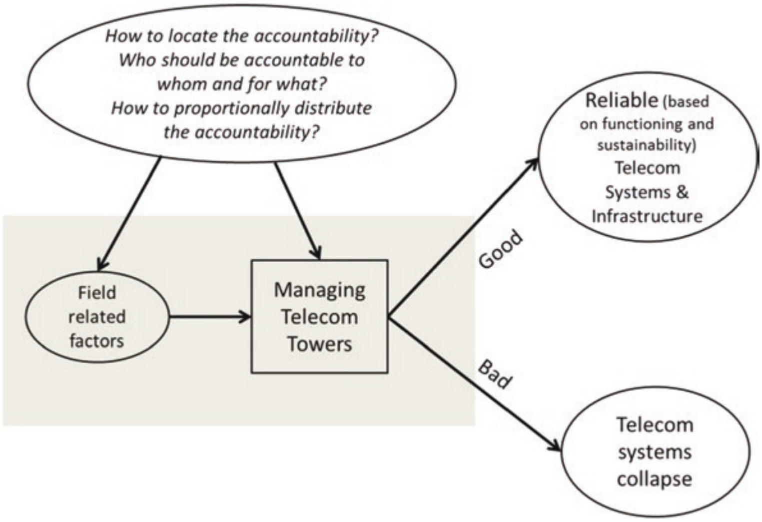

Fig. 12.2 Problem description of the case study

be accountable to whom and for what. Furthermore, such accountability is not proportionally distributed regarding the tasks should be performed by each of actors involved. This has left the issue in the adoption of solar PV in telecom towers, and thus in a broader context about the responsibility in managing telecom towers. Figure 12.2 portrays a succinct problem description of the case study.

In order to further analyse this case, in the next section we will undertake stakeholder and impact analysis. This will help us in mapping out the tasks and responsibilities of actors involved in managing telecom towers and to foresee the possible impact of solar PV adoption. Thus, it will also help us in figuring out the kind of solution or way-out that can be proposed.

12.3.3 Stakeholder and Impact Analysis

Stakeholders can be generally defined as actors that can influence or are affected by a certain problem, decision or action (Chevalier 2008; Freeman 1994). Applying this definition to our case, stakeholders are considered as actors who are involved in the adoption of solar PV in telecom towers as well as in managing telecom towers where that involvement may affect the collapse of telecommunication systems. Drawing on data from the literature and interviews, we identify and classify at least five groups of relevant actors involved in the adoption of solar PV in telecom towers that should be held accountable for their tasks, namely: telecom operators (e.g. Telkomsel, Indosat, BTel, XL-Axiata), tower infrastructure companies (e.g. Indonesia Tower, Mitratel), solar PV manufacturers (e.g. LEN, Inti, Solarens), solar PV installers (e.g. Imprima, Hariff), and regulator (Indonesian Telecommunication Regulatory Authority (BRTI)). Following actors' identification, we also identify important tasks and responsibilities related to managing telecom towers as depicted in Fig. 12.3.

As mentioned in the previous section, the field related factors pose a threat for the adoption of solar PV in telecom towers as well as for managing telecom towers. Such factors may cause the collapse of telecommunication systems if the actors involved are unable to deliberately anticipate them in the adoption. There is an indication that if and when such disaster occurs due to these factors, the actors involved will most likely be prone to finger-pointing at each other. Consumers and regulator will be the first parties that finger-pointing at telecom operators and asking for their responsibilities for the occurrence of such disaster. Most likely telecom operators do not want either to be entirely blamed, or to take the entire burden by themselves. They will be prone to point their fingers at tower companies as the parties who perform the maintenance of the towers. However, in order to deal with the field related factors, an extra (cleaning) task is needed and it requires extra maintenance cost. It raises a question: who should pay for it? As a response, tower companies tend to point their fingers back at telecom operators and saying that such task is not ours because the existing contract does not cover the cost for the extra task. They also think that such cost should be provided by telecom operators.

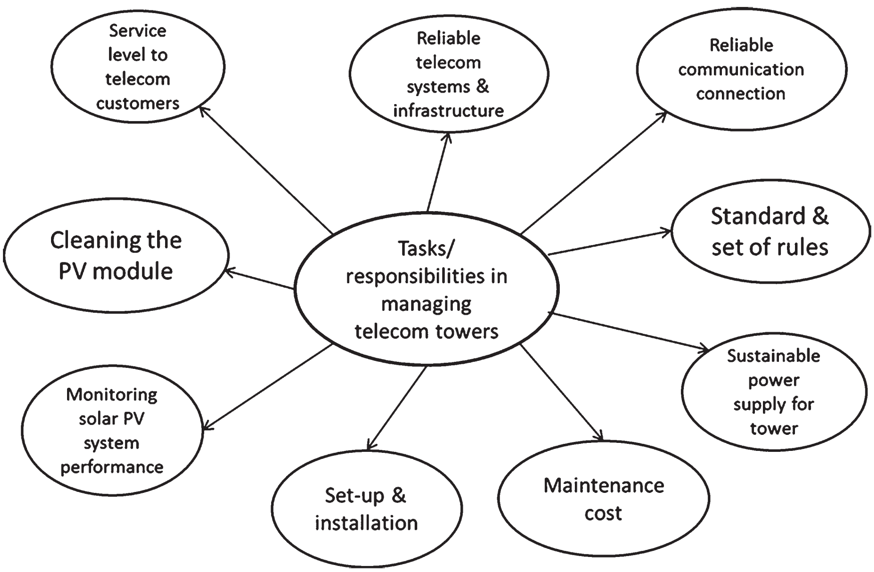

Fig. 12.3 List of tasks/responsibilities of actors involved in managing telecom towers

Hence, there is no sense of liability for tower companies to perform such task. Following this response, tower companies could also point their fingers at solar PV manufacturers as the parties who provide the PV modules. As indicated in the interviews, people from the tower companies commonly said, “Cleaning task won’t be periodically needed if the PV modules can endure with the local climate conditions, and it is the responsibility of PV manufacturers to provide such technology”. The finger-pointing does not end at this point, because tower companies as well as solar PV manufacturers could also point their fingers at solar PV installers who perform the setting and installation of PV modules. Such task can determine the operation and performance of PV modules, e.g. adequate training for tower operators concerning maintenance. Furthermore, the finger-pointing could also reach at regulator who should provide regulation or standard protocol for managing telecom towers. Yet, such specific standard protocol for managing telecom towers powered by solar PV and other renewable energy technologies is still absence in Indonesia. Telecom operators could also finger-pointing to the customers since they want only cheap price but are not willing to pay more for better service and maintenance. Figure 12.4

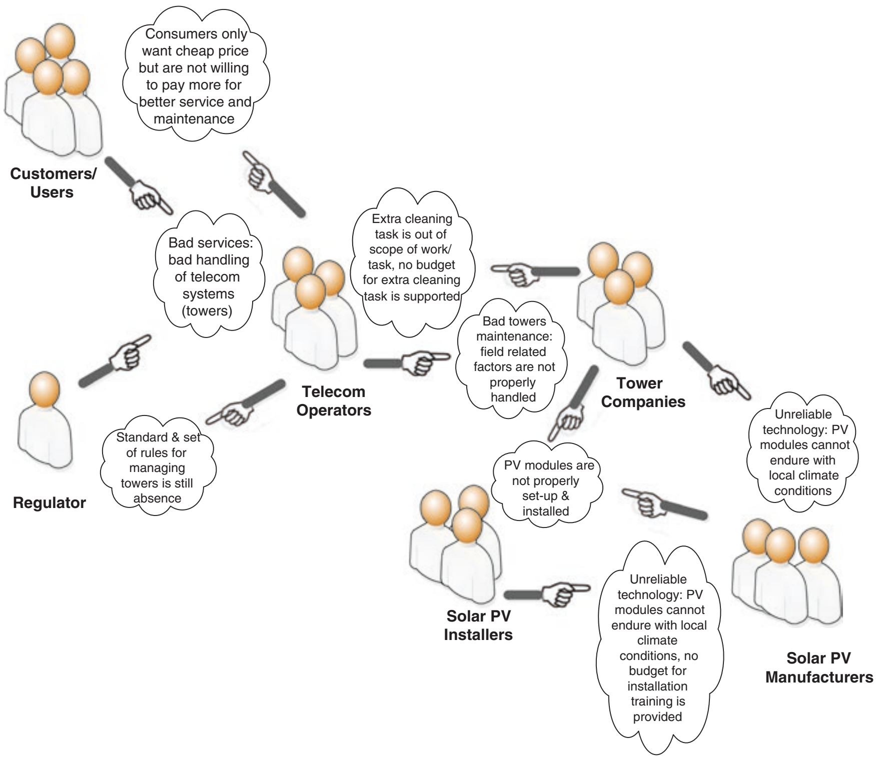

Fig. 12.4 Finger-pointing between actors involved in solar PV adoption

illustrates the finger-pointing between actors involved in solar PV adoption as the phenomenon that could emerge.

Above elucidation implies that a lack of and difficulty in locating the accountability of actors involved create the risk of failure of solar PV adoption and could bring in negative (unexpected) impact of innovation in terms of environment, social, and economy. The impact on environment is that such failure may lead to the rejection of solar PV adoption and the adoption of other renewable energy technologies. The adoption of solar PV can be seen as very complicated since it is not easy to deal with the field related factors, especially by the tower companies since they are expected to be responsible for the maintenance of the towers. Whereas it should be seen to some extent as a joint responsibility of all the actors in the system. Thus, there could be a set-back to replace the use of unsustainable and non-environment friendly technology such as DG for powering telecom towers. The impact on social is that the absence of sustainable power supply and good maintenance of telecom towers will lead to the unreliable telecom systems and infrastructure, and thus suffering the users of telecommunication services. Meanwhile the impact on economy is that the failure of the adoption of solar PV creates the risk of telecommunication systems collapse. Thus, if and when such disaster occurs, stakeholders will suffer severe economic losses.

In order to prevent the failure of the adoption as well as the negative (unexpected) impact of such adoption, the first thing to do is by clarifying the interests, tasks and responsibilities of actors involved as well as their interests. According to Indonesia's Telecommunication Law No. 36/1999, telecom operators are those who provide telecommunication network and services to users. They are responsible for reliable telecommunication systems and network infrastructure. In addition, they have to provide budget for maintenance cost and other costs related to supporting activities in managing telecom infrastructure notably the telecom towers. Ensuring the reliability of telecommunication systems and infrastructure is the main concern of telecom operators. This concern is most likely driven by their interest to ensure the business can keep running and creating profit out of it. In this case, the adoption of solar PV in telecom towers is a decision made by telecom operators as an effort in ensuring the availability of power in telecom towers, and thus a reliable telecommunication systems and infrastructure. Further, in order to perform such tasks and fulfill such responsibilities, telecom operators need to cooperate and involve other parties. Most importantly are the tower companies who provide telecom towers construction and maintenance services. Tower companies are responsible for site maintenance and monitoring system performance of telecom towers. Similar to telecom operators, their interest is to make profit from telecom sector as the telecom infrastructure provider.

Such adoption also involves solar PV manufacturers who provide solar PV modules. They are responsible for the production of reliable solar PV modules that can fit with the specification of telecom tower equipment. Next to their involvement in the adoption are solar PV installers who provide solar PV systems setup and installation services. They are responsible for ensuring solar PV modules operating in the best performance according to specification of telecom tower equipment and site condition. The setting and installation determine the performance of solar PV modules, any improper setting and installation can cause dysfunction of the tower's power system. Both, PV manufacturers and PV installers have quite similar interest to make profit from business in PV technology, as they concern in renewable energy development. In some cases, PV manufacturers could also serve as PV installers, but in this case we give our emphasis to the clarity of the tasks and responsibilities. Therefore in this case they are considered as two different actors with different tasks and responsibilities.

Furthermore, regulator provides rules and regulation regarding activities in the telecommunication sector. Regulator has interest in national economic development through telecommunication sector. In general, regulator is responsible for monitoring the functioning and performance of telecommunication sector and undertaking any necessary measures to ensure telecommunication activities can meet the standard and quality service operations (BRTI 2010). So in this case, regulator should be held responsible for providing regulation or standard for managing telecom towers.

Looking at aforementioned tasks and responsibilities of actors involved, such fulfillment of these tasks and responsibilities will lead to the successful adoption of solar PV as well as to managing telecom towers. Thus, it will create values and bring in positive (expected) impact of innovation in terms of environment, social, and economy. The impact on environment is that the adoption of solar PV provides more sustainable and environment friendly power supply for telecom towers. The impact on social can be that sustainable power supply and good tower maintenance services lead to reliable use of mobile communication, and thus keep the customers service level high. Meanwhile the impact on economy can be associated to the cost of energy provision, and to a great extent it may boost the development of renewable energy sector such as the PV industry.

Further, such positive impacts of the adoption of solar PV are expected to give benefits to each of the actors in accordance to their interests. The adoption of solar PV can reduce the cost of powering telecom towers which to a great extent benefiting telecom operators. Telecom operators are expecting to keep (and increase) the loyalty of their customers and attract new customers, and thus to make more profit out of it. Tower companies can reduce their operational cost in the long run by reducing the use of diesel generator. The success adoption of solar PV could also increase market for innovation PV technology which creates more opportunities for solar manufacturers to make profit out of it. Also, as people/society comes to trust in innovation, it could lead to the raising demand of PV technology as expected by PV manufacturers and PV installers. These benefits could induce the actors involved to engage in collaborative, responsible behavior in the fulfillment of their tasks and responsibilities. In other words such benefits could be seen as the incentives for the actors involved to cooperate each other and to participate in responsible innovation. Table 12.1 epitomizes stakeholder and impact analysis of the case study.

Table 12.1 Stakeholder and impact analysis

Actor	Motivation	Task (Responsibility)	Impact			Benefit
			Environment	Social	Economy	
Telecom operators	Business profit in telecom sector as the network and service provider	Providing telecommunication network and services (Reliable telecom systems & network infrastructure (i.e. adopting solar PV in telecom tower, providing budget for maintenance cost, etc.))	(+) The adoption of solar PV provides (more) sustainable and environment friendly power supply for telecom towers, e.g. less CO_2 emission, less noise compare to DG (−) The failure of the adoption of solar PV may lead to the rejection of using renewable energy technologies, and a setback to the use of unsustainable and non-environment friendly technology such as DG for powering telecom towers	(+) Sustainable power supply and good tower maintenance lead to reliable telecom systems & infrastructure keep the customers service level high (reliable use of mobile communication) (−) The absence of sustainable power supply and good tower maintenance will lead to unreliable telecom systems and infrastructure, low customers service level, and thus suffering the users	(+) The adoption of solar PV reduces the cost of energy for powering the towers. This may boost the development of renewable energy sector in general (−) The failure of the adoption of solar PV creates the risk of telecommunication systems collapse, at the end if and when such disaster occurs, the economic losses will be severe	Reliable telecom service and infrastructure keep (and increase) the loyalty of the customers and attracts new customers. Thus, it creates business profit
Tower companies	Business profit in telecom sector as the telecom infrastructure provider	Providing tower construction & maintenance services (Site maintenance & monitoring systems performance)				Cost reduction in business operation in the long run by reducing the use of diesel generator
Solar PV manufacturers	Business profit from renewable energy development	Providing solar PV modules (Production of reliable solar PV modules (i.e. resistance to climate condition))				Increase market for innovation PV technology creates more opportunity for making business profit
Solar PV installers	Business profit from renewable energy development	Providing solar PV system setup & installation services (Solar PV system operates in the best performance)				Rising demand for the PV technology as people /society comes to trust in the innovation
Regulator	National economic development through telecommunication sector	Providing rules and regulation for telecommunication sector activities (Regulation or standard for managing telecommunication systems (i.e. telecom towers))				High competitive and environmentally friendly telecommunication sector

Notes (+) = positive impact, (−) = negative impact

12.3.4 The Proposed Solution or Way-Out

Stakeholder and impact analysis in the previous section provides the map of tasks and responsibilities of actors involved. It also briefly portrays the possible impact of solar PV adoption in terms of environment, social, and economy. The map sheds some light on how to locate the accountability of actors involved and to clarify who should be accountable to whom. Following stakeholder and impact analysis, the five dimensions of responsible innovation are then further elaborated and analyzed. In sum, this is simply illustrated in Table 12.2.

The analysis of the five dimensions of responsible innovation suggests that in order to ensure the accountability of actors involved can be proportionally distributed, there is a need of innovative way or an adoption of technology that considers three aspects: environment, social, and economy. Therefore, by engaging these dimensions into consideration, at least two possible solutions or ways out can be proposed. The first possible solution is by identifying and developing framework which can help in assessing and fixing up the responsibility of actors involved in the adoption of solar PV as well as in managing telecom towers. However, to some

Table 12.2 Analysis of the five dimensions of responsible innovation

Dimension	Form of question	Indicative finding/answer
Anticipation	What (negative) impacts possibly emerge and should be prevented?	Impact on environment, social, and economy
Reflexivity	What causes the problem and the effects of it?	Causes of the problem: the field related factors Effects: declining in power output/production, dysfunction of towers, and risk of telecom system collapse
Responsiveness	What issue needs to be addressed?	Accountability of actors involved in dealing with the field related factors
Deliberation	What should be done to address the issue? What possible actions can be proposed?	Developing innovative way or (adoption of) technology that considers three aspects: environment, social, economy
Participation	Who should be involved and contribute to problem solving?	Telecom operators, tower companies, PV manufacturers, solar PV installers, and regulator

extent the application of such framework could be complicated and time consuming, thus makes it less practical. Therefore, as another alternative, the second possible solution can be proposed is by adopting technology that can cope with the field related factors namely the self-cleaning PV module technology. Such technology is expected to prevent power losses due to the local climate conditions as well as to reduce the cost of maintenance such as periodic cleaning task. The self-cleaning PV module technology could be the answer for the problem since it offers more practical and cost-efficient PV modules maintenance.

12.4 Conclusion

Through a case study, this paper has illustrated the application of the concept of responsible innovation in practice by undertaking stakeholder and impact analysis, and analysis of five dimensions of responsible innovation. From stakeholder and impact analysis, we get the map of tasks and responsibilities of actors involved and the possible impact of innovation adoption. This map helps us to locate the accountability of actors involved and to clarify who should be accountable to whom and for what. In addition to this map is the overview of the possible impact of innovation adoption. While innovation can provide the solution to such problem, but it can also bring another problem since the negative impact can emerge afterwards. Therefore, such overview can help us to figure out beforehand about what (negative) impacts could possibly emerge and should be prevented. Following stakeholder and impact analysis, by further elaborating the five dimensions of responsible innovation we also get some clues on how to proportionally distribute the accountability of actors involved. Therefore, by engaging these dimensions into consideration, at least two solutions or ways-out can be proposed: the identification and development of an accountability assessment framework, and self-cleaning PV module technology. This also implies that the five dimensions of responsible innovation can serve as the overarching anchor points for exploring the possible solutions or ways-out of the issue.

The case study also shows that the absence of responsibility in solar PV adoption can lead to the dysfunction of telecom towers. Also, non-consideration of post-deployment factors in designing of renewable technology applications can lead to less success at the deployment level. Whilst, due to lack of consideration of accountability, it can also lead to conclusions on part of tower companies that PV technology is not viable and not a sustainable option for powering telecom towers. This issue is becoming important to get more attention in the context of developing countries like Indonesia, since the failure in technology adoption often leads to the rejection of such technology in the later stage. This phenomenon might happen due to culture, and it is therefore different to developed countries where such failure often leads to more experimentation for improving such technology, and thus often leads to another innovation. Therefore, another lesson that can be learned from this case is that innovation without responsibility faces dangers of rejection or less acceptability of even greener and sustainable technology. To a greater extent, this case can also be considered as a starting point to discuss broader issues associated with the (non) adoption of new technologies in different contexts.

The paper also concludes that ensuring the accountability of innovation actors can be regarded as the way to prevent the unexpected impact of innovation, hence what we may term as responsible innovation. What also important from the adoption of such innovation is the adaptation with the local environment, social, and economic conditions. Furthermore, for such innovation to be responsible, it should be inclusive for environment and social life, and stakeholder participation should be co-evolved in the innovation process. However, the paper has not yet addressed the cross-cultural issue of the adoption into detail, while further elaboration of such particular issue would give more insights on the applicability of responsible innovation concept in developing countries context. Therefore, other avenues that can be suggested for future research are to examine the effect of national culture on responsibility in the adoption of solar PV as part of transitions to renewable energy; and on the role of actors and institutions in shaping responsible innovation in the context of energy transitions. Especially regarding the later avenue, it may raise the question such as who could be expected to or should take the lead in the anticipatory responsible innovation processes as outlined in Table 12.2.

Acknowledgment We would like to express our gratitude to the Indonesia Endowment Fund for Education (LPDP), Ministry of Finance of Republic of Indonesia for the scholarship support to the first author of this chapter.

References

Acosta, M.-S., and A. Hoolbrok. 2012. *Some notes on theories of technology, society and innovation systems for S&T policy studies*. Vancouver BC: Centre for Policy Research on Science and Technology Simon Fraser University.

Balkema, A., and A. Pols. this Volume. Biofuel: Sustainable innovation or gold rush? In *Responsible innovation, volume 2: Concepts, approaches, and applications*, ed. B.J. Koops, I. Oosterlaken, H.A. Romijn, J. van den Hoven and T.E. Swierstra. Dodretch: Springer.

Blok, V., and P. Lemmens. this Volume. The emerging concept of responsible innovation. Three reasons why it is questionable and calls for a radical transformation of the concept of innovation. In *Responsible innovation, volume 2: Concepts, approaches, and applications*, ed. B.J. Koops, I. Oosterlaken, H.A. Romijn, J. van den Hoven and T.E. Swierstra. Dodretch: Springer.

BRTI. 2010. Fungsi dan Wewenang BRTI. Available at http://brti.or.id/fungsi-dan-wewenang.

Chevalier, J.M. 2008. Stakeholder identification. In *Social analysis system (SAS): A guide to collaborative inquiry and social engagement*, ed. J.M. Chevalier and D.J. Buckles. New Delhi: SAGE Publications India Pvt Ltd.

Crossan, M.M., and M. Apaydin. 2010. A multi-dimensional framework of organizational innovation: A systematic review of the literature. *Journal of Management Studies* 47(6): 1154–1191.

Dosi, G. 1988. The nature of the innovative process. In *Technical change and economic theory*, ed. G. Dosi and C. Freeman et al. London: Pinter Publishers.

Edquist, C. 1997. *Systems of innovation: Technologies, institutions and organizations*. London: Pinter Publishers.

El-Shobokshy, M.S., and F.M. Hussein. 1993. Degradation of photovoltaic cell performance due to dust deposition on to its surface. *Renewable Energy* 3: 585–590.

Elminir, H.K., A.E. Ghitas, R.H. Hamid, F. El-Hussainy, M. Beheary, et al. 2006. Effect of dust on the transparent cover of solar collectors. *Energy Conversion and Management* 47: 3192–3203.

Freeman, R.E. 1994. Strategic management. A stakeholder approach. Marshfield, Massachussets: Pitman Publsihing Inc.

Grinbaum, A., and C. Groves. 2013. What is "responsible" about responsible innovation? Understanding the ethical issues. In *Responsible innovation*, ed. R. Owen, M. Heintz and J. Bessant, 119–142. London: Wiley.

Guston, D.H., and D. Sarewitz. 2002. Real-time technology assessment. *Technology in Society* 24(1–2): 93–109.

Hassan, A.H., U.A. Rahoma, H.K. Elminir, and A.M. Fathy. 2005. Effect of airborne dust concentration on the performance of PV modules. *Journal of the Astronomical Society of Egypt* 13(1): 24–38.

Hegazy, A.A. 2001. Effect of dust accumulation on solar transmittance through glass covers of plate-type collectors. *Renewable Energy* 22: 525–540.

Jong, M.de, A. Roelofsen, F. Kupper, T.d.C. Buning, and J. Broerse, this Volume. Exploring responsible innovation as a guiding concept: The case of neuroimaging in justice & security. In *responsible innovation, volume 2: Concepts, approaches, and applications*, ed. B.J. Koops, I. Oosterlaken, H.A. Romijn, J. van den Hoven and T.E. Swierstra. Dodretch: Springer.

Kompas.com. 2010. BTS Ramah Lingkungan Telkomsel Terbanyak di Asia. Available at <http://tekno.kompas.com/read/2010/02/18/1522405/BTS.Ramah.Lingkungan.Telkomsel.Terbanyak.di.Asia>.

Koops, B.J., this Volume. The concepts, approaches, and applications of responsible innovation. An introduction. In *Responsible innovation, volume 2: Concepts, approaches, and applications*, ed. B.J. Koops, I. Oosterlaken, H.A. Romijn, J. van den Hoven and T.E. Swierstra. Dodretch: Springer.

Latif, S.S. 2012. Menara Telekomunikasi Menjamur Tahun Ini. Available at <http://us.bisnis.news.viva.co.id/news/read/279808-menara-bts-menjamur-di-tahun-ini>.

Lundvall, B.-Å. 1992. National systems of innovation toward a theory of innovation and interactive learning. In *Introduction* ed. B.-Å. Lundvall. London: Pinter Publishers.

Maksudi, E., E.D. Tamtama, and E.S. Agustina. 2012. *Green BTS : Rancang bangun BTS remote solution untuk area terpencil*. Departemen Teknik Elektro: Universitas Indonesia.

Mani, M., and R. Pillai. 2010. Impact of dust on solar photovoltaic (PV) performance: Research status, challenges and recommendations. *Renewable and Sustainable Energy Reviews* 14: 3124–3131.

MCIT. 2011. *Annual Report 2011*. Jakarta: Ministry of communication and information technology.

Mitratel. 2010. Solar powered BTS is effective to produce an electricity. Available at http://www.mitratel.co.id/news/read/33.

Mytelka, L.K., and K. Smith. 2002. Policy learning and innovation theory: An interactive and co-evolving process. *Research Policy* 31(8–9): 1467–1479.

Nihlén Fahlquist, J. 2006. Responsibility ascriptions and vision zero. *Accident Analysis and Prevention* 38: 1113–1118.

Rogers, E.M. 2003. *Diffusion of innovations*, 5th ed. New York: Free Press.

Schot, J., and A. Rip. 1997. The past and future of constructive technology assessment. *Technological Forecasting and Social Change* 54(2–3): 251–268.

Schumpeter, J. 1961. *The theory of economic development*. New York: Oxford University Press.

Singh, R., and O. Kroesen. 2012. Understanding responsible innovation from developing countries perspectives. In *The 2nd conference on responsible innovation 2012*. The Hague, The Netherlands.

Sriram, S. 2009. 'Staying connected': Increasing role of solar power in telecom applications in Southeast Asia. Available at http://www.frost.com/sublib/display-market-insight-top.do?id=185477272.

Ubois, J. 2010. Responsible innovation/sustainable innovation. In *Annual meeting of the American Anthropological Association, November 2010*.

van de Poel, I., J. Nihlén Fahlquist, N. Doorn, S. Zwart, and L. Royakkers. 2012. The problem of many hands: Climate change as an example. *Science and Engineering Ethics* 18(1): 49–67.

Zorilla-Casanova, J., M. Piliougine, J. Carretero, P. Bernaola, P. Carpena et al. 2011. Analysis of dust losses in photovoltaic modules. In *World Renewable Energy Congress 2011*. Sweden.

Chapter 13
Responsible Innovation in Social Epistemic Systems: The P300 Memory Detection Test and the Legal Trial

John Danaher

Abstract Memory Detection Tests (MDTs) are a general class of psychophysiological tests that can be used to determine whether someone remembers a particular fact or datum. The P300 MDT is a type of MDT that relies on a presumed correlation between a detectable neural signal (the P300 "brainwave") in a test subject, and the recognition of those facts in the subject's mind. The P300 MDT belongs to a class of brain-based forensic technologies which have proved popular and controversial in recent years. With such tests increasingly being proffered for use in the courtroom—to either support or call into question testimony—it would behoove the legal system to have some systematic framework for ensuring that they are used responsibly, and for this framework, in turn, to play a part in future research and development of this technology. In this paper, I defend one such framework for ensuring that this is the case: the legitimacy enhancing test. According to this test, it is appropriate to make use of technologies such as the P300 MDT whenever doing so would (probably) enhance the legitimacy of the trial. I argue that this test addresses tensions between scientific and legal norms of evidence, and exhibits a number of additional virtues including unification, simplicity and flexibility.

Keywords Lie detection · Memory detection · P300 test · Scientific evidence · Forensic innovation · Legitimacy · Social systems

13.1 Introduction

The set-up is a familiar one. A nervous suspect, his body wired-up to record his autonomic responses, sits in the chair. An intimidating questioner hovers nearby, a list of questions in her hand. She starts off easy, asking the suspect to state his

J. Danaher (✉)
Department of Law, National University of Ireland, Galway, Ireland
e-mail: john.danaher@nuigalway.ie

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_13

name and date of birth. But soon things escalate in emotional intensity. She asks awkward questions about his childhood, such as “Did you ever lie to get out of trouble?”. Finally, she moves into the core of the interview: “Where were you on the night of the 25th?”, “Did you kill the victim?”.

This scenario, played out on screen in an endless series of police procedurals, describes the essence of the classic polygraph lie detector test (Iacono 2008; National Academies of the Sciences Report 2003; Furedy 1996; Furedy and Heselgrave 1988; Ben-Shakar 2002). The test, originally developed by John A. Larson in 1921 (Adler 2007) is an object of suspicion and cultural obsession. It speaks to our fears and desires. On the one hand, anything that allows us to look past the often biased and distorted nature of testimony, to the real intentions and motivations of the speaker, could be a major boon to the legal system. But on the other hand, anything that looks beneath our public expressions, and into the workings of our minds, seems to threaten our cognitive autonomy, invade our privacy, and erode our sense of trust. Little wonder then that, since the earliest days, the polygraph test has never truly been welcomed by courts of law.[1] (Note: references to courts of law throughout this paper are to those found in the Anglo-American legal systems most familiar to the author, specifically: Ireland, England and Wales, and the United States. No claims are made about continental legal systems.)

But that is the past, this paper is concerned with the future. Since 1921 things have changed. New, more sophisticated, versions of the lie-detecter test have arisen, with the latest wave incorporating brain-imaging techniques (Schauer 2010, 2012) and alternative “mind-reading” tests have been proposed, developed and investigated (Verscheure et al. 2011). One particular alternative is the P300 memory detection test (MDT[2]), which, instead of detecting deception, locates forensically valuable information in a person’s brain. Already used, with minimal success,[3] in legal trials, there are an increasing number of advocates for the forensic use of this technology (e.g. several papers with glowing accounts in Verscheure et al. 2011). The question is whether the advocates are right that this technology should be more welcome than its controversial predecessors?

This paper tries to answer that question by defending a novel framework for the responsible use of the P300 MDT in a social system like law. The framework, building on concepts from democratic theory and social epistemology, provides a simple, unifying test for the responsible use of the P300 MDT. This, it is argued, in turn feeds into an analysis of the responsible innovation of any social system that generates judgments

[1]On the early rejection of the lie detector test, see *Frye v United States* (1923) 293 F. 1013. Schauer (2012).

[2]I take the name from Meegan (2008). Roughly equivalent terms are “Guilty Knowledge Test” or “Concealed Information Test”, but both of those carry unwelcome pejorative implications (guilt in the first instance and concealment in the second). “Memory Detection” is a more neutral, yet still descriptively appropriate term.

[3]Lawrence Farwell famously started the company Brainwave Science in order to push the forensic use of his “brainfingerprinting” version of the P300 MDT. A full analysis of his attempts can be found in Rosenfeld (2005). A response to this can be found in Farwell (2011).

of truth and falsity (hereinafter called "social epistemic systems"—a phrase derived from Koppl 2006). Thus, the paper aims not only to discuss a particular technology, but also to build an evaluative bridge between norms of scientific and technological development, and norms for the successful design of social epistemic systems, such as the legal trial. The building of such evaluative bridges is crucial in responsible innovation, as Boshuijzen-van Burken (Chap. 14, this volume) illustrates in relation to military decision-making and the use of technology, and as Maslen et al. (Chap. 7, this volume) do in relation to the use of cognitive enhancing drugs by professionals.

The remainder of the paper is structured as follows. In Sect. 13.2, I briefly outline some of the salient features of the P300 MDT, explaining how it is distinct from the more widely-discussed fMRI lie detector, and highlighting the benefits and risks associated with the technology. In Sect. 13.3, I analyse some tensions between scientific and legal standards of evidential success, recently highlighted by Frederick Schauer, which can affect our willingness to make use of forensic techniques of this sort, and which are not satisfactorily addressed by existing tests for the admissibility of scientific evidence. In Sect. 13.4, I propose and defend a novel framework that tries to resolve these tensions, something I call the Legitimacy-Enhancing Test (LET). This gives a roadmap for the responsible use of emerging technologies, and a roadmap for the responsible innovation of the legal trial. Finally, in Sect. 13.5, I apply this test to the P300 MDT.

The subject matter fits within the broader landscape of responsible innovation as follows: Traditionally understood, responsible innovation is about the research and development phase of technology, not about the usage phase (the latter being where my concern lies). But this traditional dichotomy is misleading. Research and development is often an iterative process, with the attempted use of a technology playing an important role in further phases of research and development. I would argue that it is crucial that this iterative process be guided by an appropriate framework for responsible usage. In this respect, my paper fits within the product-oriented approach to responsible innovation, suggested by Koops (Chap. 1, this volume). With such a framework in place, we can learn from and prevent, failed, irresponsible attempts to use a technology, and use that information in future rounds of research and development. Indeed, as I hope to illustrate below, this has already happened, to some extent, in relation to the P300 MDT. Rosenfeld et al. (2013) provides, perhaps, the best example of this, setting out the current state of the technology and the needs of the legal system, in an effort to guide future research. The hope is that the LET can further advance this process by providing a tool for constant, responsible, iterative improvements.

13.2 The Allure of the P300 MDT

A memory detection test (MDT) is a test that purports to detect the presence of certain memories in a person's mind. The P300 MDT is simply one example of such a test (other variants are discussed by Meegan 2008, and Verscheure et al.

2011). It relies on electro-encephalographic (EEG) imaging, which is used to record variations in electrical activity across different regions of the brain (Rosenfeld 2011 and Meegan 2008). The brain displays constant and sometimes consistent variation in electrical activity (brainwave patterns), with some patterns being associated with certain types of conscious awareness, e.g. delta waves are associated with sleep.

If a discrete stimulus is presented to a subject who is hooked-up to an EEG, a "blip" in the otherwise constantly varying levels of activity can be detected. This is known as an evoked response potential (ERP). The P300 is particular kind of ERP that seems to arise whenever recognised, meaningful and rare stimuli are presented to a test subject (Polich 2007). As a result, the P300 is thought to be an ideal candidate for an MDT. If the testing paradigm is robust enough, and if it is protected from confounding variables, then the detection of a P300 in a test subject could be a reliable indicator of their recognition of certain information. If that information is taken, for example, from a crime scene, then the results of the P300 MDT might help to link a suspect to a crime scene, in a manner somewhat analogous to the use of DNA-matching. This would make the test forensically useful.[4]

It is no surprise then to learn there are those who advocate its forensic use. Perhaps the most conspicuous advocate is Lawrence Farwell, one of the original pioneers of the P300 MDT (Farwell and Donchin 1991), who developed his own patented version of the test (sometimes referred to as "brain fingerprinting")[5] and received some reasonably high-profile media exposure for his efforts to get it accepted by the courts. Those efforts were successful in one case, with a court (Iowa 2003) admitting the evidence under the so-called Daubert test (discussed below), though the technique proved less successful in a later case (Oklahoma 2005). In subsequent years, Farwell has involved himself in classified military and security service tests of his technique (Farwell et al. 2012), before recently re-emerging to the public eye to argue for the strengths of his version of the P300 MDT.

There are, of course, critics of Farwell's test (Rosenfeld 2005; Meegan 2008). One criticism concerns the forensic uses to which it has been put. In the *Harrington* decision, in which the results of a P300 MDT were admitted to court, the test was used to exculpate a prisoner who had served 24 years in jail. The results of the test supposedly showed that certain information relevant to a crime was not present in the defendant's brain, but that information relevant to his alibi was. Now, it is questionable whether this had a significant effect on the decision of the court since a key witness in the earlier case had also recanted his testimony, but in any event, as Rosenfeld (2005) points out, the forensic value of Farwell's

[4]It should be noted that, as with DNA-matching, the fact that an accused person can be linked to a crime scene does not equate with their guilt, obviously additional steps are needed for that.

[5]For general details, see Farwell's company webpage at www.governmentworks.com/bws. The unique feature of Farwell's test is that in addition to recording the P300 response it also focuses on something Farwell calls the MERMER response. The technique is promoted in Farwell et al. (2012) and Farwell (2012).

test result in this context was practically nil. Memories are constantly distorted, reconstructed and rewritten over the course of time. A person who was actually present at a crime scene 20 years ago may have forgotten certain details that the P300 MDT tries to test them on, thus resulting in a negative test that is not truly exculpatory. Similarly, a person who has rehearsed their alibi story for over 20 years could well recognise information connected with that story, despite it not being true (Allen and Mertens 2009). There are also a number of ethical concerns associated with Farwell's use of the technology. Most obviously, misapplication of an epistemically faulty or unreliable test will degrade the quality of legal decision-making, which is itself ethically undesirable. In addition to this, concerns have been expressed about the effects of such techniques on cognitive liberty and autonomy (Sententia 2001, 2004; Halliburton 2007). Nevertheless, there remains considerable enthusiasm about the forensic uses of MDTs[6] and even the critics of Farwell's test have proposed alternative, and they believe more reliable, versions of the P300 MDT.[7] This wave of enthusiasm shows no signs of abating.[8]

The back-and-forth between Farwell and his critics provides a neat illustration of the iterative process of technological research and development. Clearly, Farwell's attempt to introduce the P300 MDT in the early 2000s, in cases like *Harrington*, was a case of irresponsible innovation: he moved too quickly and too recklessly from preliminary, albeit promising, research results to forensic application. He has been rightly criticized for this. However, it would be wrong for his failures to simply close-off all future usage of the technology. His mistakes have provided the basis for more responsible research and development, which has been, and is now being, undertaken by other researchers. The question is whether this new wave of research can be translated into responsible usage. The framework I set out below tries to argue that it can.

At this juncture, and since it will become relevant later, it is worth briefly considering the differences between the P300 MDT and another type of test that has received considerable attention and (some) enthusiastic support in recent years, namely: fMRI lie detection (see Schauer 2012, for an overview of the fMRI debate). Like the P300 MDT, fMRI lie detection tests have received at least one day in court (United States v Semrau, No. 07-10074 M1/P, W.D. Tenn. May 31, 2010), though, as with the classic polygraph test, they failed to find a sympathetic audience. Still, as with the P300 MDT, there are those promote their

[6]Contributor Patrick in Verscheure et al. (2011) argues that MDTs are an "idea whose time has come" (p. 9), and another contributor to the same volume makes the case for its widespread use by law enforcement (Iacono, pp 12–27).

[7]Rosenfeld is developing a version of the P300 MDT that makes use of something he calls the complex trial protocol. This, along with certain other testing techniques, makes the test more accurate and less prone to countermeasures. This is discussed in more detail in Sect. 13.4. A list of his most recent publications can be found at: http://groups.psych.northwestern.edu/rosenfeld/publications.html.

[8]Farwell published six articles on his technique since 2011, and Rosenfeld published five in 2012 alone. See both authors' webpages for the relevant papers.

forensic use, most notably two companies Cephos Corp (www.cephoscorp.com) and NoLie MRI (www.noliemri.com). Others are deeply concerned (Greely and Illes 2007).

So what are the differences between the two tests? Well, there are many, but two are important for present purposes. First, like the classic lie detector test, fMRI techniques usually[9] rely on a testing format—the control question test—of dubious reliability. Since the goal of a lie detection test is to determine whether a suspect is lying or not, the emotional saliency of the test questions is an important feature of the test. But emotional saliency is difficult to control and is something that can be manipulated by the tester and altered by the context of the test. This has been thought to undermine the reliability of the test (Furedy 1996, Furedy and Heselgrave 1988; Ganis and Rosenfeld 2011). The P300 MDT seems to avoid these problems because the goal of the test is not to determine deception, but to determine whether certain information is recognised. It is thought to be easier to create proper "control" questions in this format,[10] which improves the reliability of the technique. Allied to this is the second main difference between the tests, which has to do with the nature of the error rates associated with them. While figures vary between experimental tests, Schauer suggests, from a review of the evidence, that the accuracy of fMRI lie detection ranges between 70–90 %. Noticeably, however, the number of false positive errors in fMRI tests can be quite high, ranging upwards of 20 % in some cases. This presents a significant contrast with the P300 MDT. Although error rates for that test vary considerably too—with several tests reporting accuracy of near 100 % (Farwell et al. 2012; and Iacoco 2011, focusing on MDTs more generally) but others offering much lower accuracy rates (as low as 27 %) in more realistic mock-crime tests or when the risk of false memories is high (Allen and Mertens 2009; Mertens and Allen 2008)—they are heavily skewed toward the false negative type of error. In other words, the test frequently says that information which should have been recognised by a test subject was not, or reaches an inconclusive outcome. At the same time, whenever it does say that information was recognised, it is a pretty good indicator that this was indeed the case. This means that the test could have high probative value, higher than fMRI lie detection, provided it is used correctly. This is a point to which I shall return below.

In sum, the P300 MDT is an emerging forensic technique with a great deal of potential, but also with significant attendant risks. It is importantly different from a similar emerging technology, fMRI lie detection, in terms of its test format and its accuracy rates. Given its potential and its risks, it forms an interesting case study in the responsible development and use of emerging technologies.

[9]Note: since fMRI is simply an imaging-technique, it could potentially be used as the basis for a MDT. This is discussed in Gamer (2011). Nevertheless, this has not been developed to a significant extent yet. The P300 EEG-based technique has been much more fully developed.

[10]See Verscheure et al. (2011) on the advantages of the MDT. Note that deception may indirectly feature as part of the test (Rosenfeld 2011).

13.3 The Law and Innovative Forensic Technologies

With a basic understanding of the P300 MDT in place, we can now consider exactly how the legal system should approach a novel forensic technology of this sort. In doing so a key theme emerges: there is a connection between the responsible use of innovative technologies of this sort, the responsible innovation of an epistemic system such as the legal trial, and the future responsible development of technologies. In the attempt to develop and analyse that theme, this section performs two key functions. First, it highlights the role of forensic technologies such as the P300 MDT in the legal system. Second, it identifies some tensions between legal and scientific standards for evidential success which may affect how and when such forensic technologies are used. This paves the way for the development of a novel framework in Sect. 13.4.

13.3.1 The Trial as an Epistemic System

Following Koppl (2006), an epistemic system is here defined as any social system that (at least sometimes) generates judgments of truth or falsity. The legal trial can be understood as a social system that both generates judgments of truth and falsity (Danaher 2011a). We can see this by considering the dynamics of the trial in more depth. In any given trial, there is usually a key fact that needs to be proved in order to establish the appropriate legal outcome. This is known as the "factum probandum". In order to prove the factum probandum, the court relies on evidence. This evidence is presented to the court in the form of testimony. Witnesses are put before the court and asked to tell the court what they saw or what they experienced or, exceptionally in the case of experts, to offer opinions about what might have happened. The problem is that, at least in common law jurisdictions, the system is adversarial. Both sides want to prove different things, they present different witnesses to help them do so, and these witnesses oftentimes contradict one another. Thus, it becomes difficult for the court to figure out where the truth really lies, which in turn makes it difficult for it to perform its epistemic function in a fair and reliable manner.

Interestingly, it is this very property of the trial that highlights the allure of forensic technologies like the P300 MDT or the fMRI lie detector. When witnesses are incentivised to either mislead the court in the fulfillment of its epistemic function— as they are given the strategic dynamics of the adversarial trial— we have the *problem of false or absent testimony*. To overcome this problem, courts can adopt two strategies: (a) incentivise the presentation of truthful testimony (which they do through the use of penalties for lying or obstructing justice and the like) or (b) use some forensic technology to bypass or correct for false or absent testimony. MDTs and lie detector tests are examples of strategy (b). An MDT can allow a court to link a suspect to a crime scene, irrespective of their actual testimony, by checking to see whether their brains respond to crime-relevant information. Similarly, a lie detector test could allow the court to infer general deceptiveness on the part of a witness,[11] which would enable them to make judgments about the reliability of their testimony.

It is important to bear in mind, however, that even though judgments of truth or falsity play an important role in the trial, and improving the accuracy of those judgments is a major policy concern, the epistemic dimension of the trial is complemented by a number of normative and ethical dimensions. The trial is not simply an instrument for reaching judgments about the truth or falsity of particular claims; it is a procedure that protects and respects the central moral properties of the people who operate within it. Thus, there are procedural norms associated with the standards of proof, the right to a fair trial, and the protection of vulnerable witnesses. These norms play a significant part in determining whether it is appropriate to make use of novel forensic technologies like the P300 MDT, as we shall see.

13.3.2 Problems with Existing Admissibility Tests

The next question is whether existing protocols and tests for deciding on the use novel forensic technologies are fit for purpose? In other words, do they allow for the effective and responsible use of such technologies? The claim I wish to defend is that they are not, and this in turn hinders the responsible innovation of the technologies. In making this argument, I wish to clarify and build upon some arguments made by Frederick Schauer in relation to the use of fMRI lie detection (Schauer 2010, 2012). Schauer's concern is to highlight important tensions between scientific and legal norms of evidential value. These tensions represent faultlines between the scientific and legal systems that will need to be addressed by any proposed framework for responsible usage.

Schauer notes that, from its earliest days, the lie detector test it has struggled to win acceptance in the courts. This trend has continued despite the advent of newer versions of the test using fMRI imaging techniques. This is also true of the P300 MDT which, despite some initial success, is generally treated with a good deal of suspicion by both researchers, lawyers and courts (Rosenfeld 2011 and in Meegan 2008). Schauer questions the tenability of this trend by defending one overarching claim, which we may call "Schauer's Thesis":

> **Schauer's Thesis**: Whether fMRI lie detection (or P300 MDT) evidence should be admitted to court is not simply a question of its scientific validity and reliability, it is also (perhaps primarily) a question of the normative and ethical function of the law. That is to say, questions of evidential utility are fundamentally determined by legal-ethical standards, not purely scientific ones.

[11]Though, problematically this is all they seem capable of doing. See Danaher (2011b).

This claim is significant in that current tests for the admissibility of scientific evidence, such as DNA fingerprinting and other forensic techniques, rely heavily on, though are not isomorphic with, scientific standards of validity and reliability. For instance the Daubert test, states that judges should assess scientific evidence by referring to the indicia of reliability that are common in the scientific world. These indicia include things like "known error rates", "general acceptance within the relevant scientific community", "testability" and "passing peer review". This might seem like an obvious way in which to ensure that technologies are responsibly deployed in the legal system. But the problem is that this approach can yield significant legal territory to the norms of scientific inquiry. What scientists rightfully deem "good evidence" and what legal theorists rightfully deem "good evidence" may be two different things. It is important not to lose sight of this.

Schauer supports his thesis with two arguments, which we shall call the "probative context" and the "epistemic progress" arguments. Let's look first at the probative context argument. The key premise of this argument is that the value of any evidence placed before a court depends largely on three factors:

Probability: Does the evidence raise or lower the probability of the factum probandum and if so, by how much does it raise or lower its probability?

Standard of Proof: What confidence threshold must the probability of the factum probandum cross in order for it to count as being proved or not proved?

Legal Purpose: Is the evidence being submitted in order to prove or disprove the factum probandum?

These factors make up the probative context. The probative context varies depending on the legal issue at stake, and the party on whose behalf the evidence is proffered. For example, in criminal cases, the standard of proof for the prosecution is beyond a reasonable doubt. This is a notoriously fuzzy standard, but suppose that it corresponds to a 95 % (0.95) probability of the factum probandum being true. In that case, in order to secure a conviction, the prosecution would need to introduce a body of evidence that (in its totality) raises the probability of the factum probandum to the 95 % threshold. Contrariwise, the defence would succeed if they could introduce evidence that prevented the probability from crossing that 95 % threshold. Thus, the probative value of the evidence varies depending on who is presenting it, which is another of the variables of the probative context.

This is important because it feeds into the assessment of the value of lie detection or MDT evidence. Schauer notes that fMRI lie detector tests have reported accuracy rates that vary from 70 to 90 % (Schauer 2012). This might seem scientifically unimpressive, but they are better than chance at identifying deceptive individuals, and in the right context this could be probatively valuable. For example, in a criminal trial, although a 70 % accuracy rate might not suffice to prove that someone is guilty, it might suffice to prove reasonable doubt. So, for instance, if I am 70 % likely to be telling the truth about my alibi when I am being charged for murder, then (Bayesian considerations to one side) it might be highly useful for the court to be made aware of this fact. The important point is that decisions about the utility of evidence need to be highly sensitive to the probative context in which that evidence is presented, which is something they cannot be if they are overly-beholden to scientific standards of epistemic success since the probative context of science is quite different.

Another problem with existing approaches to determining whether novel forensic technologies should be used is the problem of status quo bias—a general problem in applied ethical reasoning (Bostrom and Ord 2006). Schauer illustrates this with the epistemic progress argument. An epistemic system can be said to under progress whenever there is some overall improvement in its epistemic efficiency. "Efficiency" in this sense is defined by Koppl (2006) as a measure of the likelihood of the system reaching a true judgment. The basic idea is that epistemic progress is a good thing, and that any reform to the system that allows it to progress would be welcome provided it does not compromise the ethical/normative aspects of the system (a point to which I shall return). Epistemic progress is always assessed relative to the existing level of epistemic efficiency. Thus, if we wished to argue in favour of a particular reform, we would have to do so by directly referencing the current level of efficiency. This relativistic aspect of epistemic progress has one interesting effect: if the current level of epistemic efficiency is low, then a proposed reform to that system, even one with an unimpressive level of overall accuracy, may nevertheless be warranted on the grounds that it still raises the efficiency of the system.

Schauer argues that this could be true in the case of fMRI lie detection. He does so by highlighting how existing solutions to the false or absent testimony problem are lacking. For example, historically, the administration of the religious oath was thought to incentivise truth-telling. In a culture in thrall to the fear of God and hell, this may have held some sway, but in its modern secular form the oath relies on the desire to be honest and the threat of punishment to do its work. Arguably, neither of these are particularly effective and certainly have no known accuracy rates. So with just the oath to protect us from false testimony, the epistemic efficiency of the trial is unclear. In light of these comparators, the admission of fMRI lie detection would seem to represent epistemic progress. Since it does have known accuracy rates, and since it can do something to break the deadlock between contradictory testimonies, it could lead to epistemic progress. Again, something similar could be true in the case of the P300 MDT, although that technique might have a greater impact on the problem of absent testimony.

Schauer's two arguments help to readjust our perspective on the relationship between the legal trial and the usage novel forensic technologies such as the P300 MDT. They encourage us to see that the needs of the trial and the needs of scientists may not be one and the same thing. This is something that should filter back into the research and development of these technologies. Still, Schauer's arguments are not entirely satisfactory. For one thing, it is possible to dispute some of his factual claims about fMRI lie detection.[12] For another, it is not clear whether they mount a serious challenge to existing protocols or tests for determining the

[12] Schauer (2012) deals with some such criticisms.

utility of novel forensic techniques. Is a new framework really needed to address Schauer's concerns?

If we assume that the current framework is represented by some version of the Daubert test, then it is possible to argue that there isn't really a problem. A Daubert-style test, as mentioned earlier, requires judges to assess the merits of scientific evidence by reference to a number of commonly-recognised indicia of reliability: error rates, peer review, sound methodology and so on. And while these are scientific indicia, there is no reason why an attentive mind could not apply those standards in a manner that is consistent with the epistemic and normative aims of the legal system. Thus, for example, a judge could look to the known error rate of a forensic technique that falls below what might be desirable for scientific purposes, but can still accept it if the probative context of the law would benefit from it. There is nothing in a Daubert-style test that prevents him or her from doing so. The test does not subordinate the law to science.

On a purely formal level, then, a Daubert-style test is neutral with respect to the types of concerns raised by Schauer: it doesn't actively seek to address them, nor does it actively exacerbate them. It is a small part of the overall apparatus the law uses to determine which kinds of evidence are worth considering and which are not. But this formal neutrality is itself problematic. By not actively directing our attention to Schauer's concerns, it is too easy for the person adopting the test to lose sight of epistemic progress and probative context. Arguably this is what Schauer shows to have happened in relation to polygraph and fMRI lie detector tests. This is compounded by the fact that, in terms of historical impetus, Daubert was created to address the problem of "junk" science, i.e. to close the doors of the court to bad forensic evidence (such as, perhaps, Farwell's use of the P300 MDT). As Schauer himself points out, this historical background tends to support the status quo within the legal system (Schauer 2010: 1216–1217).

One might suppose that simple emendations to the Daubert-test could solve the problem. By adding additional criteria to the test, we could overcome its neutrality and encourage active engagement with epistemic progress and probative context. Perhaps we could, but this would be a partial solution at best. This modified Daubert would still only look to the epistemic dimension of the trial and how some of the particular norms of the legal system play an important role in determining the probative value of evidence. This leaves other normative considerations in the lurch, either to be ignored or to be dealt with by other evidentiary rules. This should not be the case. A theoretical framework that provided a deeper theoretical grounding for Schauer's arguments could be used to assess forensic techniques, and also to assess other normative and procedural aspects of the legal trial. It would not only give us a roadmap for the responsible use of novel technologies like the P300 MDT, but a roadmap for the responsible innovation of an epistemic system like the legal trial. This roadmap could then guide the future development and deployment of this technology. The job of the next section is to provide that theoretical framework.

13.4 The Legitimacy Enhancing Test

The proposed framework is grounded in the notion of legitimacy. Put most simply, it requires a court (or other agency) when tasked with considering reforms to the trial—where "reforms" is taken to include decisions as to evidential admissibility—to ask whether that reform would enhance the legitimacy of the trial. The result is the legitimacy enhancing test (LET). This deceptively simply test allows the court to draw upon a rich theoretical literature about the nature of legitimacy conditions, which in turn bring together the disparate normative considerations highlighted in the preceding section. The remainder of this section will defend this test. It does so in two parts. First, it sets out the basic elements of the test, simplifying it to an argument with two premises: a normative premise and a factual premise. It then discusses each premise in sequence, noting how they supply the theoretical depth and unity absent from Schauer's analysis. The arguments here are programmatic, intending to provide stimulus for future debate and research, not a comprehensive defence of the test. Areas in which the theoretical basis of the test needs further elaboration will be highlighted, though in doing so I try to indicate how I think that elaboration could work.

13.4.1 The Test Itself

A central concept—perhaps the central concept—in contemporary political theory is that of legitimacy (Peter 2007, 2008a, b). The concept addresses the basic challenge of political authority, namely that the exercise of authority tends to be coercive and autonomy-undermining. That is to say: whenever the government exerts its authority, the result is typically that somebody, somewhere, is forced to do something they would otherwise have preferred not to do. Prima facie, this is unjustifiable. The notion of legitimacy is what rescues political authority from this prima facie unjustifiability because a coercive rule or practice becomes justifiable if it satisfies a set of legitimacy conditions.

The LET takes advantage of this concept in formulating a test for determining whether novel forensic technologies should be used in the legal trial. What's more it does so in a way that is sensitive to the probative context, the relativistic nature of epistemic progress, and the other normative dimensions of the trial. To see this, we can start by sketching the structure of the test, which can be reduced to a simple syllogistic argument.

1. If a change to an epistemic system with coercive powers (such as the legal trial) would enhance the legitimacy of that system, then that change is to be welcomed (normative premise)
2. Change X would/would not enhance the legitimacy of the trial (factual premise).
3. Therefore, change X is/is not to be welcomed.

The word “change” is used in the broadest possible sense. An epistemic system such as the trial could be changed by a number of things, with the use of novel forensic technologies being but one example of a change. It is this broadness that allows the LET to not only provide guidance about the responsible use of a novel forensic technology, but also to provide guidance for the responsible innovation of the epistemic system as a whole.

In practical terms, the LET works like this: whenever a court (or other agent within an epistemic system) must consider a change to that system, they simply ask themselves whether the above argument is true, or probably true, in the case of the proposed change. By focusing on legitimacy enhancement the LET will direct their attention to the need for epistemic progress relative to the status quo, and by focusing on legitimacy it will direct their attention to the epistemic and normative concerns mentioned in the preceding section. In doing so, the normative premise supplies the theoretical background missing from Schauer’s analysis and the modified Daubert test: it grounds, unifies and expands our assessment in a way that those proposals do not. Let’s unpack this in a little more detail.

13.4.2 The Premises in Practice

Although the truth of the normative premise need not be considered in practice—it can be taken as a background normative presupposition—some defence of it is in order here. In brief: legitimacy conditions must be met in order for a coercive act of political authority to be justified. The rationale for this is grounded in a liberal model of the state, according to which the default moral position in society is one of non-intervention in individual autonomy. Deviations from the default position need to be justified to the citizenry (Gaus 2003, 2010). Legitimacy conditions are what determine whether this can be done. Hence, the method of justification is through the satisfaction of legitimacy conditions. Since the trial plays an important role in the exercise of coercive political authority (see Danaher 2013, for a lenghtier discussion), it follows that it needs to be justified in light of legitimacy conditions.[13] This is why the normative premise is appropriate in this context. Interestingly, this highlights a limitation of the LET: it only applies to epistemic systems with some coercive power. If the system has no coercive power, then the legitimacy test is inappropriate. However, that limitation is not particularly severe since a good number of social epistemic systems have coercive power and it is oftentimes the fact that they deliver judgments of truth or falsity that grants them this power.

The strengths of the normative premise lie in how the concept of legitimacy gets fleshed out. There is a rich philosophical debate about which conditions

[13]This is not to say that the normative function of the trial reduces to that of legitimacy, merely that legitimacy is an important part of the normative justification of the trial, whatever its normative function may be.

supply the legitimacy needed for political justification. This debate has generated two distinct schools of thought: proceduralism and instrumentalism. According to proceduralism (Machin 2009; Estlund 1993, 2003; Peter 2008a, b), what legitimates a coercive decision are the features of the procedure through which that decision was made; according to instrumentalism (List and Goodin 2001; Marti 2005), what legitimates a coercive decision are the outputs of the procedure, specifically whether those outputs reach or get close to the right conclusion, normatively and/or epistemically speaking. Applying this to the trial, proceduralism would be concerned with the propriety of the trial procedure itself, i.e. whether it provides a forum in which the accused can be heard, whether it is biased and unfairly prejudicial, whether it respects individual autonomy, privacy and so on. On the other hand, instrumentalism would be concerned with whether the trial reaches the right verdict. This would be the case, in part at least, if the correct verdict about the factum probandum is reached.

The goal here is not to privilege one particular theory of legitimacy over another. Quite the opposite in fact. I suspect all the major theories highlight properties that normatively ideal epistemic systems exemplify and the goal is to bring all of those properties into consideration under the common umbrella of "legitimacy conditions". Thus, I propose that in using the LET we adopt a "mixed" account of legitimacy. In doing so, we provide a unifying framework for assessing the concerns in Sect. 13.3. Specifically, by paying attention to both the procedural and instrumental dimensions of legitimacy, and to the "enhancing" potential of any proposed reform, we can do justice to Schauer's probative context and epistemic progress arguments, as well as the other normative requirements of the trial such as respect for autonomy, privacy and moral equality. This can be done by asking the following three, deceptively simple, questions:

(A) Does the proposed change enhance the non-strategic epistemic efficiency of the trial?
(B) Does the proposed change enhance the strategic epistemic efficiency of the trial?
(C) Does the proposed change enhance the procedural virtues of the trial?

The first two questions deal with the instrumental side of legitimacy, focusing as they do on whether the change helps the system reach the right verdict about the factum probandum. The third question deals with the procedural side of legitimacy, focusing on whether the proposed change respects the rights of the people operating within the system. If one can answer "yes" to each question, then the change is acceptable. If the answers vary, then some careful balancing of the procedural and instrumental advantages will need to be considered.

On this point, it is important to bear in mind that the LET proposes a multidimensional test for epistemic reform. Broadly speaking, the instrumentalist and proceduralist concerns represent the two major dimensions to the test, but these (in particular the proceduralist dimension) can probably be broken down into a number of separate concerns. This raises the spectre of relativism and pluralism in the application of the test. It is possible that the many different kinds of legitimacy condition covered by the test are not reducible to one another, and therefore that reforms which satisfy different sets of those conditions cannot really be assessed relative to one another. For instance, maybe a reform which scores highly in terms of epistemic efficiency cannot be compared to a reform that scores highly in terms of its respect for privacy. This is indeed a serious concern and I can only gesture at a response to it here. The response would be that the plurality and irreducibility of legitimacy conditions does not entail their incommensurability, nor their incomparability (on this point, see Chang 2013 and forthcoming; Gaus 2003). As others have pointed out (Chang 2013; Gaus 2003), it is possible to rank plural values relative to one another. If that's right then it is also possible to rank different proposed reforms, relative to the status quo and to one another, even if those proposed reforms exemplify different sets of legitimacy condition. Admittedly, however, this is something that would require further specification if the LET is to become fully persuasive.

Finally, one potentially confusing aspect of this test is the reference to "strategic" epistemic efficiency in the second question. What differentiates this from non-strategic epistemic efficiency? The idea is that since epistemic systems involve interactions between strategic agents, we can never consider the epistemic virtues of a proposed change without also considering the effect of that change on the incentives of those agents. I consider an illustration of this problem below.

To summarise, the LET provides a simple, unifying framework for determining the responsible use of innovative forensic technologies, and the responsible innovation of the legal trial as a whole. It incorporates the concerns developed by Schauer in his analysis of fMRI-based lie detection, but supplies a deeper theoretical underpinning for those concerns (viz. the concept of legitimacy) which can in turn incorporate a broader, pluralistic set of concerns (procedural and instrumental). The final point to make is that the LET aids the iterative process of responsible research and development. It gives researchers a roadmap of what is needed when researching and developing the P300 MDT. It gives them a better sense of the values they need to respect when developing novel protocols for the test. In this manner, constructive "dialogue" between science and law can be facilitated.

13.5 Conclusion

I conclude by simply applying the LET to the case of the P300 MDT. I do so by asking and answering the three questions posed above. The application is not exhaustive, and should not be considered to definitively determine whether or not the P300 MDT should be used. Rather, it provides an illustration of how the framework defended in this article can be used in practice (for a lengthier illustration, see Danaher 2015).

First, would the use of the P300 MDT enhance the non-strategic epistemic efficiency of the trial? The answer to this is "it depends" on the probative context since that determines the epistemic goals of the trial and the measure of efficiency.

The clear strength of the technique is that it addresses the false or absent testimony problem, it does so in a manner that is better than chance (as suggested by Iowa 2003; Oklahoma 2005; Farwell et al. 2012; Rosenfeld 2005; Meegan 2008; and Verscheure et al. 2011), and it could be particularly effective when other methods of addressing the testimony problem (e.g. fingerprint or DNA evidence) are themselves absent. Furthermore, it addresses these problems in a manner that enhances epistemic efficiency relative to the status quo. This positive assessment must be tempered by some concerns about the ability of the test to distinguish true from false memories, and to detect recognition long after the crime has occurred (tests have proven reliability up to one month afterwards, e.g. Hu and Rosenfeld 2012). Protocols that determine when it is not safe to use the test could, however, correct for these problems. Allen (2008) for example suggests a possible protocol, relating to false memory.

Second, would the use of the P300 MDT enhance the strategic epistemic efficiency of the trial? If use of the P300 MDT becomes common place in law enforcement, it might incentivise changes in behaviour that thwart the utility of the test. A classic illustration of this might be the fact that glove-wearing was incentivised following the advent of fingerprint matching. Classic lie detector tests exhibit this kind of strategic weakness because they are vulnerable to countermeasures. That is: test subjects can undermine the test by performing voluntary acts that reduce its accuracy. This problem has been studied by P300 researchers Meixner and Rosenfeld (2010) and it has been found that complex testing protocols or the subliminal presentation of signals to the test subject make it far less vulnerable to countermeasures. This suggests that strategic epistemic weakness may not be a problem for this technique.

Finally, would the use of the P300 MDT enhance the procedural virtues of the trial? A more appropriate question might be: would it reduce or undermine those virtues? The answer is far from clear. One major concern might be that the test is used in practice as an interrogation prop. That is: the threat of the test is something a police investigator uses to force a confession, with the confession then being the main piece of evidence used at trial, not the result of the test. This might be thought to undermine the accused's right to a fair trial. It is suggested that strict rules for the use of the test could help to overcome this problem, though oversight of its actual use will be important. Similarly, although many are concerned about the risk of unfair prejudice (Goldberg 2008) when it comes to brain-based forensic techniques, this is overstated (Klaming 2011): such risks are present with other kinds of evidence that are routinely used (Schauer 2012), and the risk could be minimised by parallel reforms to the jury system (Danaher 2013). The test may also raise privacy concerns,[14] depending on the nature of the evidence collected. But one suspects those are relatively minor since the test, at best, may be used to link a suspect to a particular piece of information. This would be analogous to the dangers to privacy from matching DNA or fingerprints from the crime scene and

[14] Vedder and Klaming (2010) discuss privacy problems in relation to another possible reform of the legal epistemic system: eyewitness enhancement; I respond to their arguments at length in Danaher (2013).

so the concerns would be no more serious than those raised by those techniques. Still, some caution must be expressed. It is possible that the test inadvertently uncovers other details about the suspect that have yet to be fully appreciated. For example, perhaps information from the EEG-scans could be used to make inferences to cognitive or intellectual ability. Although I can find no reports of this, it is nevertheless something to be wary about in the future.

In sum, there might be case to be made in favour of the forensic use of the P300 MDT, but this case is best made within the framework supplied by the LET. This is because the LET creates a systematic and unifying framework for determining the responsible use of novel forensic technologies, and provides guidelines for the responsible innovation of social epistemic systems like the legal trial. Furthermore, the LET does this while remaining sensitive to the epistemic and procedural needs of the trial.

Acknowledgements I would like to thank two anonymous reviewers for comments on a previous draft.

References

Adler, K. 2007. *The lie detectors: the history of an American obsession*. New York: Free Press.

Allen, John J.B. 2008. Not devoid of forensic potential, but…. *American Journal of Bioethics* 8(1): 27–28.

Allen, John J.B., and R. Mertens. 2009. Limitations to the detection of deception: true and false recollections are poorly distinguished using an event-related potential procedure. *Social Neuroscience* 4(6): 473–490.

Ben-Shakar, G. 2002. A critical review of the control questions test. In *Handbook of polygraph testing*, ed. M. Kleiner. London: Academic Press.

Bostrom, Nick, and Toby Ord. 2006. The reversal test: eliminating status quo bias in applied ethics. *Ethics* 116: 656–679.

Chang, R. 2013. Incommensurability and incomparability. In *International encyclopedia of ethics*, ed. H. La Follette. Oxford: Blackwell.

Chang, R. forthcoming. Value pluralism. In *International encyclopedia of the social and behavioral sciences*, edited by Wright, J.D, 2nd Edn. Elsevier. Preprint available at: http://ruthchang.net/wp-content/uploads/2012/11/VALUEPLURALISMsubmittedupdateforweb1.pdf (accessed May 29, 2013).

Danaher, John. 2011a. Blind expertise and the problem of scientific evidence. *International Journal of Evidence and Proof* 15(3): 207–231.

Danaher, John. 2011b. The future of brain-based lie detection and the admissibility of scientific evidence. *Irish Criminal Law Journal* 21: 67–76.

Danaher, John. 2013. On the need for epistemic enhancement: democratic legitimacy and the enhancement project. *Law, Innovation and Technology* 5(1): 85–112.

Danaher, John. 2015. The comparative advantages of brain-based lie detection: the P300 CIT and pre-trial bargaining. *International Journal of Evidence and Proof.* 19(1): 52–66.

Estlund, David. 1993. Making truth safe for democracy. In *The idea of democracy,* ed. D. Copp, J. Hampton and J.E. Roemer. Cambridge: Cambridge University Press.

Estlund, David. 2003. Why not epistocracy?. In *Desire, identity and existence: essays in honour of T.M. Penner*, ed. Reshotko, N. Kelowna, B.C.: Academic Printing and Publishing.

Farwell, Lawrence. 2011. Brain fingerprinting: corrections to rosenfeld. *Scientific Review of Mental Health Practice* 8: 56–68.

Farwell, Lawrence. 2012. Brain fingerprinting: a comprehensive tutorial review of detection of concealed information with event-related brain potentials. *Cognitive Neurodynamics* 6: 115–154.

Farwell, Lawrence, and E. Donchin. 1991. The truth will out: interrogative polygraphy ("lie detection") with event related potentials. *Psychophysiology* 28: 531–547.

Farwell, Lawrence, D. Richardson, and G. Richardson. 2012. Brain fingerprinting field studies comparing P300-MERMER and P300 brainwave responses in the detection of concealed information. *Cognitive Neurodynamics*. doi:10.1007/s11571-012-9230-0.

Furedy, John. 1996. The north American polygraph and psychophysiology: disinterested, uninterested and interested perspectives. *International Journal of Psychophysiology* 21: 97.

Furedy, John, and R. Heselgrave. 1988. The validity of the lie detector: a psychophysiological perspective. *Criminal Justice and Behavior* 15: 219.

Gamer, Matthias. 2011. Detecting of deception and concealed information using neuroimaging techniques. In *Memory detection: theory and application of the concealed information test*, ed. Verscheure, B., Gershon Ben-Shakar and Ewout Meijer. Cambridge: Cambridge University Press.

Ganis, G.J.P., and J.P. Rosenfeld. 2011. Neural correlates of deception. In *The Oxford handbook of neuroethics*, ed. J. Illes, and B. Sahakian. Oxford: OUP.

Gaus, Gerald. 2003. *Contemporary theories of liberalism*. London: Sage Publications.

Gaus, Gerald. 2010. *The order of public reason*. Cambridge: Cambridge University Press.

Goldberg, D. 2008. The detection of constructed memories and the risks of undue prejudice. *American Journal of Bioethics* 8: 23–24.

Greely, Hank T. and Illes, J. 2007. Neuroscience-based lie detection: the urgent need for regulation. *American Journal of Law and Medicine* 377–431.

Halliburton, C. 2007. Letting Katz out of the bag: cognitive freedom and fourth amendment fidelity. *Hastings Law Journal* 59: 309.

Hu, X., and J.P. Rosenfeld. 2012. Combining the P300-complex trial-based concealed Information Test and the reaction time-based autobiographical implicit association test in concealed memory detection. *Psychophysiology* 49: 1090.

Iacono, W.G. 2008. Effective policing: understanding how polygraph tests work and are used. *Criminal Justice and Behaviour* 35: 1295.

Iacoco, W.G. 2011. Encouraging the use of the Guilty Knowledge Test (GKT): what the GKT has to offer law enforcement. In *Memory detection: theory and application of the concealed information test*, ed. Verschuere et al. Cambridge: Cambridge University Press.

Iowa (2003) Harrington v. State of Iowa 659 N.W.2d 509.

Klaming, Laura. 2011. Does neuroscientific evidence bias legal decision-making? some preliminary findings. *Law, Innovation and Technology* 3: 303–317.

Koppl. Roger. 2006. Epistemic systems. *Episteme: A journal of social epistemology* 2: 91–106.

List, C., and R. Goodin. 2001. Epistemic democracy: generalizing the condorcet jury theorem. *Journal of Political Philosophy* 9: 277.

Machin, D. 2009. The irrelevance of democracy to the public justification of political authority. *Res Publica* 15: 103.

Marti, J.L. 2005. The epistemic conception of deliberative democracy defended. In *Deliberative democracy and its discontents*, ed. Besson and Marti. London: Ashgate Publishing.

Meegan, D. 2008. Neuroimaging techniques for memory detection. *American Journal of Bioethics* 8: 9–20.

Meixner, J., and J.P. Rosenfeld. 2010. Countermeasure mechanisms in a P300-based concealed information test. *Psychophysiology* 47: 57–65.

Mertens, R., and J.J.B. Allen. 2008. The role of psychophysiology in forensic assessments: deception detection, ERPs and virtual reality mock crime scenarios. *Psychophysiology* 45: 286–298.

National Academies of the Sciences Report. 2003. *The polygraph and lie detection*. Washington, DC: National Academies Press.

Oklahoma. 2005. Slaughter v. State of Oklahoma 105 P.3d 832.

Peter, Fabienne. 2007. Democratic legitimacy and proceduralist social epistemology. *Politics, Philosophy and Economics* 6(3): 329.

Peter, Fabienne. 2008a. Pure epistemic proceduralism. *Episteme* 5: 33–55.

Peter, Fabienne. 2008b. *Democratic legitimacy*. London: Routledge.

Polich, J. 2007. Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology* 118(10): 2128–2148.

Rosenfeld, J. 2005. Brain fingerprinting: a critical analysis. *Scientific Review of Mental Health Practice* 4: 20–37.

Rosenfeld, J. Peter. 2011. P300 in detecting concealed information. In *Memory detection: theory and application of the concealed information test*, ed. Verscheure, B., Gershon Ben-Shakar and Ewout Meijer. Cambridge: Cambridge University Press.

Rosenfeld, J.Peter, H. Xiaoqing, E. Labkovsky, J. Meixner, and M. Winograd. 2013. Review of recent studies and issues regarding the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology* 90(2): 118–134.

Schauer, Frederick. 2010. Can bad science be good evidence: neuroscience, lie detection and beyond. *Cornell Law Review* 95: 1191.

Schauer, Frederick. 2012. Lie detection, neuroscience, and the law of evidence. *Virginia public law and legal theory research paper 2012–49.*

Sententia, Wyre. 2004. Cognitive liberty and converging technologies for improving human cognition. *Annals of the New York Academy of Sciences* 1013: 221–228.

Sententia, Wyre. 2001. Brain fingerprinting: databodies to databrains. *Institute for ethics and emerging technologies*, available at http://ieet.org/index.php/IEET/print/1412.

Vedder, A., and L. Klaming. 2010. Human enhancement for the common good—using neurotechnologies to improve eyewitness memory. *AJOB: Neuroethics* 3(1): 22–33.

Verscheure, Bruno, Gershon Ben-Shakar, and Ewout Meijer (eds.). 2011. *Memory detection: theory and application of the concealed information test*. Cambridge: Cambridge University Press.

Chapter 14
Morally Responsible Decision Making in Networked Military Operations

Christine Boshuijzen-van Burken and Bart van Bezooijen

Abstract Introducing responsible innovations on the battlefield requires a rethinking of social and psychological aspects of moral decision making on the battlefield, and in particular, including how these aspects are influenced by technology. In this chapter, the social aspects of moral decision making are accounted for in terms of the normative practices in which soldiers do their jobs. Soldiers on the battlefield are embedded in a very specific structure, and are expected to act according to rules, norms and procedures. Their actions are inspired by a certain worldview, which influences the way in which the rules, norms and procedures are interpreted. Technology, especially ICT, connects different practices on the battlefield, thereby creating a network of different (sub-)practices. This may cause a blurring or clashing of different normative practices, which affects moral decision making. In this chapter, Remotely Piloted Aircrafts (RPAs) are used as a case in point for technologically mediated moral decision making. The normative practice model gives insights in the social aspect of decision making in networked missions, but it does not pay attention to the role of the individual soldier in an in-depth way. Therefore an addition is needed, which focusses on the individual soldiers themselves. For the individual level, we take the psychological component of moral decision making and explain how this aspect is affected by technology. The model of normative practices is thus informed by insights from empirical psychology. Moral psychologists have empirically investigated how certain cues influence moral decision making. Some of the cues can be effectuated through technology. Social cognitive theory, as developed by Bandura (Social foundations of thought and action: A social cognitive theory. Prentice-Hall, Englewood Cliffs, 1986), (Personality and Social Psychology Review 3(3):193–209, 1999)

C.B. Burken (✉) · B. van Bezooijen
Philosophy and Ethics, School of Innovation Sciences,
Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: christinevanburken@gmail.com

B. van Bezooijen
e-mail: vanbezooijenbart@gmail.com

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_14

and moral intensity theory developed by Jones (Academy of Management Review 16(2):366–395, 1991) are theories that explain moral decision making mechanisms in terms of respectively moral (dis)engagement mechanisms and the perceived moral intensity of a situation. From both theories we infer how visual data sharing technologies can increase or decrease morally appropriate decision making in networked enabled operations.

Keywords Moral decision making • Normative practices • Network enabled operations • Remotely piloted aircrafts (RPAs) • Social cognitive theory • Moral intensity theory

14.1 Introduction

Is it less morally acceptable for a soldier to kill a man either in firefight in Afghanistan than through pushing a button drone pilot cubicle in the homeland? While the net result is identical (one man has died), there is compelling empirical evidence that the underlying moral decision making processes differ. When technologies make it easier to kill, the latest military innovations such as drones have an inherent moral dimension. The question therefore is not arguing whether it is morally better to kill someone with or without a high tech weapon system, but to see how technology influences military personnel that make such decisions. In order to design technological innovations on the battlefield in a responsible manner, research on the impact on its users is needed. This fits in a product approach to responsible innovation (see Chap. 1, this volume), in the sense that during the development of new technologies on the battlefield, social and psychological factors should be included, in order to make military innovations more responsible. Modern soldiers do not work with one sole technology, but are embedded in a working environment in which technology is ubiquitous. The technological make-up of military working environments may influence their moral decision making. Possible effects of technology on moral decision making may however not be easily understood. Let us introduce the comments about the use of force in recent operations in Afghanistan by Britain's Prince Harry, who was a helicopter gunner, and that of a US drone pilot in the same military mission. While Prince Harry described his tasks as: "It's a joy for me because I'm one of those people who loves playing PlayStation and Xbox, so with my thumbs I like to think I'm probably quite useful," (see Crilly 2013), the drone pilot described his tasks as "I see mothers with children, I see fathers with children, I see fathers with mothers, I see kids playing soccer," before the call comes to fire a missile and kill the man (see Bumiller 2012). Clearly the role of technology is more complex than that pressing buttons make it easier to fire lethal missiles.

So how does technology affect decision making? We aim to answer this question by first providing a structural analysis of the technologically mediated context in which soldiers make decisions. This structural analysis has been used before to understand the nature of complexities in military contexts. We then enrich the theoretical analysis with empirical research from the realm of psychology that encompasses technology and decision making. We focus our inquiry on military operations in which information and communication technology (ICT) is being used, so-called networked military operations, because in these settings technology and moral decision making are deeply intertwined. Information is shared between different members of the network, on the basis of which moral decisions are made. As a case in point for innovations on the battlefield, we use Unmanned Aerial Vehicles (UAVs).

The aforementioned examples of Prince Harry and the drone pilot were unthinkable a decade ago. The fact that one can draw an analogy between the killing of a real person with playing a video game—at least if we phrase the analogy in terms of bodily movements, namely manipulating a stick and buttons while watching a video screen in an airconditioned room—says something about how military practice is changing. Military practice is technologically mediated by all sorts of innovations. In the two examples (Harry and the drone pilot) in the introduction this became clear. The analogy with the play station shows that innovations on the battlefield are often mediated by the screen and that soldiers are acting from a geographic distance. Albert Borgmann (1984), a philosopher of technology, brought these developments under attention by phrasing them in terms of alienation powers of technology. Royakkers and van Est state: "For a soldier in combat fighting an enemy is something that costs a lot of effort, literally 'blood, sweat and tears'. Remote control warfare has gone rid of the 'blood and sweat'." (2010: 292). Borgmann considers it a loss to be detached from the means that bring certain ends. It may be that network enabled operations lead to a furthering of what he names 'moral commodification'. How can we further understand the ways soldiers deal with innovations on the battlefield, taking it beyond a rather pessimistic view of how technology negatively affects individuals and society? The 'empirical turn' in philosophy of technology attempts to approach technology from an internal perspective, emphasizing empirical facts about technology that can be used in ontological, epistemological, ethical or more general discussions in the philosophy of technology (Kroes and Meijers 2000). In this article, the empirical turn is taken a step further by bringing empirical data from a different discipline, namely psychology, to discuss responsible innovations on the battlefield. We take the case of UAV pilots and perform a structural analysis. Later we enrich the structural analysis with empirical data from psychology. The structural analysis starts from the assumption that soldiers do not decide and act in a 'void', but in a very specific context. We call this context a normative practice (more on this later) and our structural analysis is therefore called a practice-analysis. A soldier is embedded in a normative structure of a practice, namely military practice. A practice-analysis has previously provided insights in complexities in network enabled operations (van Burken and de Vries 2012). In this chapter we aim to enrich the practice analysis with empirical data.

The chapter is organized as follows. We provide a philosophical account of the UAV practice in the first part of the chapter. It is an evaluation of moral decision making in light of the conceptual framework of normative practices. It shows the structure of decision making processes from a theoretical perspective and is, to some extent, prescriptive. In the second part of the chapter we add two theories or lines of research from psychology that give insights into how moral decision making actually takes place in a technological mediated environment. From the experimental data we extract what knowledge this may bring us for understanding how technology changes military practices in networked military operations and if this change is desirable.

14.2 Normative Practice

In the introduction we remarked that soldiers do not make their decisions in a void, but that they are embedded in a specific social practice, namely military practice. Military practice (as well as other practices, such as medical practice, business practice), comprises norms, rules, and procedures which people who work in this practice (ought to) adhere to. Therefore we call them normative. The concept of 'normative practice', was developed by Jochemsen and Glas (1997) (see also Jochemsen 2006). Their understanding of social practices is that they have specific law-spheres of their own and that the people who work in the practice are therefore structurally bound by specific rules, norms and principles. The normative structure of a social practice is not only the boundary of the practice, but also has a constitutive function. Some rules and norms and procedures are what makes the practice that specific practice, in the same way as the rules of the game of chess make what chess is a game. The rules, norms, procedures are called the structural side of a practice. The end of the practice, or telos, also belongs to the structural side of the practice. For the military practice the rules and norms are related to the use of force and often reveal the hierarchical nature of the organizational aspect of the practice. A practice has a directional side too. This is the way in which people who work in the practices 'open up' the structure, in other words, how they interpret the rules and norms that hold for the practice. This is often inspired by a persons' cultural or religious background and is related to worldview, or ethos. A soldier may, for example, 'use force for the protection of one's country', or 'fight for freedom'. This affects the way moral decisions on the battlefield are made.

We further propose that soldiers work in different 'sub-practices', for example the sub-practices of pilots, reconnaissance soldiers, ground commanders, etcetera. For each of the sub-practices there are different norms, rules and procedures. For example, a reconnaissance soldier has different rules and norms with regard to safety (e.g., there are rules on how to search a village) than a pilot (e.g., there are rules on where to bomb and where not). Some rules and norms hold for all sub-practices, for example the Geneva Conventions. It is obvious that the available technologies demand new norms or adaptation of norms, given that these technologies are embedded in a practice. For example, when social media became available for sharing information, a rule was established on how to deal with confidential information on the internet in military practice.

14.3 Case Presentation: Remotely Piloted Aircraft

The following case, described by Laster and Iannotta (2012), will be used to analyze a specific case of network enabled operations, namely that of UAV. In the military, the preferred way of speaking about UAVs or drones, is Remotely Piloted Aircraft (RPA). The type of RPA that was used is a Predator, which is an armed aircraft, piloted from a distance.

> A Predator pilot in California saw a "splash" on his video screen as a Hellfire missile slammed into a small group of people thought to be Taliban fighters. Almost immediately, he knew something was wrong.[…] The morning of April 6, 2011, in Afghanistan's Sangin River Valley turned out to be a Predator pilot's worst nightmare, and one that prompted intelligence analysts to gain a greater voice during the fast-paced discussions preceding the firing of weapons. […]. [Given the situation, the Predator crew was furiously scanning for targets to strike with one or more of the plane's Hellfire missiles. […]
>
> The Hellfire missile killed [a US marine and sailor]. An [investigative report] redacts the names of those involved and finds no one negligent. But it recommends changes in Predator processes that in January were blessed in a formal Joint Urgent Operational Need statement. […]. In the Sangin case, the mission intelligence coordinator hesitated to bring the controversial whisper chat [about conflicting assessment regarding the direction of fire away from friendly forces] to the attention of the pilot. The coordinator told investigators the crew was trained not to address whisper chats "during a dynamic situation." […] "You have to understand, the Sangin Valley is bad-guy territory," […] "If you don't know where the friendlies are, it's pretty difficult for you to know [what] to overturn based on what one of the supporting forces thought." (Laster and Iannotta 2012: 24–27).

In order to be able to make a structural analysis of the incident, we quote Fitzsimmons and Sangha (2013), who describe a typical RPA crew.

> Both [Predator and Reaper] aircraft are operated from ground control stations by teams consisting of a pilot, a sensor operator (SO), and a mission intelligence coordinator (MIC). Seated side-by-side in a trailer on any one of the eight USAF facilities on U.S. soil where combat RPA units are currently based, each of these individuals faces three monitors: one displaying live footage from the RPA's cameras, another displaying data pertaining to flying the aircraft, such as its altitude and fuel level, and another displaying an array of other data. While the pilot is responsible for flying the aircraft and launching weapons, the SO conducts reconnaissance and, using the aircraft's laser targeting system, guides weapons into targets on the ground. The MIC coordinates the flow of information between the aircrew (the pilot and the SO) and outside sources, such as any ground units the RPA is supporting in theatre. Collectively, these individuals perform a number of valuable functions, including providing real-time surveillance to personnel around the globe; guiding and protecting ground forces from enemy attacks; locating and eliminating weapons caches and enemy combatants; and conducting damage assessments following airstrikes…. (Fitzsimmons and Sangha 2013: 3).
>
> [I]t is important to recognize that RPA operators do not decide if and when to launch hunter-killer missions. Rather, they conduct the missions and fire weapons at the targets that their commanding officers order and authorize them to engage with. (Fitzsimmons and Sangha 2013: 12).

14.4 Practice Analysis of the RPA Case

In this ‘friendly fire’ incident, resulting in the death of two allied soldiers, some unfortunate decisions were made. How can we understand why and how this decision to engage was made? And how did technology play a role? For this, we turn to a practice analysis. We are not claiming to give a full account of the incident, but we use it to illustrate how a practice analysis sheds light on why people (had to) make decisions, instead of going to a quick conclusion about the lack of technological assets, such as Laster and Iannotta (2012) suggest. We find that in the Sangin case, there are different sub-practices at work. There is (at least) the pilot-practice, the sensor-operator-practice, mission intelligence coordinator and a joint tactical air controller (JTAC, or ground controller). Below we give a structural analysis of each of the practices and distinguish between the structural and directional side of the practice. Most information for this analysis is derived from newspaper articles or online sources, from both critical and non-critical nature, to ensure a more or less balanced view on the practices.

RPA Pilot:
Structure:
- flying the aircraft and launching weapons
- rules for classification [unlike other military operations (Fitzsimmons and Sangha 2013: 11)]
- Geneva Conventions and other legal procedures
- under authorization of commanding officers
- working in shifts
- gut-feeling check

Direction:
- protect the people on the ground (Schogol and Ricks 2012)
- fear of repercussions in case of collateral damage (Groetken 2010)

Sensor Operator:
Structure:
- conducts reconnaissance
- guides weapons into targets on the ground
- Geneva Conventions and other legal procedures
- working in shifts
- under authorization of commanding officers

Direction:
- making a difference in the world (ibid)
- job satisfaction from releasing bombs (Schogol and Ricks 2012)
- fear of repercussions in case of collateral damage (Groetken 2010)

Mission intelligence coordinator:
Structure:
- coordinating flow of information between the air-crew (the pilot and the SO) and outside sources

- rules for classification
- under authorization of commanding officers
- Geneva Conventions, Rules of Engagement (ROE) and other legal procedures (Walker 2010)
- working in shifts

Direction:
- coalition forces should not be injured or killed
- trust but verify information (Laster and Iannotta 2012: 26)

Ground Commander:
Structure:
- release authority
- standing operation procedures

Direction:
- priority is saving lives of own troops (Corcoran 2012)

A practice analysis emphasizes that decision making in these practices should be seen in the context of rules and norms and procedures and also in the context of a specific aim (telos) of the sub-practice. People in the different sub-practices have a degree of freedom to interpret the structural side of the practice. For example, the mission intelligence coordinator had the freedom to withhold or present the whisper chat from or to the pilot. Especially with new technologies such as RPA, sometimes the rules and norms are insufficient for guiding what is the best decision. In this case, the friendly forces divided into two groups, but from a black-and-white screen in California it was unclear which group was now the friendly and which was the opposing force. The coordinator chose to withhold the whisper chat about this confusion, since the attitude amongst the coordinators was to 'trust but verify' the information (Laster and Iannotta 2012: 26). This trust attitude belongs to the directional side of the practice. It is suggested that, depending on one's cultural background, people are more or less likely to trust others, especially those one has never met (Fukuyama 1995). In networked operations, decisions may be influenced by the different levels of trust people have towards others on the network, especially when the sub-practices are geographically and culturally apart. When investigators asked the pilot in the Sangin case what he had done, had they presented the whisper chat about the concerns, the pilot told them he was "90 percent likely" to have informed others and would have "considered putting a hold on the engagement." (Laster and Iannotta 2012: 26).

This last remark from the pilot is an important aspect of the structural side of the pilot practice. As part of the procedure, a pilot needs to perform a 'gut-feeling check'. If the pilot has any doubts about the planned attack, he should abort the engagement. It seems that for the others in the RPA crew this gut-feeling check is not part of their procedure. A former sensor operator confesses that during the many engagements he took part in, some situations could not have passed this 'gut-feeling' check, had it been in his procedure. He recalls one engagement and states that he "wasn't convinced that they were bad guys.", but as a young airman, "he didn't think he had the standing to ask questions" (Engel 2013). What a practice analysis reveals here, is that in traditional piloting, the ground commander and the pilot could abort the mission if they had doubts regarding the engagement (for example, in case they suspected that the strike was not according to the rules of engagement). In the RPA missions however, with much more people involved, who have diffused tasks and responsibilities, an appeal to the gut-feeling check may be hindered. This can be due to hierarchical reasons, such as the young airman suggested above, or simply because it is not accounted for in the structural side of the sub-practice. In other words, there is no part in the procedure that calls attention to the gut-feeling check. Decision behavior of soldiers and airmen cannot be accounted for by the rules and norms, or the structural side of the practice alone. The directional part plays an important role too. If the young airman from the quote had had a different understanding of his position and role in remotely piloted missions in general (i.e. direction, or ethos), instead of doing his job from the perspective that his role wouldn't matter in the decision making process, he may have raised questions, and thus perhaps have prevented airstrikes in which civilians died.

Much more can be said about the friendly fire case described above from a normative practice approach, but we conclude this section with two points. First, that soldiers and airman are embedded in normative practices, in which certain rules and norms are constitutive for the decisions and actions in that practice. The way in which the rules and norms are interpreted are the directional side of the practice, which allow for differences in decision making. These differences in decision making can become problematic and difficult to overcome in case one does not understand that different worldviews may underlie a specific interpretation of the rules and norms (note that there is a difference between interpreting a rule or norm differently and actual breaking a rule or norm). Second, that due to the innovations on the battlefield new (sub-)practices emerge. These new sub-practices do not always have their structure in place (yet) for the difficult task that they are supposed to do. The view that the RPA pilots and analysts who are working on the remote end do exactly the same things as manned planes or analysts close to the battlefield is too simple. The practice of an RPA pilot is different from a traditional pilot, because a technology mediated environment requires different rules and norms for responsible decision making, as the case made clear.

People who are working in the newly evolving sub-practices, which are inevitably networked, give normative structure and direction to their practices. If this is done in a just manner, it supports responsible decision making in networked missions. Clashes between practices, or one practice potentially overruling another practice, is then recognized as undesirable before big accidents happen, such as the friendly fire incident.

The theoretical model that we used to point to problems in this case is a normative one, since it lays bare the normative structure (norms, rules, ends, etcetera) of social practices. In this chapter, it functioned in a descriptive-analytical way. It can have a prescriptive function too, if we decide that the people who are working in the practices should act according to the normative structure of the practice.

Our practice analysis does not say anything about how people actually behave in technologically mediated practice. Therefore we need empirical evidence from psychological research.

14.5 Moral Psychological Theories

As has been shown in the previous section, a practice has a structural side (the rules, norms, procedures and telos), but also a directional side (is connected to a certain worldview, or ethos). People with different perspectives on what is good and what is bad are working in the practices. Analyzing moral decision making in terms of this conceptual framework gives us a broader understanding of why soldiers (have to) make certain decisions. We concluded that some rules or norms are lacking or are inappropriate for network enabled missions (for example the gut-feeling check rules that are absent for sensor operators, or the norm about whisper chats is often difficult to interpret) and therefore people make unfortunate decisions. What is missing in this perspective actually is that not only there are rules, norms and underlying worldviews that steer behavior in the practices, but there is also a psychological aspect that plays a role in acting and deciding. For example, technology enabled the soldiers in the Sangin case to stay physically distant, while having lethal capacity to destroy persons or villages that they assigned as 'bad' and that were thousands of miles away. The concepts of good and bad may be inspired by a more fundamental view of the world (direction), however, this does not explain how assigning certain people as 'bad', resulting in specific behavior, is influenced by technology. For these questions we need psychological tools that can point out behavioral patterns. Also, whether the remote way of working, away from where the bullets fly, makes it easier or less easy to use force are questions that can be answered by looking at behavioral responses from empirical psychology. Another example from the Sangin case, in favor of our point to add psychological reflection, is that the behavior described as 'furiously scanning' (Laster and Iannotta 2012), which reflects a specific cognitive state, may be activated and upheld by technology. Therefore technology plays an active role in decisions that follow from the 'furious scanning' that cannot be explained in terms of structure and direction alone. In the following sections we clarify some relationships between the way the military practice is technologically mediated and psychological processes that partially underlie the decision behaviors of soldiers who work in technologically mediated environments. It adds empirical reality to the normative practice model. It should be noted that in the discussion of the psychological theories, the Sangin case will get little attention as a case, but in the section below we show how technology affects the practice of RPA in a more general way. It is a practice in which the use of lethal force is sometimes demanded. Can it be the case that technology such as UAVs makes it easier to use lethal force, thus influencing moral decision making?

14.5.1 Social Cognitive Theory

We searched within moral psychology for theories, concepts, or lines of research that could help us to think about the use of technology in network enabled military operations. Below we present two theories and lines of research, discuss our findings and formulate a subconclusion on the relevance of these theories for networked military operations.

One psychological theory that has been applied often in the context of military operations is Bandura's (1986, 1991) social-cognitive theory (SCT). The basic premise of SCT is that: "(…) moral thinking is a process in which multidimensional rules or standards are used to judge conduct. Situations with moral implications contain many decisional ingredients that not only vary in importance, but may be given lesser or greater weight, depending on the particular constellation of events in a given moral predicament" (Bandura 1991, pp. 64–65). In other words, what keeps people from behaving inhumanely and stimulates people to behave humanely is not only determined by moral reasoning, but other factors such as standards, motivation, and self-control also play a large role. An important distinction between SCT and other well-known moral psychological theories (e.g. Kohlberg 1969; Rest 1979, 1986) is the emphasis on mechanisms other than moral reasoning for describing why people behave (im)morally. Bandura proposed that people have self-influence mechanisms that control behavior, for instance violating one's moral standards will bring self-condemnation. These self-regulative mechanisms, either positive or negative, are a key component of SCT.

Self-regulation, however, requires effort, and people may not be motivated to engage in self-sanctioning at all times. From this it follows that selective activation of self-regulatory mechanisms and disengagement will lead people to immoral behavior. Bandura formulated seven ways in which people can disengage their moral self-sanctioning from their behavior. We will briefly discuss these disengagement mechanisms below (based on Bandura 1999).

The first three mechanisms are related to the reconstrual of behavior (Bandura 1986). People may make immoral behavior acceptable by portraying it as socially worthy or serving moral purposes (moral justification). People may sanitize or camouflage their immoral behavior (euphemistic labeling). People may further compare their behavior with other behavioral options that would have been worse (palliative comparison). The second set of disengagement mechanisms is related to the structuration of personal responsibility. People may obscure or minimize the effects of their behavior (displacement or diffusion of responsibility). The power of this mechanism, some have argued, has been demonstrated in atrocities such as the My Lai massacre, the Nazi mass executions, and the Milgram studies. Here, people displaced the responsibility for the immoral conduct. The third set of mechanisms is to minimize, disregard, or misconstrue the consequences of one's behavior. It is easier to pursue harmful actions when minimizing the effects of actions, or to discredit the sources of the reported harm. The final set of disengagement mechanisms is related to the recipients of the harmful acts, for instance by treating the recipients as being less than human (de-humanization) and making recipients partially responsible for the harmful actions (attribution of blame).

14.5.1.1 Relevance for Networked Military Operations

All disengagement mechanisms have been used by others to think about the morality of actions in military operations (for a discussion, see Bandura 1999). When zooming in on the mechanisms that may be affected by technology, we consider three mechanisms to be relevant: diffusion of responsibility, disregarding harmful consequences, and dehumanization. We briefly discuss these mechanisms below.

First, technology offers new possibilities to organizations for diffusing or displacing personal accountability for immoral actions. For instance, dividing labor so that work becomes more routinized is a powerful way to diffuse personal accountability. We can learn from Corcoran (2012) that there is at least a team of 10 people involved in flying an unmanned aircraft in combat, each with their own tasks and responsibilities. "In the Kandahar control room are a pilot, a pay load operator and an electronic warfare section comprising intelligence officers, technicians and locally hired "cultural advisers" and linguists". That some of those members get routinized in their job, potentially leading to a dehumanization process, becomes clear from the following quote: "After participating in hundreds of missions over the years, Bryant said he "lost respect for life" and began to feel like a sociopath. He remembers coming into work in 2010, seeing pictures of targeted individuals on the wall—Anwar al-Awlaki and other al Qaeda and Taliban leaders—and musing, "Which one of these f_____s is going to die today?" (Engel 2013). This last remark entails a part of the process of dehumanization that is not necessarily influenced by technology, but ICT makes this scene a real option. Bandura calls it 'euphemistic labeling', and here it is the effect of sticking pictures of people on the wall of your office, whom you will never meet face to face, that you may kill the same day.

Further, relocating the 'dirty' work to one group of people, let us say drone pilots, make that only drone pilots have to cope with the outcomes of their offensive acts. There are numerous quotes available from drone pilots who say that they are the ones that see the nasty part, for example this former sensor operator: "People say that drone strikes are like mortar attacks," Bryant said. "Well, artillery doesn't see this. Artillery doesn't see the results of their actions. It's really more intimate for us, because we see everything." (Engel 2013).

Second, technology may also be used for disregarding or distortion of consequences. To use Bandura's own words: "Our death technologies have become highly lethal and depersonalized. We are now in the era of faceless warfare, in which mass destruction is delivered remotely with deadly accuracy by computer and laser-controlled systems' (1999: p. 199). Not only does Bandura emphasize the role of technological mediation between actor and recipient in these words, he also illustrates the working of a third mechanism of moral disengagement, dehumanization. Dehumanization occurs when technology estranges military personnel from the people who are affected by the outcomes of their actions. In other words, this mechanism could occur for military personnel that are working with 'depersonalized' technologies, fighting a 'faceless warfare' watching dots or thermal images on a screen. Grossman has referred to this as 'mechanical distance' created by technology, enabling soldiers to deny that they are killing humans but instead are attacking a group of pixels (Grossman 2009: 181). In sum, SCT generally holds that military personnel in networked military operations may disengage their moral standards when fighting a faceless enemy.

However, SCT also accounts for one way in which technology leads to moral engagement: humanization (see Bandura 1999). Within SCT, humanization holds that when people see the suffering that they have caused, and when people are made aware of the social linkages between moral actor and recipient, this makes actions personal (thereby inhibiting disengagement mechanisms). One example of this would be one of Milgram's (1974) studies in which participants directly witnessed the outcomes of their actions for the recipient of electric shocks. This humanization of the recipient dramatically lowered the tendency of participants to cooperate with the experimenter. In networked military operations, humanization would explain why the hair on the back of a drone operator's neck stood up when he had to fire a missile, just as it did when he used to line up targets in his F-16 fighter jet: one can get immersed in a digitized version of reality (see Bumiller 2012). An illustration of this humanization aspect of technology is given by an official, who did not want to be identified, in Schogol and Ricks (2012). As UAV operators, they observed a bomb-maker for weeks to find out who was financing the bombs and providing explosive materials.

> "We watched him wake up in the morning; we watched him leave for work in his vehicle; we tracked him to where he was building these weapons; we watched him eat lunch; we watched him go home and play soccer in his yard with his family—with his two little girls," the official said. "We watched him live with his wife, watched him sleep, we watched him sleep; we watched him get up in the middle of the night, go to the back of his house and build weapons."
>
> So the drone operators knew this man well when the time came to kill him.
>
> "We've been watching him for so long that we have that part of the history with our operators, who are having the thought in their head of, 'I don't care what you think of this individual, he does have two daughters; I have seen him with his family,'" the official said. (Schogol and Ricks 2012).

Unfortunately, we were unable to find studies that estimated the relative weights of disengagement mechanisms, or empirical studies on humanization, to date. Studying (de-)humanization of technology and the relative effects of following a person for a longer time, (thus getting familiar with the daily routines of those whom are potential targets) needs to be done. Apart from quotes such as given above, the claims of SCT for the effects of technology on moral decision making remain theoretical at this point.

14.5.2 Moral Intensity

Similar to Albert Bandura, Thomas Jones emphasized that moral actions do not merely follow from moral cognitions. Jones (1991) developed an issue-contingent model and recognized the influence of multiple characteristics that are associated with the ethical situation itself that influence moral decision making, which make up the dimensions of moral intensity. Moral Intensity theory holds that six characteristics collectively comprise moral intensity: (1) Magnitude of Consequences (2) Probability of Effect (3) Proximity (4) Temporal Immediacy, (5) Social Consensus and (6) Concentration of Effect. According to Jones' model, "[a]s the moral intensity of a situation increases, awareness of the ethical nature of the situation should increase; judgments regarding the appropriate action to be taken in the situation should lean more toward ethical action; behavioral intention should be to act in a more ethical manner; and behavior should be more ethical" (McMahon and Harvey 2006: 352). A research on factor structure of moral intensity revealed that moral intensity has three, rather than six dimensions. McMahon and Harvey (2006) distinguished between: probable magnitude of consequences (Jones's dimensions 1, 2 and 4), proximity (Jones's dimension 3), and social consensus (Jones' dimension 5). Moral intensity holds that insofar consequences of moral actions are: severe, certain, and immediate (Factor 1), psychologically close (Factor 2), and socially acceptable (Factor 3), this increases the likelihood that individuals will behave morally.

14.5.2.1 Relevance for Networked Military Operations

Technology potentially plays a role in all of these dimensions. This can be exemplified when this model is taken into the context of networked military operations.

First, the probable magnitude of consequences is directly related to the type of weapon or ammunition, which defines the magnitude of the consequences and the probability of effects. Further, the networked military operations are focused on increasing the speed of decision making through the use of ICT systems, which raises the temporal immediacy of outcomes. This was the case in the friendly fire incident described in the previous section. "With the clock ticking down toward the missile launch, the mission intelligence coordinators scurried to assess the conflicting information about the direction of fire" (Laster and Iannotta 2012: 26). The intelligence coordinators knew that the impact of the decision would be very immediate. These outcomes increase moral intensity.

Second, proximity holds that physical distance between actor and recipient is negatively related to moral intensity. Since spatial distance between actor and recipient (or other actors) is dramatically increased in current operations, this suggests decreasing moral intensity in networked operations. In literature, this objection is often heard, see for example Royakkers and van Est (2010). However, looking at what the UAV pilots themselves state about this distance, the opposite may be the case. Technology may increase the physical distance between the pilot and the target, but people do not perceive this as a great distance. Rather the opposite, as the following quote illustrates. "You are 18 inches away from 32-inch, high-definition combat, where you are in contact [by headset with] the guys on the ground," the official said. "You are there. You are there. You fly with them, you support them and a person you are tasked with supporting gets engaged, hurt, possibly killed, it's a deeply, deeply emotional event. It's not detached. It's not a video game. And it's certainly not 8000 miles away" (Engel 2013). Therefore, the moral intensity may be increased due to technology, because the perceived distance is 'only 18 inches'.

Third, 'mixing and matching' of military units such as soldiers and sailors, or such as different nationalities in networked operations, leads to a potential decrease of social consensus between involved parties, thereby lowering moral intensity of the situation. However, one of the quotes from an RPA official suggests that there is not much concern for the decrease of social consensus amongst those involved in missions: "We have a lot of other agencies that we're working with, whether we're supporting them from above with them on the ground or other agencies that we're coordinating with during the missions—not too concerned that amongst the agencies that we're working with or supporting that they will be able to make right decisions." (Schogol and Ricks 2012). From this quote we learn that within network enabled missions, members assume that the parties in the network are able to make right decisions, like they would make themselves. However, an empirical test whether this is really the case has not been done.

14.5.3 What have we Learned from Psychological Theories?

The previous section discussed potentially relevant psychological theories contributing to a better understanding of moral decision behaviors of soldiers in a networked mission. These concepts were chosen because they reveal potential changes in behavioral patterns of people when they are exposed to situations that are morally laden and in which technology is or can be involved. Depending on the specific dimension under study, the moral intensity of the situation can either be increased (e.g., using weapons with enhanced power and precision) or decreased (e.g. using weapons operated from the homeland). Moral intensity offers good ways to link morality, decision making, and technology, and the theory was identified as a powerful theory to describe moral behavior in organizations (for a meta-analysis, see Kish-Gephart et al. 2010). Similar to SCT, moral intensity theory does account for the different effects of technology on moral decision making, but the relative weight of the dimensions for our inquiry remain unclear.

The empirical research further provides insights that are helpful for designers of innovations on the battlefield. Taking into account the way in which (characteristics of) technologies affects decision making can inform designers in such a way that technologies support responsible decision making on the battlefield. This is our proposal in support of realizing responsible innovations on the battlefield.

The focus of most empirical research has been on what technologies do to individuals. We should keep in mind that behavior of soldiers cannot be derived from a narrow focus on psychological mechanisms alone. Therefore, we introduced the framework of a normative practice. People do not make decisions 'out of the blue', but they are often embedded in a practice. The technologies that are used on the battlefield do not stand on their own, nor in a one-to-one relationship with an individual, but they are embedded in the practice. We have enriched the model of a normative practice with empirical psychological data. It broadens the discussion on moral discussion making from the social level to the level of the individual in the practice.

14.6 Conclusion

This chapter aimed to provide insights in moral decision making on the battlefield where technology is ubiquitous. It added an empirical component, derived from the realm of psychology, to the normative practice model. People who work in a normative practice behave not only according to rules, norms and procedures, inspired by a certain worldview, but there is also a psychological component to their behavior and which plays a role in decision making. In case of technologically mediated practices, this psychological component may play an important but neglected role in the way people make decisions on the battlefield. In this chapter we have informed the model with regard to which kinds of behavior people are likely to expose in a highly technological environment. The military (sub-)practices can then improve the rules, norms and procedures accordingly and perhaps in an early phase discard certain technologies in the practice, since their (subconscious) effect on decision making behavior is detrimental for the practice. If for example, it turns out that soldiers are more likely to bypass certain rules and procedures that are important to distinguish between soldiers and civilians, these technologies may do more harm than good in the practice. This chapter addressed which psychological responses can be expected from people in general, when they are exposed to certain technologies, or ways of working in a technologically mediated environment. The rules and norms for hiring soldiers, as well as the rules and norms of the practice, may need to change given the insights from psychological research.

There are different ways of working at the intersection of philosophy and psychology. In this chapter we have discussed the issue of moral decision making in the context of military practice where there is an increased use of technologies that connect people and enable technologically mediated actions. A short introduction of two psychological concepts was given. Social cognitive theory and moral intensity theory provided us with clear descriptions of how people are likely to respond to certain morally problematic situations. It remains difficult to integrate the findings into the case of network enabled operations, because the psychological concepts seem unable to grasp that soldiers do not make decisions ‘out of the blue’, but they are embedded in a practice, namely the military practice. This is a normative practice where specific rules and norms hold and where there is an underlying worldview to the practice. In military practice there exist multiple ‘sub-practices’ that are brought together in new ways in a networked mission. Unforeseen conflicts may arise at the level of structure and direction. The psychological theories could be enriched by doing specific research of which some is suggested in the next section.

14.7 Recommendations

If we take seriously the claim that soldiers are embedded in a practice, research for understanding the moral behaviors of soldiers in network enabled missions should move towards lab research with people in the practice that answers specific questions, such as the questions of proximity. Also, anthropological research, focused on soldiers in their practices, could give better insights in the way soldiers act and decide in the face of an increasingly technologically dominated working environment. Another helpful type of research is case studies and interviews. Current military missions host multiple novel technologies to deal with pressing problems, such as improvised explosive devices, suicide attacks and tracking down of terrorists. Soldiers who work with these technologies, either for reconnaissance or attack, could be a source of information for better understanding the way the military practice changes, making it more explicit what this means for (interpretation of) the rules and directions of sub practices. Interviews may spark off empirical psychological research to study how technology affects moral decision making in military practices.

Some interesting suggestions are given by Sripada and Stich (2006) who bring the interplay between the normative structure of social practices and psychology of individual decision making to the fore. They argue that the philosophical intuitions about the subjective power of norms reflect an empirical truth about the psychology of norms (2006). According to Sripada and Stich, norms have powerful motivation effects on the people who hold them (2006). "People are disposed to comply with norms even when there is little prospect for instrumental gain, future reciprocation or enhanced reputation, and when the chance of being detected for failing to comply with the norm is very small" (Sripada and Stich 2006). This suggests that both structure and direction may be an important intrinsic motivator for moral decision making. Psychological research should move towards investigating how people deal with conflicting norms and underlying worldview of people who work in military practice, because the future of military missions will be in networked missions, where rules, norms and directions can clash.

This chapter may further be informative for reconsidering the psychological profile for soldiers that work in highly technological environments. On the level of mastering physical and intellectual skills there is already a shift going on in the hiring of, for example, UAV pilots. The 'functional' profile that the US Army is looking for nowadays, on the level of intellectual, physical and technical skills, are gamers (The Canadian Press 2013). Knowing better in which ways soldiers are affected by technology in their decision making may inform the hiring policy for people in the RPA job or networked operations in general. Missy Cummings, a researcher at MIT, studies which personality traits would be helpful to deal with boredom, which is a key aspect of flying UAVs and usually gamers are not very good at dealing with boredom. Cummings found that the only personality trait to help predict who will be better UAV operators was conscientiousness. "Those who scored high on conscientiousness did better when the environment became boring" (The Canadian Press 2013). One should further study which 'psychological profile' fits best in this technological mediated battlefield. Currently before a soldier is hired, she undergoes a psychological test to see if the required character traits fit with the job, for example if, according to her psychological profile, she is likely to be trustworthy. In this chapter we suggested that not only the profile on the level of technical and intellectual skills should be reconsidered, but also the level of psychological skills, call it their 'moral fitness' (cf. Richardson et al. 2004).

Acknowledgments This article was written by the authors as part of the research program 'Moral fitness of military personnel in a networked operation environment', which is supported by the Netherlands Organization for Scientific Research (NWO) under grant number 313-99-110.

References

Bandura, Albert. 1986. *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, Albert. 1991. Social cognitive theory of moral thought and action. In *Handbook of moral behavior and development*, ed. W.M. Kurtines, and J. Gewirtz, 45–103. Hillsdale, NJ: Erlbaum.

Bandura, Albert. 1999. Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review* 3(3): 193–209.

Borgmann, Albert. 1984. *Technology and the character of contemporary life: A philosophical inquiry*. Chicago: University of Chicago Press.

Bumiller, Elisabeth. 2012. *A day job waiting for a kill shot a world away*. The New York Times, 29 July 2012.

Corcoran, Mark. 2012. *The kill chain: Australia's Drone War*. Australian Broadcasting Company (ABC) News, 27 June 2012.

Crilly, Rob. 2013. *Prince Harry 'Driving Wedge between forces and Afghan locals'*. The Telegraph, 22 Jan 2013.

Engel, Richard. 2013. *Former Drone operator says he's haunted by his part in more than 1,600 deaths*. NBC News, 17 June 2013.

Fitzsimmons, Scott, and Karina Sangha. 2013. *Killing in high definition*. San Francisco, International Studies Association Annual Convention, 3rd–6th April 2013.

Fukuyama, Francis. 1995. *Trust: The social virtues and the creation of prosperity*. New York: Free Press.

Groetken, Steven. 2010. *Military operations half a world away: The 24 hour use of unmanned vehicles controlled from remote locations*. University of Arizona. http://next.eller.arizona.edu/courses/24hourKnowledgefactory/Spring2010/student_papers/StevenGroetkin.pdf. Accessed 23 Aug 2013.

Grossman, Dave. 2009. *On killing: The psychological cost of learning to kill in war and society*, Rev. ed. New York: Little, Brown and Co.

Jochemsen, Henk. 2006. Normative practices and theoretical ethics and morality. *Philosophia Reformata* 71(1): 96–112.

Jochemsen, H., and G. Glas. 1997. *Verantwoord Medisch Handelen. Proeve Van Een Christelijke Medische Ethiek*. Amsterdam: Buijten & Schipperheijn.

Jones, T.M. 1991. Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review* 16(2): 366–395.

Kish-Gephart, Jennifer J., David A. Harrison, and Linda Klebe Treviño. 2010. Bad apples, bad cases, and bad barrels: Meta-analytic evidence about sources of unethical decisions at work. *Journal of Applied Psychology* 95(1): 1–31.

Kohlberg, Lawrence. 1969. *Stage and sequence: The cognitive-developmental approach to socialization*. New York: Rand McNally.

Kroes, P.A., and A.W.M. Meijers. 2000. *The empirical turn in the philosophy of technology*, vol 20. Jai.

Laster, Jill, and Ben Iannotta. 2012. Learning from fratricide-US deaths in Sangin valley bring long-sought tactical intel changes. *C4ISR-Journal of Net-Centric Warfare* 11(2): 24–27.

McMahon, J.M., and R.J. Harvey. 2006. An analysis of the factor structure of Jones' moral intensity construct. *Journal of Business Ethics* 64(4): 381–404.

Milgram, Stanley. 1974. *Obedience to authority: An experimental view*. New York: Harper & Row.

Rest, James R. 1979. *Development in judging moral issues*. Minnesota: University of Minnesota Press.

Rest, James R. 1986. *Moral development: Advances in theory and research*. New York: Praeger.

Richardson, Rudy, Desiree Verweij, and Donna Winslow. 2004. Moral fitness for peace operations. *Journal of Political and Military Sociology* 32(1): 99–113.

Royakkers, Lamber, and Rinie van Est. 2010. The cubicle warrior: The marionette of digitalized warfare. *Ethics and Information Technology* 12(3): 289–296.

Schogol, Jeff, and Markeshia Ricks. 2012. *Demand grows for UAV pilots, sensor operators*. Air force Times, 12 April 2012.

Sripada, C.S., and S. Stich. 2006. A framework for the psychology of norms. In *The innate mind: Culture and cognition*, ed. P. Carruthers, S. Laurence, and S. Stich, 280–301. Oxford: Oxford University Press.

The Canadian Press. 2013. *Turning video gamers into the ultimate drone pilots. Researchers find that distractions boost performance*. CBC News, 9 May 2013.

Van Burken, Christine G., and Marc J. De Vries. 2012. Extending the theory of normative practices: An application to two cases of networked military operations. *Philosophia Reformata* 77(2): 135–154.

Walker, Andrew. 2010. *1N0 Mission Intel Coordinator*. United States Air Force Intelligence. USAF.IN > 1N0X1 > 1N0 Mission Intel Coordinator ed. 29 July 2010. http://usaf.in/viewtopic.php?id=117. Accessed 23 August 2013.

Chapter 15
Biofuels: Sustainable Innovation or Gold Rush? Identifying Responsibilities for Biofuel Innovations

Annelies Balkema and Auke Pols

Abstract Based on fieldwork and literature review we have investigated the rise and fall of jatropha cultivation in Tanzania over the last decade, and its negative socio-economic and environmental impacts. Based on the fact that the most vulnerable actors, small farmers, were affected the most, through loss of land and income, we conclude that biofuel innovations have so far been irresponsible. In this chapter we draw lessons for future biofuel innovations through the identification of stakeholder responsibilities. We do so by developing a framework which is based on current discussions on the meaning of 'sustainability' and recent ethical work on moral responsibility. In addition, we use the framework to reflect on the jatropha biofuel innovation experiences. Additional fieldwork will be done to gather information on visions and expectations and to discuss responsibilities for sustainable biofuel innovations in Tanzania. Our preliminary conclusion is that stakeholder participation and a clear demarcation of responsibilities are preconditions for sustainable biofuel innovations.

Keywords Biofuels · Sustainability · Moral responsibility · Discourse ethics · Tanzania

15.1 Introduction

Over the last decade, the visions and expectations with regards to jatropha biofuel shifted from a "miracle crop" that could provide sustainable biodiesel for Europe through large-scale cultivation in plantations in the global south, into a "biofuel

A. Balkema (✉)
Technology, Innovation and Society, School of Innovation Sciences,
Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: abalkema@clm.nl

A. Pols
Philosophy and Ethics, School of Innovation Sciences,
Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: a.j.k.pols@tue.nl

B.-J. Koops et al. (eds.), *Responsible Innovation 2: Concepts, Approaches, and Applications*, DOI 10.1007/978-3-319-17308-5_15

niche crop" with relatively low yields suitable for bio-oil production for local use that can best be cultivated on a small scale preferably by intercropping or in hedges on marginal lands (van Eijck et al. 2010; Beyene et al. 2013; Locke and Henley 2013; Segerstedt and Bobert 2013). Looking back at the rise and fall of the jatropha biofuel cultivation in Tanzania, and the associated negative social-economic and environmental impacts, we conclude that these biofuel innovations have been unsustainable and irresponsible. Indeed, jatropha biofuel innovation can be characterised by many of the difficulties that Blok and Lemmens signal in Chap. 2 of this book: major power differences between stakeholders, leading to conflicts and specific problem framings, a narrow focus on technological and economic aspects, and technological optimism and insufficient efforts to protect the most vulnerable stakeholders from the destructive or 'Faustian' aspects of this innovation.

The rise of jatropha was triggered by the European biofuel directive (2003/30/EC) that led to large-scale investments, mostly by foreign investors in large jatropha biofuel plantations producing for export. Few investments have been made in more inclusive business models such as the outgrower model, in which small farmers cultivate jatropha on their land besides food crops for extra income. Most investments relied on subsidies, never made a profit, never reached envisioned cultivation targets, and stopped operating within a few years. We identified 20 jatropha biofuel initiatives in Tanzania in the period of 2008 up to 2013, of which 14 never came off the ground or stopped operation (see also Beyene et al. 2013: p. 11; FAO 2010: p. 37, Table 3.2; Locke and Henley 2013: p. 15; van Teeffelen 2013: pp. 10–11). Two outgrower projects are still in operation at the time of writing, but have not achieved the envisioned upscaling of production. Especially the closing down of the large plantations has led to local farmers losing their food security and their source of income; as they first sold their land and then lost their jobs on the plantation. Even the more inclusive business models did not achieve substantial contributions to poverty alleviation (van Eijck et al. 2010). In addition, there is a negative impact on the environment especially as a result of land clearing for plantations or indirect land use change, affecting biodiversity, soil fertility and causing initial carbon emissions (van Eijck et al. 2010; Bergius 2012).

As such, biofuel innovations encouraged by the EU have contributed to a host of problems that conflict with ideals of sustainability, such as food insecurity, marginalisation of small farmers, loss in soil fertility and biodiversity, and climate change. As we argue in this chapter, the opportunism of scientists, policymakers, and investors has put the food security of vulnerable farmers at risk. This raises the question how the actors involved could avoid similar mistakes in future biofuel innovations.

Sustainability requires us to act in ways such that problems are actually solved, not just being exported to become the burden of another stakeholder group in another geographical location or another time period. For greening the EU transport sector, this means that a sustainable solution should result in less emissions without causing other problems such as food insecurity, marginalisation of small farmers, loss in soil fertility or biodiversity, in the EU or other continents, now or in the future. So far voluntary standards, defining sustainability criteria, voluntary social corporate responsibility, etc. have proven to be not good enough (Partzsch 2011; Levidow 2013). Therefore, well-defined responsibilities are needed—an observation also made elsewhere in this book with respect to responsible innovation (Setiawan and Singh, Chap. 12). This raises the question of who can be held responsible for what in these international innovations?

In order to answer this question, in this chapter we develop a theoretical framework based on current discussions on the meaning of 'sustainability' and recent ethical work on moral responsibility. We use this framework to reflect on the experiences in Tanzania. We will do additional field work to gather more information and to propose and discuss responsibilities for sustainable biofuel innovations in Tanzania, in the next phase of our research. As such, this chapter takes a product approach to responsible innovation, where it draws on theorizing on sustainability and moral responsibility to give recommendations for different stakeholders (cf. Koops, Chap. 1 of this book). Some of these recommendations, though, pertain to processes, such as the need for the EU to reflect more critically on its own values during the policy-making process.

15.2 Theoretical Framework to Identify Responsibilities for Sustainable Innovation

We have developed the theoretical framework that helps us to identify responsibilities for sustainable innovations in three steps, namely by describing the concept of sustainability at first, followed by reflecting on moral responsibility and finally by using the insights in a reflection on the past experiences with jatropha biofuel in Tanzania. These steps are described in the following three paragraphs.

Step 1: **Operationalization of the sustainability concept**

The concept of sustainable development is based on the observation that economy, environment and well-being can no longer be separated. Definitions for sustainable development often sketch a concept rather than giving an unambiguous restrictive parameter that can be applied right away. For instance, the definition of sustainability formulated within the framework of the World Conservation Strategy is: 'Improving the quality of human life while living in the carrying capacity of supporting ecosystems', see Fig. 15.1 (UNEP 2012). As such, sustainability can be interpreted differently by different people, evoking the critique that the term sustainability could mean almost anything (Mitcham 1995). However, the room left for interpretation can also be valuable as ideas about sustainability are destined to be discussed over time and place, as different generations will have to deal with different problems and different cultures and local circumstances will give a different perspective on these problems.

For simplification, we use the sustainability concept with the three dimensions; (1) Social-cultural (People), (2) Environment (Planet), and (3) Economics (Profit). This simplification helps us in identifying the differences in sustainability priorities

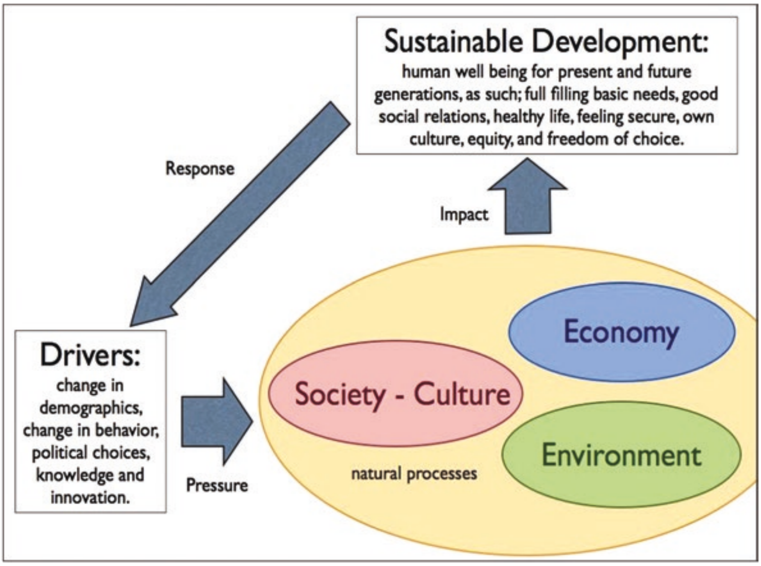

Fig. 15.1 Conceptual framework for sustainable development (based on the DPSIR in UNEP 2012: p. XX)

by different stakeholders or over time in an innovation trajectory. For instance, as Rosling points out, capitalists favour economy, as in Western societies we are still steering development on economic growth while this is a means, not a goal, for development; goals are for instance health, culture, protection of human rights and biodiversity (Rosling 2006, 2007). Similarly Jackson (2009), in his book "Prosperity without Growth", states that our addiction to economic growth, based on debts, and our overconsumption has resulted in social exclusion (unemployment) and inequality (poverty as a result of uneven distribution of wealth) as well as depletion of natural resources. His alternative, the New Green Deal, is a stable economy investing in renewable resources, labour-intensive services and social cohesion. According to Jackson, political intervention is an important means to reach this goal. As in the current western economic system, profit making (economic dimension of sustainability) will be reflected upon as a private responsibility mainly, while the people (socio-cultural dimension) and planet (environmental dimensions) are mainly seen as societal responsibility. This is important in our stakeholder assessment later on.

Another complex question regarding sustainability is how to deal with risks and uncertainties with respect to future generations. Hermele et al. (2009), raise the question why the future is less valuable than the present. However, we would like to draw attention to the fact that it is estimated that one billion people will still be living in extreme poverty in 2015. Hermele et al. (2009) try answering their question on future values by looking at the choice of discounting in reports such as Stern (2007, discount rate of less than 1 %) and Nordhaus (2008, a discount rate of 4 %), pointing out that by opting for a high discount rate, we in effect 'place a larger weight on consumption now over the effects on future generations', but so might future generations, which means that we could be 'perpetuating the delay for significant reductions' forever (Stern 2007). The demands of intergenerational justice simply prescribe more action today (cf. Jonas 1979/1984; Tremmel 2006).

The need to weigh the sustainability dimensions as well as the need to deal with risks and uncertainties with respect to future generations indicate that defining sustainable innovations requires making normative choices, and should therefore be a political process rather than a purely scientific one. Therefore, an essential and valuable aspect of sustainable innovation is to make normative choices, trade-offs and risks and uncertainties explicit. This means that democratic policy making is essential in setting the direction and pace of development. As such policy-makers have the responsibility to lead the other stakeholders in making normative choices. Scientists do play an important role in policy making by creating knowledge and gathering evidence on sustainable innovation. With respect to sustainability scientists will provide valuable insights into impacts of the innovation on the different dimensions of sustainability and trade-offs that may occur, as well as on the risks and uncertainties. Other stakeholders, such as investors, farmers and non-governmental organisations should have access to knowledge and insights generated and participate in politics and decision making to bring in their opinion and defend their interests. It is therefore vital that responsibilities are clearly defined and demarcated in order to ensure that each stakeholder has the opportunity to play its required role in the decision making process, and avoid situations where relevant responsibilities are not taken up (van de Poel et al. 2012) or where decision making becomes characterised by 'organised irresponsibility' (Beck 2000).

Step 2: **Reflection on ethical and moral responsibility**

How should responsibilities be taken up in the global biofuel chain from an ethical point of view? This section addresses this question by first untangling the meanings that 'moral responsibility' can have, following the taxonomy of Vincent (2011). Note that this taxonomy is designed for notions of responsibility that apply to individual human beings: we assume that its core structure can be applied to organisations and institutions as well (cf. Goodpaster 1983). Second, we use this taxonomy to identify the conditions under which policy makers can 'take responsibility' for setting a policy to stimulate responsible biofuel innovation in the absence of global governance structures. We look specifically at European and Tanzanian policy makers. We argue that stakeholder involvement in policy-setting is an essential part of morally responsible innovation. This includes ensuring that different stakeholders can, in turn, 'take responsibility' for playing their role in the policy-making process. Note that, our ethical framework should not be considered the final word on how responsible innovation should proceed in general. Rather, it indicates what practical conditions for policy making can be derived from current work on moral responsibility and discourse ethics, and what recommendations this normative work can offer us for current biofuel policy making.

The notion 'moral responsibility' can be used in several different ways. In this chapter, we will focus on the forward-looking role responsibilities of different stakeholders (Pols 2010; Vincent 2011). When it comes to responsible innovation, the question becomes what role responsibilities different stakeholders should take on in order to make the innovation work and avoid unwanted effects. In the case of biofuels, 'work' includes making sure that biofuel innovation is sustainable, where defining its sustainability is, as we have noted in the previous section, the result of a political rather than a purely scientific process.

In order to successfully take on such a role responsibility, however, certain capacities are needed in order to ensure that the set goal can actually be achieved. For this, we reach back to Aristotle (1985 translation), who has proposed two threats to moral responsibility: ignorance and force. That is, you cannot be held responsible for consequences of your actions that you could not reasonably have anticipated, or that you were unable to prevent. Of course, you can be held responsible for failing to anticipate or to take adequate precautions. This implies that taking the role of responsibility for achieving an end requires that you have the (physical, mental, institutional, etc.) capacities to achieve that end and know how to achieve that end—or at least that you have the capacities to get to that point. Furthermore, avoiding undesirable side effects as much as possible—what we consider a key component of responsible innovation—also requires the capacities for responsiveness; to monitor the effects of one's actions and intervene if necessary.

Also note that especially the early stages of innovation are characterised by learning and experimentation, in which actors have limited knowledge of the impacts while risks of failure are relatively high, furthermore involved stakeholders may act opportunistically. Therefore, innovating responsibly requires intensive stakeholder participation and may require special capacities on the part of the innovator. The four dimensions of responsible innovation defined by Stilgoe et al. (2013) (see Textbox 15.1) indicate capacities that they judge to be required for responsible innovation.

As described in the previous paragraph on sustainability, policy-makers play a special role in responsible innovation by setting the goals for development, a decision making process in which stakeholder participation is essential. Not only is stakeholder involvement necessary for discharging the other role responsibilities of monitoring, controlling, etc., stakeholder involvement in decision making offers various practical advantages, though it introduces risks as well: see Table 15.1 (Luyet et al. 2012). Full stakeholder involvement offers three advantages from an ethical perspective over no involvement or informed consent. First, different stakeholders tend to bring different values and goals to the table, which enables a more *comprehensive and inclusive* ethical evaluation. It also increases the chance that no important values are overlooked in the decision making process. Second, discourse ethicists (Habermas 1990, 1991; Apel and Kettner 1992) have argued that discussing norms and arguments in an open, democratic process is a precondition for *just or fair* policy-setting. According to them, a policy cannot be just if it is not the result of a democratic process where all relevant perspectives are heard, and no force is exerted except that of the better argument. van Buren (2001) has used a similar argument in the context of business ethics; he claims that 'mere' informed consent in stakeholder relations cannot guarantee a fair treatment of stakeholders who have urgent and legitimate moral claims but lack the power to walk out of an unfair arrangement (e.g. child labourers or landless farmers). Third, it can *contribute to the legitimacy* of

Table 15.1 Advantages and risks of participation (Luyet et al. 2012: p. 214; Chambers 2009)

Advantages
• building trust and a feeling of ownership
• improving project design with local knowledge
• adaptation of project design to local circumstances and priorities
• integration of various interests and opinions
• better understanding issues and project
• public acceptance of decisions
• fostering and developing local learning
Risks
• time consuming and expensive
• potential stakeholder frustration
• possibly identification of new conflicts
• difficult to involve good representatives for all stakeholders groups
• empowerment of already important stakeholder
• potential increase of vulnerability through sharing information

a policy that has effects beyond state boundaries (Partzsch 2011). In this case, a new source of legitimacy is needed that involves approval from the global community, and stakeholder involvement can play a role in this legitimisation.

(1) **Anticipation** - systematic thinking aimed at increasing resilience, while revealing new opportunities for innovation and the shaping of agenda's for social robust risk research

(2) **Reflexivity** - (self)-referential critique as an organising principle;

(3) **Inclusion** - including new voices in the governance of science and innovation as part of the search for legitimacy;

(4) **Responsiveness** - changing shape or direction in response to stakeholder and public values and changing circumstances.

Textbox 15.1 The four dimensions of the responsible innovation framework by Stilgoe et al. (2013)

Stakeholder involvement in decision making can offer those three ethical advantages, but in order to do so, as discourse ethicists have claimed, certain procedural values such as inclusivity and fairness have to be upheld: stakeholder involvement in itself is not necessarily morally good (Greenwood 2007). The responsible innovation community fortunately is quite aware of this: Koops (Sect. 1.3, this volume) notes the prevalence of discussions on procedural values in this book.

Upholding these values can be framed in terms of responsibilities as well. They require the policy maker to take on more role responsibilities, namely those with regard to enabling a proper stakeholder discussion. Those include making sure that actual stakeholders are identified as such and are invited into the discussion, stakeholders representing all relevant views can participate in the discussion, and that power differences between stakeholders are compensated for as much as possible, in order to create a level playing field. Similarly, the participating stakeholders have to take up role responsibilities as participants in the discussion. Those include giving all values and arguments due consideration, and not suppressing or excluding any relevant argument. Again, this requires that stakeholders have (or are given by the policy maker) certain capacities, namely, the capacities required to participate in a discussion based on rational arguments. This may be especially challenging for broader definitions of 'stakeholders' that include future generations and/or ecosystems, threatened species and other nonhuman actors. In practice, nonhuman actors are often indirectly represented, by NGOs such as the WWF, as direct representation would be impossible. Representation of future generations has its own difficulties (e.g. Ekeli 2005; Norton 2005, Chap. 8).

Upholding the values associated with responsible stakeholder involvement in decision making is practically very difficult when stakeholders are spread across the globe and range from political institutions to multinational companies to small farmers. The greater and wider the impact of a certain policy is, the more and more different parties will become stakeholders, and the harder and more expensive it will become to take up one's role responsibilities as a policy-maker. This is not to say that any policy that does not live up to these responsibilities will be flat-out irresponsible. Rather, these responsibilities can, as Koops (Sect. 1.1, this volume) puts it, serve as both an ideal and a project. An ideal, because it sets a standard by which to measure current policy-making procedures in the global biofuel trade. A project, because it suggests options for improvement of those procedures—and because the responsibility framework itself is continually being investigated and improved as well.

Step 3: **Reflection on jatropha biofuel experiences in Tanzania**

In this section we reflect on the rise and fall of the jatropha biofuel cultivation (also illustrated by Fig. 15.2), and describe the role of the selected stakeholder groups. We have selected the following stakeholder groups; policy makers (European and Tanzanian), scientists, international investors, NGOs, and farmers. For these selected stakeholder groups we describe what stakes they have, what priorities they envision with respect to sustainable development, what capacities they have (see also Kroesen et al., Chap. 11, this volume), and which responsibilities they have taken up.

The driving force for biofuel innovations was the **European Union (EU)**, through enacting the Directive for the promotion of biofuels (2003/30/EC), containing a voluntary target of 5.75 % share of renewable energy in the transport sector by 2010. In 2009, the EU agreed on the Renewable Energy Directive (RED) (2009/28/EC) with the ambitious binding target that 10 % of the final consumption

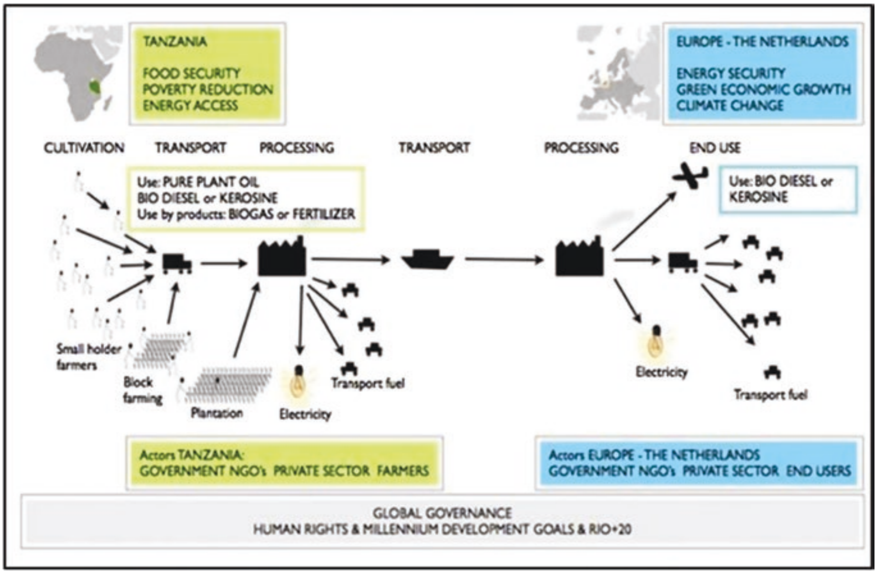

Fig. 15.2 Overview of different business models for jatropha biofuels cultivated in Tanzania for local use or for export to Europe

of energy in transport in the EU should come from renewable sources by 2020. This directive also establishes sustainability criteria for biofuels (EASAC 2012: p. 5). In response to the emerging negative consequences of this policy (e.g. rising food prices and land grabbing), the EU took responsibility and published in October 2012 proposals to limit the range of biofuels that can be counted against the 2020 10 % target. The aim of the new proposals is to limit the proportion of food-based (primary-crop) biofuels that can contribute towards the target to a maximum of 5 %. The remainder of the target, it is expected, would be met by fuels derived from waste or other renewable sources (EASAC 2012: p. 6). In January 2013 the EU published the Clean Power for Transport: A European Alternative Fuels Strategy, which encompasses biofuels as well as gas, electricity and hydrogen (IP/13/40). The strategy document advocates support for sustainable advanced biofuels produced from lignocellulose and wastes, as well as algae and microorganisms. It recommends no further public support for first generation biofuels produced from food crops after 2020.

What is at stake? The current annual EU demand for road transport energy is about 300 Million tonnes of oil equivalent (Mtoe). Thus, to achieve the 10 % aims of the Renewable Energy Directive for 2020, about 30 Mtoe of fuel from renewable sources will be required annually by 2020 (EASAC 2012: p. 9). At present, biofuels contribute about 10 Mtoe to the EU road transport energy mix. In 2008 about 40 % of this was imported into the EU, either as biofuel or feedstock for manufacture in the EU, mostly from the USA and Brazil (EASAC 2012: p. 9). Land use for current levels of EU biofuel demand is estimated at 7 million hectares

(Mha), of which 3.6 Mha is within the EU (EASAC 2012). Because of this occupation of arable land biofuels are considered to be a major driver of food prices (Mitchell 2008), and thus, a major threat to the human right to food.

Furthermore, current first-generation biofuels appear to provide little or no greenhouse gas emission reduction once all impacts of biomass cultivation, including indirect land use change (ILUC), and fuel production are taken into account. Despite the development of criteria for protection of biodiversity and ecosystem services and certification schemes, there are serious concerns about the sustainability of biofuel production and its impacts on the natural environment (EASAC 2012: p. 25).

With regard to stakeholder involvement in policy-making, the EU can be said to have neglected two major responsibilities that could have mitigated or prevented this outcome. The first is the focus on procedures rather than goals, the second is its biased stakeholder selection.

On procedures and goals, according to Norton (2005), stakeholder policy deliberation should alternate between an action phase, in which criteria and indicators are chosen and action is undertaken, and a reflective phase, in which set goals and wishes are evaluated. This is meant to ensure that the ends of the initiating stakeholder (in our case the EU) serve as input rather than as constraint on the discussion. In the global biofuel chain, however, discussion has often been about the means rather than also about the ends. For example, though the EU RED's 10 % target was challenged, this challenge became overshadowed by a discussion on the means, particularly whether ILUC effects should be accounted for (Levidow 2013). Similarly, the Dutch Cramer Criteria (Cramer 2007) focused on the means, conditions for the sustainable production of biomass, but assumed the ends, having large quantities of biomass available for energy production, fixed from the start. Serious, multi-stakeholder discussions on the EU's goals could have contributed to priority-setting and goal re-evaluation.

With regard to selective stakeholder involvement, Partzsch (2011) has examined the procedures by which the Dutch Cramer Criteria (Cramer 2007) and the Roundtable on Sustainable Palm Oil (RSPO) sustainability criteria have been drawn up. She concludes that neither procedure has adequately involved stakeholders from the Global South. Similarly, Romijn et al. (2013) observe that the procedure to establish the Dutch NTA 8080/81 biofuel sustainability norm did not involve African small farmers as stakeholders. Of all (living, human) stakeholders, African small farmers might be the most difficult to involve in a meaningful way in stakeholder discussions set up by Dutch or European policy makers. Their interests are similarly easily overlooked due to their relative powerlessness, compared to other stakeholders in the global biofuel trade. Ethically seen, however, this powerlessness combined with the legitimacy and urgency of their claims makes them the most important stakeholders to consider (van Buren 2001). On a very practical level, small farmers are responsible for a large part of global agriculture (Vorley et al. 2012) and for Tanzanian agriculture in particular (Sulle and Nelson 2009). Seeing as how biofuel production is essentially an agricultural activity, not involving small farmers as stakeholders in a biofuel policy-making process can lead to missed opportunities, especially regarding how biofuels could be used for poverty alleviation. Similarly, not involving small farmers as stakeholders in the discussion can lead to practical problems in implementing certification systems or sustainability criteria in local practices (Romijn et al. 2013).

What are priorities? The EU prioritised the 'greening' (read 'sustaining') of its transport sector and the interests of the fossil fuel establishment above food security in the global south. This is evident from the fact that the EU chose to direct its policies towards large-scale biofuel investments rather than limiting transport or shifting to transport options incompatible with the continued usage of fossil energy sources.

What capacities are in place? The EU lacks the land area to sustain the current EU lifestyle with food and energy production within her borders, creating dependency on land elsewhere in the world. Compared to other stakeholders in the global biofuel chain, the EU especially has the capacity in terms of knowledge and economic power to take role responsibility for sustainable development. She can create insight into the impacts on the different dimensions of sustainability as result of the policies directed, and take the responsibility to prevent damaging impacts. On paper the EU took this responsibility by formulating the Policy Coherence for Development (PCD) approach in 2005, requiring that EU policies have a positive impact on vulnerable groups in order to accelerate progress on the Millennium Development Goals. Nevertheless, the EU approach remained technocratic with a focus on GHG-balances and crop yields rather than opting for democratic decision making and small-scale decentralized production for local energy demand (Leopold and Diets 2012: p. 5). Furthermore, the EU itself wrote a reflection on the production of biofuels in developing countries from the point of view of PDC (Diop et al. 2013). This evaluation acknowledges problems such as rising food prices and large land acquisitions, although it is stresses that there are other causes besides the EU biofuel policies. A recommendation to the European Commission is to expand the sustainability criteria to include social criteria, food security, and access to natural resources for communities affected by land acquisitions.

With respect with the four dimensions of responsible innovation defined by Stilgoe et al. (2013, see Textbox 15.1) and our identified capacities for responsibility, the EU failed to anticipate on the impacts of her biofuel policy, and has not included the voices of the global south in her policy. She did, however, reflect on the results of her policies and will hopefully be responsive.

Which responsibilities have been taken so far? One may argue that the enacting of the Biofuel Directive (2003/30/EC) and later on the RED (2009) shows that the EU is willing to act responsibly for reducing the emissions of transport in line with its aim of climate change mitigation. But how sustainable is it to promote biofuels? Why not simply tax fossil fuels or transport? How realistic is it to try and handle the resulting problems such as food insecurity and sustained poverty through the definition of sustainability criteria? Non-food crops do still occupy land area! The EU should have realized that the huge scale and relatively short time frame laid down in its directives, combined with the lack of land area within EU borders as well as the economic power of EU would lead to an unsustainable biofuel innovation trajectory. In general, monitoring and control become more difficult the greater the scale of the innovation and the more drastic the changes are. In the case of biofuel innovations, the EU could be said to "have bitten off more than it could chew"; it has tried to take role responsibility for an innovation trajectory that it (apparently) did not have the capacities to properly and promptly monitor and control.

Until recently, the **Tanzanian government** welcomed new investors for biofuel cultivation, both using the outgrower and plantation model, but since many biofuel projects in Tanzania failed, the government halted new investments and is in the process of defining a biofuel policy.

What is at stake? And *what are priorities?* These are expressed in a vision defined in the draft Tanzanian liquid biofuel policy (MEM 2012: p. 10; van Teeffelen 2013: p. 56): to contribute significantly to local energy security through sustainable production and utilisation of biofuels for socio-economic development in Tanzania.

What capacities are in place? The Tanzanian government may have less economic power and less access to knowledge than the EU, but the land area and climate are favourable for biofuel cultivation.

What responsibilities have been taken up so far? By halting investment and drafting a biofuel policy the Tanzanian government has shown to be willing to take responsibility for sustainable biofuel development. As defined in the draft policy, five specific objectives are; (1) facilitate feedstock production in mapped and zoned areas; (2) promote and facilitate value addition; (3) strengthen institutional, legal and regulatory frameworks, (4) ensure environmental sustainability; and (5) ensure equitable benefit sharing (MEM 2012: p. 12 in line with FAO/UNEP 2010: pp. 7–8).

The main policy instrument to be implemented is zoning and mapping of biofuels and restrictions on land tenure [max. 33 years with a 5-year probation and a maximum of 20.000 ha and for foreign investors only through derivative rights from the Tanzania Investment Centre (TIC) (MEM 2012: pp. 14, 15)]. The draft policy acknowledges that edible feedstock for biofuels is a risk to food security, as well as that there is limited market access when choosing a non-edible biofuel crop (MEM 2012: p. 13). Medium- and large-scale biofuels investments have to contribute to increased food production (p. 14). The government will formulate regulatory frameworks and will ensure effective enforcement, strengthen coordination, facilitate stakeholder involvement, stimulate the private sector to invest in biofuels and development through corporate social responsibility (investing in infrastructure, health and gender equality, p. 22+25), stimulate local communities to get involved, and establish a technical body responsible for advice on liquid biofuels investment and development. An interesting part of the draft policy document is the chapter that defines the roles of stakeholders, in which communities are named to have a role in environmental protection, NGOs and CSOs are assigned the task of defending interests of vulnerable and disadvantaged groups, and the media is assigned the role of watchdog (6.0 MEM 2012: pp. 26–27).

On a more critical note, Vorley et al. (2012) observe the overall lack of inputs in forms of ideas of small farmers in general, and women in particular, in the process of formulation, implementation, monitoring and evaluation of policies in Tanzania. The processes are not participatory enough to capture the views, outcries, concerns, frustrations and burning issues of small farmers in general, and women in particular. These are processes that typically take place in urban ministries' offices, and hotels in male-dominated decision making environments (Vorley et al. 2012: p. 26). This is reflected by the fact that the draft Tanzanian liquid biofuel policy is still very open to investors and not aiming to solve the problems occurring due the vulnerable position many farmers are in. The concentration of land in the hands of very few landowners (land grabs) has driven small farmers from their land, undermining the basis of their existence (Ottinger 2007 in León-Moreta 2011: p. 112). Therefore it is important both to legally recognise the right of indigenous people to land ownership and to protect farmers against the urgings by domestic or transnational companies to sell their land at low prices (León-Moreta (2011): p. 113). Recognition of land rights and protection of vulnerable groups against powerful stakeholders such as large investors deserves extra attention in defining this policy. As with the EU, the goal of defining a sustainable biofuel policy gives the Tanzanian government the responsibility to work actively to involve stakeholders in the policy making process, in order to make it more just and fair.

Scientists could have played a more critical role in both the biofuel policy definition as well as in the implementation of jatropha biofuel production by indicating the potential impacts and risks more clearly. Scientists have been opportunistic in publishing and quoting high yields, though they have also indicated that this would be hard to achieve under field conditions. In our literature review we found seed yields from 1 up to 6 tons per hectare per year (t/ha/y); however many publications lacked reflection on deviation from what other scientists had found, while this is one of the most decisive variables (van Eijck et al. 2010: pp. 74–75).

What is at stake? The stakes for this actor group seem to be low, however, scientists have to be careful to keep their impartiality and their reputation (and perhaps their research funds). Scientists play an important role as honest knowledge brokers (Pielke 2007), which means that they should take the responsibility to critically create and apply knowledge on sustainable biofuel innovations.

What are priorities? Priorities are not strongly defined for this actor. Research funding may play a role. Based on the definition of sustainability, one may claim that multidisciplinary research is needed for sustainability as well as poverty related field research (such as barefoot economics; Neef 2010).

What are the capacities? Scientists do have the capacity to understand the sustainability issues, although this requires multidisciplinary research. Scientists have access to information and the capacity to create insight into the trade-offs in sustainability dimensions, can map the stakes and priorities of stakeholder groups and comprehend long time scales. However, scientists should recognise that in sustainability some decisions are based on moral choices rather than on scientific facts. To quote Giampietro and Mayumi (2009: pp. 256–258), no matter how smart the scientists making the predictions are, our future is necessarily uncertain. When deciding about a desirable common future, decisions should be based on what civil society wants, not on what the academic establishment thinks should be done, as we have argued in the section on sustainability. Similarly, Norton (2005, Chap. 7) argues that scientists bring valuable expertise to the negotiating table, but that there is no reason to place their values above those of other stakeholders. For Norton, valuing and weighing evidence are democratic rather than technocratic processes.

What responsibilities have been taken up so far? Scientists did gather evidence on sustainable biofuel innovations and brought up critical issues with respect to social and environmental dimensions of sustainable development. However, scientists should take care in their publications to present assumptions regarding quantitative information such as yields and modelling outcomes together with the assumptions, uncertainties and reflection on conflicting outcomes of other research projects, as these can be easily misinterpreted or misrepresented. White and Dasgupta (2010: p. 597), for instance, rightfully conclude that the claim that biofuels have the potential to revive peasant agriculture and stimulate rural development, which has been made time and again in popular as well as academic writing, should be examined against the backdrop of persistent agrarian underdevelopment (or uneven development). Also, it should be kept in mind that current policies favour certain business models over others, such as the requirement for sustainability certification that due to its administrative and financial burden is easier to implement for large-scale plantations than for small farmers. Both scientists and policy-makers as stakeholders should take the responsibility to draw attention to uncertainties and controversies in the scientific debate when engaging in the decision making process. This will facilitate creating an evidence-based policy rather than leading to policy-based evidence gathering, of which current EU biofuel targets have been accused (Sharman and Holmes 2010).

The huge demand created by the European biofuel directive (2003/30/EC) and subsidies made **international investors** enthusiastic for large-scale biofuel projects.

What is at stake? And what are priorities? Most investors did define social and environmental business objectives, aiming at creating jobs, generating extra income for farmers, and CO_2 emission reduction, besides profit making. However, none of the large-scale biofuel projects in Tanzania managed to significantly contribute to poverty reduction (van Teeffelen 2013: p. 57). For some projects the CO_2 emission reduction is doubtful due to large initial carbon debts created through the extensive land clearing required to start the plantations, and none of the large-scale biofuel projects managed to generate a profit. Most activities are still heavily depending on subsidies and so far many initiatives have failed (FAO 2010).

What capacities are in place? The investors had access to investment capital and knowledge, and in several cases they were able to lease extensive tracts of land in Tanzania for as long as 99 years at bargain prices.

What responsibilities have been taken up so far? Some companies did invest in extension work for the small-scale farmers, and were able to create some extra income for these farmers. However, in other cases, the risk of the innovation and the cost of failure have harmed the small-scale farmers and the local environment (Hooijkaas 2010; Bergius 2012). In responsible innovation the learning trajectory of innovation should be made explicit, by acknowledging that the first projects are experiments and that the risk of failure is high. Anticipating on failures should include ensuring that especially vulnerable groups are protected against the possible consequences of failure. The investors could have done this by more carefully defining their social and environmental objectives, involving those groups more in decision making and defining an exit strategy so that, in a worst case scenario, at least the most vulnerable stakeholders would not be harmed, that is, be left no worse off than they were before. They should also have allocated funds towards these objectives to avoid harming vulnerable groups or damaging the environment in the initial learning phase of the biofuel innovation.

Non-Governmental Organisations (NGOs), both local and international, played a critical role in the biofuel developments in Tanzania.

What is at stake? As biofuels pose a threat to food security and income of poor farmers, and endanger ecosystems, NGOs have a crucial role in getting these stakes on the agenda of policymakers.

What are their priorities? NGOs typically have priorities deviating from policy makers, putting the emphasis on livelihoods of the poor (people dimension) and/or the environment (planet dimension), while policy makers are more engaged with balancing over all three, profit, people, and planet, dimensions.

What capacities are in place? The strength of NGOs is that they have a good connection with their target group and a large input in the public debate.

What responsibilities have been taken up so far? NGOs took their responsibility in critically reflecting on the policies and activities in the biofuel sector, through publications and campaigns. To name some: ActionAid, Fair Fuels, Fair Politics, Friends of the Earth, Haki Ardhi, Oxfam, Max Havelaar, TaTEDO, WWF. Local NGOs played a crucial role in halting large investments in biofuels in Tanzania and played an important role in policy definition as well. Their views and activities get considerable press coverage. Some local NGOs have invested in inclusive biofuel projects as well.

Although living in marginal conditions **small farmers** are willing to invest land area and labour in experimental biofuel production for extra income generation.

What is at stake? While in the perspective of the global biofuel chain their stakes are relatively low, the stakes for the individual farmers are high relative to their income and food insecurity. Despite the attention in biofuel policies given to commercial investment, the bulk of investment in agricultural production is still made by small farmers themselves (Vorley et al. 2012), who invest their own savings as well as small loans. Note that a high number of outgrowers is involved in the outgrower model studied by Balkema and Romijn (2012). In scaling up for export the company estimated that it would have to reach out to over 50.000 farmers by 2013, up from 5.000 contracted farmers in 2010 (collectively owning approximately 3.500 ha of jatropha hedges) (Balkema and Romijn 2012: p. 9).

What are priorities? Priorities for small farmers are food security and extra income generation, and in the long run sustaining, or even better, improving soil fertility and water conservation to increase yields. In a baseline study (Heijnen 2010; Balkema et al. 2010), small-scale farmers indicated that they do experience hunger for a few months per year (93 %, Heijnen 2010, Fig. 51) and that only about half of them managed to eat three meals a day on regular basis (48 %, Heijnen 2010, Fig. 52). The farmers indicate that recent yields of food crops have been unsatisfactory and the number of cattle has gone down due to severe

Table 15.2 Generalized characteristics of selected stakeholders groups with respect to responsibilities for sustainable biofuel innovation

	Policy makers	Scientist	Investor	NGO's	Farmers
Interest	Setting agenda for development	Generating knowledge and insights	Income generating activities for profit making	Protecting vulnerable groups or the environment	Lowering risks
Stakes	Trust	Trust	Investment	Trust	Food security and income
Resources	Social network and budget	Knowledge, social network, and budget	Capital, social network, knowledge	Social network of followers and budget	Land and labour
Strengths	To define policies, has access to capital	Knowledge creation and gathering evidence	Generating economic activity and the access to capital	Contact with societies and knowledge of environment	Knowledge of local society and ecosystem
Vulnerabilities	Reputation, wants to be re-elected	Reputation	Need stable investment climate	Reputation and budget	May have an insecure basis of existence
Sustainability priorities					
Dimensions	Profit + People + to lesser extent Planet?	Profit + People + Planet	Profit	People + Planet	People
Spatial overview	National or international	National or international	Local, national or international	Local, national or international	Local
Timescale overview	Medium term	Long term	Short term	Medium term	Short term
Role in responsible innovation					
Setting direction and pace	Leading	Providing knowledge and insights	Lobbying and investing	Lobbying	Participating
Capacities in place	Economic power and knowledge	Knowledge, and insights	Investments	Social network	Agricultural knowledge
Responsibilities taken	Defined and adapted policies	Created knowledge and gathered data	Generated economic activity	Critical reflection on policies	Participating in experiments

drought (Heijnen 2010: pp. 76–81). On the question whether they would be interested to plant jatropha on their own land, the large majority of the respondents (88 %) answered to be interested to engage in jatropha cultivation. The main reason to consider jatropha cultivation is additional income (62 %) (Heijnen 2010, Appendices p. 73). Besides advantages, the respondents were also asked about possible disadvantages of jatropha cultivation. The farmers indicated the following: uncertain markets (28 %), poor prices (13 %), loss of pasture land (7 %) and food insecurity (5 %). However, most of them were not aware of disadvantages (58 %) (Heijnen 2010, Appendices p. 74). As reflected by the quotes, the expectations of the smallholders vary, but may be unrealistic.

What capacities are in place? The capacities of small farmers are mainly in agricultural cultivation. Income as well as access to knowledge is generally low. However, farmers tend to have indigenous knowledge on local ecosystems, for instance with respect to soil fertility, biodiversity and perhaps local climate.

What responsibilities have been taken up so far? The farmer takes responsibility for the cultivation of jatropha biofuels, sustaining their families and often the local ecosystem.

> *I don't know how many kilo's per tree I can harvest, but from next year I will keep an eye on it. For now, I'm already happy that there are seeds growing on my trees, although they are not numerous.* ***Mr. Zebazeba, outgrower Orkesumet***

> *I have already divided my farm amongst my wives. I will instruct them how to grow and harvest Jatropha, and convince them to build a house from the money from harvest they will have so I can live in one of the houses.* ***Lucas Ngukuu Karomo, Orkirun'urung***

Textbox 15.2 Quotes noted down during interviews with outgrowers in (Heijnen 2010: p. 85)

Table 15.2 summarises the generalised characteristics of selected stakeholder groups. The table is based on the concepts of sustainability and moral responsibility and on the reflection on the jatropha biofuel innovation trajectory experiences in Tanzania.

15.3 Conclusions

Essential in defining sustainability is that the impacts on its different dimensions (social-cultural, environmental and economic) are balanced according to the stakes and priorities of the different stakeholders. Our observation is that, despite the increasing knowledge on impacts and the formulation of criteria for sustainable biofuels, it is still difficult to implement sustainable biofuel innovations on a large scale. In order to be just and sustainable, the decision making process should incorporate all stakes and priorities, rather than simply being a discussion on means, where the ends have been pre-determined by the most powerful stakeholder.

In the case of current biofuel innovations, the unprecedented large scale and high pace of the introduction of biofuels, as well as the power differences between the stakeholders and the lack of stakeholder participation and clearly defined responsibilities, have resulted in irresponsible non-sustainable innovation.

Based on our conceptualisation of sustainability, our ethical framework on moral responsibility, and our reflection on the rise and fall of the jatropha biofuel innovation in Tanzania, we have explored the responsibilities that could be taken up by the stakeholders involved to come to a more sustainable biofuel innovation process.

With respect to the EU we conclude that the EU is exporting her problem of high CO_2 transport emissions to the global south in the form of marginalisation of small-scale farmers and depletion of soil fertility. This means that the preferred option of the most powerful stakeholder in the chain is being implemented. The EU could take responsibility for truly sustainable innovation here by questioning its own ends and reducing or taxing emissions, rather than enforcing its ends on stakeholders worldwide. Scientists could take responsibility by recognising that innovating sustainably requires a multidisciplinary approach and democratically made moral choices as well as scientific research. The Tanzanian government takes responsibility with its draft policy, that should focus especially on stakeholder involvement and protection of vulnerable parties. Investors, finally, could take responsibility by emphasising the experimental character of biofuel innovations and taking precautions to protect vulnerable parties against the costs of failure. Further research based on interviews of stakeholders in the Netherlands and Tanzania has to be undertaken to give a more thorough stakeholder analysis and a more detailed description of responsibilities and possibilities for responsible stakeholder participation.

Acknowledgments We want to thank everybody who made this research possible. During the field work this included several students, colleagues, and our interviewees. During writing we benefited from comments by our colleagues. This article was produced as part of the project "Biofuels: Sustainable Innovation or Gold Rush?", funded by the Responsible Innovation Programme (MVI) of the Netherlands Organisation of Scientific Research (NWO).

References

Apel, K.O., and M. Kettner. 1992. *Zur Anwendung der Diskursethik in Politik, Recht und Wissenschaft*. Frankfurt am Main: Suhrkamp.

Aristotle. 1985. *Nicomachean ethics*. Trans. Terence Irwin. Indianapolis: Hackett.

Balkema, A.J., S. van Dijk, S. Heijnen, G.P.J. Verbong, H.A. Romijn, and E. Huntjens. 2010. 'An impact assessment methodology for small scale renewable energy projects in developing countries funded under Dutch policies defined to contribute to the millennium development goals'. *Proceedings of the Berlin conference on the human dimensions of global environmental change*. 8–9 October 2010, 1-11. Berlin: Environmental Policy Research Centre (FFU).

Balkema, A.J., and H.A. Romijn. 2012. Innovative business models for sustainable biofuel production: The case of Tanzanian smallholder jatropha farmers in the global biofuel chain. In *The business of social and environmental innovation. New Frontiers in Africa*, eds. Hamann, R., M. Hall, E. Wosu Griffin-EL, and V. Bitzer, 75–100. London: Springer, and Cape Town: UCT Press.

Beck, U. 2000. Risk society revisited: Theory, politics and research programmes. In *The risk society and beyond: Critical issues for social theory*, ed. B. Adam, U. Beck, and J. Van Loon, 211–230. London: Sage.

Bergius, M. 2012. *Large scale agro investments for biofuel production in Tanzania—impact on rural households*. Norway: University of Agder.

Beyene A., C.G. Mungongo, A. Atteridge, and R.K. Larsen. 2013. *Biofuel production and its impact on local livelihoods in Tanzania, a mapping of stakeholder concerns and some implications for governance.* Stockholm Environmental Institute, working paper 2013-03.

Chambers, R. 2009. Dr. Robert chambers elaborates on good practices for participatory GIS. http://www.youtube.com/watch?v=awL1Hpy7spA (video).

Cramer, J. 2007. Testing framework for sustainable biomass. Available at http://www.sustainable-biomass.org/dynamics/modules/SFIL0100/view.php?fil_Id=857. Accessed 9 April 2015.

Diop, D., M. Blanco, A. Flammini, M. Schlaffer, M.A. Kropiwnicka, and M.M. Mankhof. 2013. *Assessing the impact of biofuels production on developing countries from the point of view of policy coherence for development.* European Commission, Final Report February 2013.

EASAC. 2012. The current status of biofuels in the European Union, their environmental impacts and future prospects, European Academia Science Advisory Council EASAC policy report 19 December 2012 ISBN: 978-3-8047-3118-9. This report can be found at www.easac.eu.

Ekeli, K.S. 2005. Giving a voice to posterity—deliberative democracy and representation of future people. *Journal of Agricultural and Environmental Ethics* 18: 429–450.

FAO. (2010). Jatropha: A smallholder bioenergy crop, the potential for pro-poor development. *Integrated Crop Management* 8.

FAO/UNEP. 2010. *Decision support tool for sustainable bioenergy*, UN Energy report by UN Food and Agricultural Organisation and the UN Environmental Programme, ISBN: 978-92-5-106638-6 (see box 2 for opportunities and impacts of biofuel projects on global, national and local scale).

Giampietro, M., and K. Mayumi. 2009. *The biofuel delusion, the fallacy of large scale agro-fuel production*. UK and USA: Earthscan.

Goodpaster, K.E. 1983. The concept of corporate responsibility. *Journal of Business Ethics* 2(1): 1–22.

Greenwood, M. 2007. Stakeholder engagement: Beyond the myth of corporate responsibility. *Journal of Business Ethics* 74: 315–327.

Habermas, J. 1990. *Moral consciousness and communicative action*. Cambridge, MA: MIT Press.

Habermas, J. 1991. *Erläuterungen zur Diskursethik*. Frankfurt am Main: Suhrkamp. (English, 1993).

Heijnen. 2010. *The impact of small scale renewable energy projects in least developed countries, a baseline study*. M.Sc. thesis, Eindhoven University of Technology, The Netherlands.

Hermele, K., L. Olsson, and A. Jerneck. 2009. Justice and fairness in resource governance: Conflicting views on allocation and access. *Paper presented to the conference on the human dimensions of global environmental change*. Amsterdam, 2–4 December 2009, Theme 5, Panel 5: Allocation and Access in Earth System Governance.

Hooijkaas, N. 2010. *The impacts of large-scale Jatropha production on the GHGbalance, carbon stock and socio-economics in Kilwa, Tanzania.* M.Sc. thesis, Copernicus Institute, Utrecht University, The Netherlands.

Jackson, T. 2009. *Prosperity without growth; economics for a finite planet*. Book published 16th October 2009 by Routledge.

Jonas, H. 1979/1984. *Das Prinzip Verantwortung*. Frankfurt am Main: Suhrkamp; *The imperative of responsibility: In search of an ethics for the technological age*. Chicago: The University of Chicago Press.

León-Moreta, M. 2011. Biofuels—A threat to environmental and human rights? An analysis of the impact of production of feedstock for agrofuels on the rights to water, land and food, European. *Journal of Legal Studies* 4(1): 102–120.

Leopold, A., and K. Diets. 2012. *Transnational contradictions and effects of Europe's bioenergy policy, evidence from Sub-Saharan Africa*, Fair Fuels? Working paper 4.

Levidow, L. 2013. EU criteria for sustainable biofuels: Accounting for carbon, depoliticising plunder. *Geoforum* 44: 211–223.

Locke, A., and G. Henley. 2013. *Scoping report on biofuels projects in five developing countries*. London: UKaid and Overseas Development Institute.

Luyet, V., R. Schlaepfer, M.B. Parlange, and A. Buttler. 2012. Review; A framework to implement stakeholder participation in environmental projects. *Journal of Environmental Management* 111(2012): 213–219.
MEM (Ministry of Energy and Minerals). 2012. *First draft; National liquid biofuel policy*, August 2012. The United Republic of Tanzania.
Mitcham, C. 1995. The concept of sustainable development: Its origins and ambivalence. *Technology in Society* 17(3): 311–326.
Mitchell, D. 2008. *A note on rising food prices.* World Bank Policy Research, Working paper 4682.
Neef, M.M. 2010. *Outside looking in: Experiences in barefoot economics*. Book and video (http://www.democracynow.org/2010/11/26/chilean_economist_manfred_max_neef_on).
Nordhaus, W. 2008. *A question of balance, weighing the options on global warming policies*. Yale: Yale University Press. ISBN 9780300137484.
Norton, B.G. 2005. *Sustainability: A philosophy of adaptive ecosystem management*. Chicago: University of Chicago Press.
Ottinger, R.L. 2007. Biofuels—potential, problems and solutions. http://cmsdata.iucn.org/downloads/cel10_ottinger_2.pdf. Accessed 20 May 2011.
Partzsch, L. 2011. The legitimacy of biofuel certification. *Agriculture and Human Values* 28: 413–425.
Pielke Jr, R.A. 2007. *The honest broker. Making sense of science in policy and politics*. Cambridge: Cambridge University Press.
Pols, A.J.K. 2010. Transferring responsibility through use plans. In *Philosophy and engineering: an emerging agenda*, ed. I. van de Poel, and D. Goldberg, 189–203. Dordrecht: Springer.
Romijn, H.A., S. Heijnen, and S. Arora. 2013. Standardizing sustainability: Certifying Tanzanian biofuel smallholders in a global supply chain. In *Sustainable value chain management: Analyzing, designing, implementing, and monitoring for social and environmental responsibility*, ed. A. Lindgreen, S. Sen, F. Maon, and J. Vanhamme, 473–487. Gower: Burlington.
Rosling, H. 2006. Hans Rosling shows the best stats you've ever seen (video) at the 2006 TED Conference in Monterey, CA.
Rosling, H. 2007. Hans Rosling's new insights on poverty (video) from TED 2007.
Segerstedt, A., and J. Bobert. 2013. Revising the potential of large-scale Jatropha oil production in Tanzania: An economic land evaluation assessment. *Energy Policy* 57(2013): 491–505.
Sharman, A., and J. Holmes. 2010. Evidence-based policy or policy-based evidence gathering? Biofuels, the EU and the 10 % target. *Environmental Policy and Governance* 20: 309–321.
Stern, N. 2007. *The economics of climate change, the Stern review*. Cambridge: Cambridge University Press.
Stilgoe, J., R. Owen, and P. Macnaghten. 2013. Developing a framework for responsible innovation. *Research Policy* 42: 1568–1580.
Sulle, E., and F. Nelson. 2009. *Biofuels, land access and rural livelihoods in Tanzania*. London: IIED.
Tremmel, J.C. 2006. *Handbook of intergenerational justice*. Cheltenham: Edward Elgar.
UNEP. 2012. *GEO5, global environmental assessment*. UNEP.
van Buren III, H.J. 2001. If fairness is the problem, is consent the solution? Integrating ISCT and stakeholder theory. *Business Ethics Quarterly* 11(3): 481–499.
van de Poel, I., J. Nihlén Fahlquist, N. Doorn, S. Zwart, and L. Royakkers. 2012. The problem of many hands: Climate change as an example. *Science and Engineering Ethics* 18: 49–67.
van Eijck, J.A.J., E.W.M. Smeets, H.A. Romijn, A.J. Balkema, and R.J.J. Jongschaap. 2010. *Jatropha assessment: Agronomy, socio-economic issues, and ecology: Facts from literature*. Utrecht: NL Agency.
van Teeffelen, J. 2013. *Fuelling progress or poverty? The EU and biofuels in Tanzania*. Evert Vermeer Foundation and Fair Politics.

Vincent, N.A. 2011. A structured taxonomy of responsibility concepts. In *Moral responsibility: beyond free will and determinism*, ed. N.A. Vincent, I. van de Poel, and J. van den Hoven, 15–35. Dordrecht: Springer.

Vorley, B., L. Cotula, and M.-K. Chan. 2012. *Tipping the balance: Policies to shape agricultural investments and markets in favour of small-scale farmers*. Oxford: Oxfam GB/IIED.

White, B., and A. Dasgupta. 2010. Agrofuel capitalism: A view from political economy. *Journal of Peasant Studies* 37(4): 539–607.