Philosophy of Engineering and Technology 16

Neelke Doorn
Daan Schuurbiers
Ibo van de Poel
Michael E. Gorman *Editors*

Early Engagement and New Technologies: Opening Up the Laboratory

Springer

Early Engagement and New Technologies: Opening Up the Laboratory

Philosophy of Engineering and Technology

VOLUME 16

Editorial Board

Editor-in-chief

Pieter E. Vermaas, *Delft University of Technology, The Netherlands*
General and overarching topics, design and analytic approaches

Editors

Christelle Didier, *Lille Catholic University, France*
Engineering ethics and science and technology studies
Craig Hanks, *Texas State University, U.S.A.*
Continental approaches, pragmatism, environmental philosophy, biotechnology
Byron Newberry, *Baylor University, U.S.A.*
Philosophy of engineering, engineering ethics and engineering education
Ibo van de Poel, *Delft University of Technology, The Netherlands*
Ethics of technology and engineering ethics

Editorial advisory board

Philip Brey, *Twente University, The Netherlands*
Louis Bucciarelli, *Massachusetts Institute of Technology, U.S.A.*
Michael Davis, *Illinois Institute of Technology, U.S.A.*
Paul Durbin, *University of Delaware, U.S.A.*
Andrew Feenberg, *Simon Fraser University, Canada*
Luciano Floridi, *University of Hertfordshire & University of Oxford, U.K.*
Jun Fudano, *Kanazawa Institute of Technology, Japan*
Sven Ove Hansson, *Royal Institute of Technology, Sweden*
Vincent F. Hendricks, *University of Copenhagen, Denmark & Columbia University, U.S.A.*
Don Ihde, *Stony Brook University, U.S.A.*
Billy V. Koen, *University of Texas, U.S.A.*
Peter Kroes, *Delft University of Technology, The Netherlands*
Sylvain Lavelle, *ICAM-Polytechnicum, France*
Michael Lynch, *Cornell University, U.S.A.*
Anthonie Meijers, *Eindhoven University of Technology, The Netherlands*
Sir Duncan Michael, *Ove Arup Foundation, U.K.*
Carl Mitcham, *Colorado School of Mines, U.S.A.*
Helen Nissenbaum, *New York University, U.S.A.*
Alfred Nordmann, *Technische Universität Darmstadt, Germany*
Joseph Pitt, *Virginia Tech, U.S.A.*
Daniel Sarewitz, *Arizona State University, U.S.A.*
Jon A. Schmidt, *Burns & McDonnell, U.S.A.*
Peter Simons, *Trinity College Dublin, Ireland*
Jeroen van den Hoven, *Delft University of Technology, The Netherlands*
John Weckert, *Charles Sturt University, Australia*

For further volumes:
http://www.springer.com/series/8657

Neelke Doorn • Daan Schuurbiers
Ibo van de Poel • Michael E. Gorman
Editors

Early Engagement and New Technologies: Opening Up the Laboratory

Springer

Editors
Neelke Doorn
Department of Technology, Policy
and Management
Delft University of Technology
Delft, The Netherlands

Ibo van de Poel
Department of Technology, Policy
and Management
Delft University of Technology
Delft, The Netherlands

Daan Schuurbiers
De Proeffabriek (The Pilot Plant)
Delft, The Netherlands

Michael E. Gorman
Department of Science, Technology
and Society
University of Virginia
Charlottesville, VA, USA

ISSN 1879-7202 ISSN 1879-7210 (electronic)
ISBN 978-94-007-7843-6 ISBN 978-94-007-7844-3 (eBook)
DOI 10.1007/978-94-007-7844-3
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013957803

© Springer Science+Business Media Dordrecht 2013
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Acknowledgments

The majority of this volume is based upon work that was originally presented at the workshop "Ethics in the Laboratory: Encouraging Reflection through Interdisciplinary Collaboration," which was hosted at the Philosophy Section of the Technical University of Delft (October 2010). The workshop was organized as part of the research programme "Moral Responsibility in R&D Networks," supported by the Netherlands Organization for Scientific Research (NWO) under grant number 360-20-160.

The editorial work by Ibo van de Poel and Neelke Doorn was made available through the research programme "New Technologies as Social Experiments: Conditions for Morally Responsible Innovation," supported by the Netherlands Organization for Scientific Research (NWO) under grant number 016.114.625.

The chapter written by Erik Fischer and Daan Schuurbiers (Chap. 5) and the chapter written by Michael E. Gorman *et al.* (Chap. 8) are based upon work supported by the National Science Foundation under grant number 0849101.

The editors wish to thank all presenters at the workshop and the additional authors for their contribution. We gratefully acknowledge the effort that all authors mustered throughout the entire process of planning, composition, and revision of the manuscript.

Contents

Part I Introduction

Part II Approaches to Early Engagement

Part III Reflections

Part IV Conclusions

Contributors

Matthias Achternbosch Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology, Karlsruhe, Germany

Alan Borning Department of Computer Science and Engineering, University of Washington, Seattle, WA, USA

Antonio Calleja-López Department of Sociology, Philosophy and Anthropology, University of Exeter, Exeter, UK
Department of Metaphysics and Current Trends in Philosophy, Ethics and Political Philosophy, University of Seville, Seville, Spain

Jane Calvert Science, Technology and Innovation Studies, University of Edinburgh, Edinburgh, UK

Shannon N. Conley School of Politics and Global Studies, Arizona State University, Tempe, AZ, USA

Neelke Doorn Department of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands

Erik Fisher CSPO, Arizona State University, Tempe, AZ, USA

Batya Friedman Information School, University of Washington, Seattle, WA, USA

Michael E. Gorman Department of Science, Technology & Society, University of Virginia, Charlottesville, VA, USA

Armin Grunwald Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology, Karlsruhe, Germany

Alina Huldtgren Interactive Intelligence Group, Delft University of Technology, Delft, The Netherlands

Peter H. Kahn Jr. Department of Psychology, University of Washington, Seattle, WA, USA

Farzad Mahootian Liberal Studies Program, New York University, New York, NY, USA

Eleonore Pauwels Science and Technology Innovation Program, Woodrow Wilson International Center for Scholars, Washington, DC, USA

Arie Rip Department of Science, Technology and Policy Studies, School of Management and Governance, University of Twente, Enschede, The Netherlands

Douglas K.R. Robinson TEQNODE Limited, Paris, France
IFRIS-LATTS, Université de Paris-Est, Paris, France

Daan Schuurbiers De Proeffabriek (The Pilot Plant), Delft, The Netherlands

Rinie van Est Technology Assessment Department, Rathenau Instituut, The Hague, The Netherlands
School of Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands

Ibo van de Poel Department of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands

Simone van der Burg IQ Healthcare, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands

Author Bios

Dr. Matthias Achternbosch has a background in Chemistry. Between 1988-1992, he was academic assistant and researcher at the Institute of Inorganic Chemistry at the University of Karlsruhe (TH). He joined the Institute for Technology Assessment and Systems Analysis (ITAS) at Karlsruhe in 1992. He is currently leader of several system analytical projects focussed on material flows, innovative materials and cements. Since 2009, he is leader of the KIT system group Cementitious Materials at the Karlsruhe Institute for Technology (KIT).

Alan Borning is a Professor in the Department of Computer Science and Engineering at the University of Washington and an adjunct professor in the Information School. He received a B.A. from Reed College in 1971, and a Ph.D. from Stanford University in 1979. His current research is primarily in human-computer interaction and designing for human values, in particular on systems to support civic engagement and deliberation, and on tools to make public transportation easier and more fun to use.

Antonio Calleja-López is a researcher doing Ph.D. work in Philosophy and Sociology at the Universities of Sevilla (Spain) and Exeter (UK). After undertaking several laboratory studies with a Socratic edge as member of the STIR project (http://cns.asu.edu/research/stir), he has moved onto research on sociotechnical movements, such as 15M in Spain, and Occupy internationally. He is currently involved in different projects at the intersection of technology, art and activism, and has become a member of the research collective Datanalysis15M (http://datanalysis15m.wordpress.com/).

Jane Calvert is Reader in Science, Technology and Innovation Studies at the University of Edinburgh. She is particularly interested in the social dimensions of systems and synthetic biology, including attempts to make biology into an engineering discipline, the role of social scientists in emerging fields, intellectual property and open source, and design and aesthetics in synthetic biology.

Shannon N. Conley is a Socio-technical Integration Research (STIR) Fellow and doctoral candidate in Political Science at Arizona State University (ASU). Shannon studies public policy and political theory, specifically focusing on the politics of new and emerging technologies. Shannon's research is comparative in nature, and delves into the governance of genetics and assisted reproductive technologies at both the national and laboratory levels. Shannon also serves as an adjunct Political Science faculty member at Rio Salado Community College.

Neelke Doorn has a background in civil engineering and philosophy. She is currently affiliated to Delft University of Technology and the 3TU.Centre for Ethics and Technology. For her Ph.D. research, she was involved in an ethical parallel research on Ambient Intelligence Technology. Neelke Doorn has published on topics within the field of applied ethics, notably engineering ethics and medical ethics, responsible innovation, and water governance. Her current research focuses on distributive issues in technology regulation and water governance.

Erik Fisher is an Assistant Professor at Arizona State University with a joint appointment in the School of Politics and Global Studies and the Consortium for Science, Policy and Outcomes. He also serves as the Associate Director of Integration at the Center for Nanotechnology in Society (CNS-ASU). Fisher leads the Socio-Technical Integration Research project as its PI and leads RTTA 4: Integration & Reflexivity at the CNS-ASU. Fisher studies the multi-level governance of emerging technologies, spanning the nested chains of agency from "lab to legislature." Fisher's work has appeared in various journals including *Research Policy*, *Science and Public Policy*, *Technology in Society* and *NanoEthics*. He guest edited a special issue of *Science and Engineering Ethics* on "Science and Technology Policy in the Making: Observation and Engagement." Fisher holds a Ph.D. in Environmental Studies (University of Colorado, Boulder), an M.A. in Classics (University of Colorado, Boulder), and a B.A. in Philosophy and Mathematics (St. John's College in Annapolis, MD).

Batya Friedman is a Professor in the Information School, adjunct professor in the Department of Computer Science, and adjunct professor in the Department of Human-Centered Design at the University of Washington where she directs the Value Sensitive Design Research Lab. She received both her B.A. and Ph.D. from the University of California at Berkeley. Dr. Friedman pioneered value sensitive design. Currently she is working on multi-lifespan information system design and on methods for envisioning information systems to shape human futures.

Michael E. Gorman earned a Ph.D. (1981) in Social Psychology at the University of New Hampshire. He is a Professor in the Department of Engineering and Society at the University of Virginia, where he teaches courses on ethics, invention, psychology of science and communication. He worked for two years as a Program Director in the Science, Technology & Society program at the National Science Foundation and is President of the International Society for Psychology of Science and Technology. His research interests include experimental simulations of science, described in *Simulating Science* (Indiana University Press, 1992) and case studies that combine ethics, invention and design, described in *Ethical and Environmental Challenges to Engineering* (Prentice-Hall, 2000). He has edited volumes on *Scientific and Technological Thinking* (Lawrence Erlbaum Associates, 2005), and *Trading Zones and Interactional Expertise: Creating New Kinds of Collaboration* (MIT Press, 2010). His current research is on the kinds of interdisciplinary trading zones that will be needed for scientists, engineers and other stakeholders to collaborate.

Professor Dr. Armin Grunwald is physicist by education. After occupations in industry and research institutes he is now full professor of philosophy and ethics of technology at Karlsruhe Institute of Technology (KIT), director of the Institute for Technology Assessment and Systems (ITAS) at KIT and director of the Office of Technology Assessment at the German Parliament at Berlin. His main research areas are theory and methodology of Technology Assessment, theory and methodology of sustainable development, and ethics of technology.

Alina Huldtgren is a postdoctoral researcher in the field of Human-Computer Interaction at the Interactive Intelligence Group at Delft University of Technology, where she received her Ph.D. in 2012. Dr. Huldtgren's research interests include human-centered design, co-design and value sensitive design. Her work focuses on socio-technical solutions in the area of health and active ageing. In particular, she investigates how to involve stakeholders in ICT design and how to account for stakeholders' values, needs and perspectives throughout the design process.

Peter H. Kahn, Jr. is a Professor in the Department of Psychology and Director of the Human Interaction With Nature and Technological Systems (HINTS) Lab at the University of Washington. His publications have appeared in such proceedings as CHI, Human-Robot Interaction, and Ubicomp. He is the author of *Technological Nature: Adaptation and the Future of Human Life* (MIT Press, 2011). His research projects are currently being funded by the National Science Foundation.

Farzad Mahootian, Ph.D., is a member of the Liberal Studies faculty of New York University, and an affiliated scholar with the Consortium for Science Policy and Outcomes at Arizona State University. He received his PhD in Philosophy from Fordham University, and an MS in Chemistry from Georgetown University. Farzad's interest in the role of metaphor in science and culture drives his interdisciplinary teaching and research activities ranging from science and technology studies to the philosophy of chemistry.

Eleonore Pauwels is a research associate at the Woodrow Wilson International Center for Scholars, affiliated with the Center's Science and Technology Innovation Program. Pauwels conducted research funded by the US National Foundation and the Alfred P. Sloan Foundation. Her research analyses how life sciences are becoming foundational epistemologies of our times. She reflects on the extent to which this production of epistemologies influences who gets to anticipate our biotechnical futures and the "matters-of-concern," which are emerging in the aftermath.

Arie Rip was Professor of Philosophy of Science and Technology at the University of Twente. After his retirement he continued to lead the Technology Assessment of Nanotechnology Program in the Dutch R&D Program NanoNed, which developed an approach for Constructive TA of emerging technologies. His other work focuses on new modes of knowledge production and changing institutions of science, also in relation to science and technology policy.

Douglas K.R. Robinson is Managing Director of TEQNODE Limited, Paris which provides strategic technology intelligence and future-oriented analysis on emerging science and technology fields. Alongside this, Douglas currently participates as a part-time post-doctoral researcher at IFRIS-LATTS. These two activities combine fruitfully to further his research interests in technology and innovation dynamics, the emergence and construction of new markets and tool development to capture traces of impacts of science and technology (to inform decision making).

Daan Schuurbiers is Director of the Pilot Plant (*De Proeffabriek*), a consultancy for responsible innovation. Daan studied chemistry and philosophy at the University of Amsterdam and has a Ph.D. in ethics of technology from Delft University of Technology. His research centers on the social and ethical dimensions of newly emerging science and technologies. In his work for the Pilot Plant, he advises knowledge-intensive organizations on ways to broaden reflection and strengthen stakeholder engagement with research and innovation.

Rinie van Est is research coordinator and "trendcatcher" with the Rathenau Institute's Technology Assessment (TA) division. He has a background in applied physics and political science. At the Rathenau Institute he is primarily concerned with emerging technologies such as nanotechnology, cognitive sciences, persuasive technology, robotics, and synthetic biology. In addition, to his work for the Rathenau Institute, he lectures Technology Assessment and Foresight at the School of Innovation Sciences of the Eindhoven University of Technology.

Ibo van de Poel is Anthonie van Leeuwenhoek Professor in Ethics and Technology at Delft University of Technology in the Netherlands. He has published on engineering ethics, the moral acceptability of technological risks, values and engineering design, moral responsibility in research networks, and ethics of new emerging technologies like nanotechnology. His current research focuses on new technologies as social experiments and conditions for morally responsible experimentation.

Dr. Simone van der Burg is senior researcher at IQ healthcare, Radboud University Nijmegen Medical Center (the Netherlands). Her research projects are primarily focused on the development of a cooperative and constructive ethics, which enhances ethical reflection among scientists, medical professionals and patients about the technologies that they develop/use. Themes of interest include: genomics, newborn screening, personalized medicine, patient involvement, lay ethics and moral imagination.

Part I
Introduction

Chapter 1
Mandates and Methods for Early Engagement

Daan Schuurbiers, Neelke Doorn, Ibo van de Poel, and Michael E. Gorman

Abstract This introduction to the volume "Early engagement and new technologies: Opening up the laboratory" sets out how recent policy developments have demonstrated a growing interest in early engagement with technology, and identifies the various ways in which scholars from the social sciences and humanities have responded to these policies. The five main approaches elaborated in this volume are introduced: Constructive Technology Assessment (CTA), Value Sensitive Design (VSD), Socio-Technical Integration Research (STIR), Network Approach for Moral Evaluation (NAME), and Political Technology Assessment (PTA). A range of broader issues related to technology engagement is identified and an outline of the volume chapters is presented.

1.1 Why Early Engagement?

Recent years have seen a notable rise in attempts at "early engagement" with science and technology, often in the form of interdisciplinary interactions between researchers from the social sciences and humanities with those in science and engineering. The aim of these interventions has been to attune research and innovation

---

D. Schuurbiers (✉)
De Proeffabriek (The Pilot Plant), Lookwatering 36, 2614 KA Delft, The Netherlands
e-mail: daan@proeffabriek.nl

N. Doorn • I.R. van de Poel
Department of Technology, Policy and Management, Delft University of Technology,
PO Box 5015, 2600 GA Delft, The Netherlands
e-mail: n.doorn@tudelft.nl; i.r.vandepoel@tudelft.nl

M.E. Gorman
Department of Science, Technology & Society, University of Virginia,
351 McCormick Road, Charlottesville, VA 22904-4744, USA
e-mail: meg3c@virginia.edu

N. Doorn et al. (eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*,
Philosophy of Engineering and Technology 16, DOI 10.1007/978-94-007-7844-3_1,

processes to societal needs. In response to science policy calls in the US, Europe and elsewhere, interdisciplinary collaborations are emerging at the heart of research and innovation – in laboratories, at design tables and on production floors – with the aim to open up the laboratory to social deliberations and concerns.

These engagements extend the ethnographic approaches in Science and Technology Studies (STS) that have become known as "laboratory studies." Although laboratory studies in a sense also open up the laboratory by disclosing "the process of knowledge production as 'constructive' rather than descriptive" (Knorr Cetina 1981: p. 140), the engagement studies described in the current volume explicitly aim at opening up the *processes* that take place in the laboratory. By broadening the reflective processes of the actors involved, these approaches "open up the laboratory" (broadly conceived) to a wider set of disciplines and views than those normally engaged in techno-scientific rationality. In other words, where traditional laboratory studies predominantly aim at *observation*, the engagement studies described here explicitly *engage* with the processes that take place in the laboratory.

Several scholars have recently reflected on the impact of attempts at socio-technical integration. A collection of viewpoints edited by Peter Stegmaier appeared in EMBO Reports as part of a Science & Society Series on Convergence Research in 2009. The contributions in the series explore opportunities and challenges for "convergence work," integrating social and humanist research into large research programs. More recently, a special issue appeared in Science and Engineering Ethics entitled "Science and Technology in the Making: Observation and Engagement." Edited by Erik Fisher, the special issue explores the various exercises by science studies scholars to attune science to its public contexts (Fisher 2011).

What is still missing is an overview of the various approaches to early engagement, an identification of their relative strengths and weaknesses, and the discussion of a range of broader issues that need to be addressed to further develop comprehensive forms of technology engagement. This volume aims to fill that gap. It examines both the mandates for early engagement and the methods that have emerged in the wake of this interest. Importantly, it surveys the current state of the art of approaches for early engagement in the laboratory. The volume brings together leading scholars in the field of early engagement with new technologies such as nanotechnology, synthetic biology, biotechnology, and Information and Communication Technology (ICT). It provides an overview of the approaches and methods for early engagement with new technologies (at the "laboratory floor") and discusses some of the major challenges. The volume aims to define how each of these approaches enables engagement, what sets them apart and, perhaps more importantly, what binds them. To what extent do these approaches, individually or collectively, speak to the broader challenge of science policy to embed science in society?

1.2 Policy Calls

Noting the recent rise of interdisciplinary collaborations, one might ask: whence this growing interest in early engagement with science and technology? This section will trace the relevant policy origins of engagement work. Admittedly, there is no single historical rationale that traces the development of methods back to one specific point in time. Research policies, practices and reflexive processes co-evolve: they mutually shape each other over time in complex ways. Policies are shaped as much by the latest academic insights as by the political preferences and public perceptions of the day, and evolve within existing practices and power divisions; "new" methods in social research draw on their academic legacies (often including political views on the role of research in society as well); and research and innovation practices respond and adapt to these influences as they arise (while often maintaining a remarkably stable identity). When discussing the main approaches to early engagement in this volume, we acknowledge the historical intertwinement of research practices and reflexive processes, and we will trace some of the relevant historical contexts in Part II of this volume.

That said, a number of research policies in the early years of the twenty-first century have been so influential in steering attention towards interdisciplinary engagement that they deserve a special mention in this introduction. They have led to a noticeable response, at least from scholars in the social sciences and humanities. In the US, the twenty-first century Nanotechnology Research and Development Act mandates the integration of nanotechnology R&D with research on societal, ethical and environmental concerns (Fisher and Mahajan 2006). This Act proved particularly important for engagement work, because its particular wording seems to have responded directly to earlier calls for engagement by academics such as Davis Baird, in testimony before the Senate Committee on Commerce, Science and Transportation, May 1, 2003: "Ethicists need to go into the lab to understand what's possible. Scientists and engineers need to engage with humanists to start thinking about this aspect of their work. Only thus, working together in dialog, will we make genuine progress on the societal and ethical issues that nanotechnology poses." Again, it is not the first time that this view has been brought forward; interdisciplinary research has a much longer history, and so has the view that ethical reflection should become part and parcel of research and innovation. Ever since C.P. Snow famously identified the "Two Cultures" (Snow 1959), many have tried to bridge the gap. Ziman for instance has argued in a Science commentary (Ziman 1998) that: "ethical reflection should become a part of the "ethos" of science." Mitcham (2003) has identified a "co-responsibility" for a broadened notion of research integrity.

Still, the twenty-first century R&D Act is noteworthy because it has inspired scholars from the social sciences and humanities to actively seek out research practitioners within research settings. The general idea behind these calls for early intervention is that socio-technical systems are easiest to modify in their early stages, before "lock in" has occurred – but the least is known about the effects of technologies on the environment, health and economies. They are hard to modify later because they are so woven into systems. At that point, the remediation of ill effects may be the only strategy. Early engagement, the argument goes, creates the possibility of getting engineers and scientists involved with social scientists and humanists while there are still maximum degrees of freedom for choosing technological directions. Of course, there is a downside to this, as was noted by Collingridge (1980) in his book *The Social Control of Technology*. What has become known as the Collingridge dilemma is that while technologies are malleable in their early stages, their societal consequences are very hard, if not impossible to predict.

So apart from the methodological question of how to *organize* early engagement so as to modulate research and innovation in early stages, it is an open question to what extent we can *anticipate* the societal consequences of future research advances. That said, collaborative inquiry on socio-technical scenarios could well serve to establish a robust vision for technological development and guide our collective imagination. As will be seen in the remainder of this volume, each of the approaches discussed in this volume formulates its own response.

Calls for cooperation have become more prominent in European research policies as well. The European Commission for instance envisages: “the initiation of new forms of partnerships between researchers and others actors through “co-operative research” in its Science in Society Work Programme (European Commission 2007: 10): “The challenge today is to encourage [scientific and technological] actors in their own disciplines and fields to participate in developing Science in Society perspectives from the very beginning of the conception of their activities” (European Commission 2007: p. 6). More recently, mandates for early engagement were enshrined in the notion of Responsible Research and Innovation. The concept of Responsible Research and Innovation marks the evolution of the European Commission’s Science in Society Programme. It represents a more integrated approach that addresses the whole innovation process and engages stakeholders in early stages of research though a collaborative approach, seeing ethical considerations not as constraints, but as drivers for innovation. Responsible Research and Innovation will be a cross-cutting issue in Horizon 2020, the Commission’s major funding program for research and innovation for the period 2014–2020. One might expect attention to collaboration in European research to increase as a result.

In addition to European policies, there is a range of national programs within Europe that call for early engagement with science and technology. The Netherlands Organisation for Scientific Research has established a research programme for “Responsible Innovation” in 2008 that performs studies of the ethical and societal aspects of technological innovations in interaction and cooperation with the technical scientists involved (NWO 2008). According to NWO, the key to responsible innovation lies in adjusting the innovation process by early recognition of social, ethical and acceptance issues in interaction and cooperation with scientists.

1.3 The Legacy of Technology Assessment

This global trend of research policies to become more focused on increased collaboration and integration finds its roots in earlier traditions. Newly emerging concepts of engagement with technology find their background in the tradition of Technology Assessment (TA) and related traditions like Risk Assessment (RA), and, more recently, ELSI (ethical, legal and social implications of technology). The notion of engagement also chimes with movements that aim to include minority views in design and technology development, like the appropriate technology movement (e.g., Nieusma 2004), frugal design (e.g., Bhatti et al. 2013), and inclusive or universal design (e.g., Connell and Sanford 1999).

To understand the influence of these traditions, the chapter following this introduction will therefore sketch this historical background, with an emphasis on TA (cf. Chap. 2).

Traditional TA was originally far removed from the laboratory – it was primarily aimed at informing policy makers and politicians about the impacts of technology, so that they could make better decisions. The idea in the early days of TA was that social consequences of new technologies could be foreseen. In the course of time, it has turned out that assessment of technologies is fraught with uncertainties and unforeseeable variables. Technology and society turned out to be intertwined in complex ways. Consequently, the emphasis in TA moved from objective assessment of expected consequences to *anticipation* of possible consequences--for example by sketching possible scenarios.

This shift in emphasis from prediction to anticipation has been accompanied by three other shifts in the focus of TA. One is that in the early days TA almost exclusively focused on the role of the government, while in the course of time much more emphasis has been placed on the role of companies and research organizations. A second shift is that TA has become much more proactive and aimed at constructively influencing technology in the making, rather than reactively focusing on technology that has already been developed and its expected impacts, which then are to be addressed (if necessary) by government policy. Third, although TA has always been interdisciplinary, it has increasingly engaged in interdisciplinary cooperation with technological researchers and scientists.

In recent years, several studies have aimed to integrate this threefold shift in the design of next-generation TA approaches, exploring the potential of proactive, practice-based, interdisciplinary collaborations between social and natural researchers for integrating wider ethical and societal considerations in research decisions (Gorman et al. 2004; Van de Poel and Van Gorp 2006; Zwart et al. 2006; Doubleday 2007; Fisher 2007; Consoli 2008; Robinson 2009; Van der Burg 2009; Schuurbiers 2011; Doorn 2012). These studies share a general commitment to "opening up the innovation process, rather than managing it after-the-fact" (Sarewitz 2005: 20). They could therefore be characterised as "lab studies 2.0," extending the traditional laboratory ethnographies of the 1970s and 1980s (Latour and Woolgar 1979) to include distinct engagement tools in addition to observation. On this view, reflection should be integrated in the practices of R&D and become, as Berloznik and Van Langenhove call it, "a built-in monitoring of the R&D process" (Berloznik and Van Langenhove 1998: 27). Technology assessment thus becomes internal to the process of R&D itself, which allows for more reflexive participation by the scientists and engineers under the assumption that this may lead not only to more responsive research outcomes, but possibly to more efficient and effective research as well (ibid.: 30).

Both in the US and in Europe, several attempts have been made to implement and further develop these engagement approaches. In the US, the National Science

Foundation has funded approximately six million dollars to prototype this research at the Center for Nanotechnology in Society at Arizona State University (Sarewitz 2005). Preliminary results show "a receptiveness to collaboration on TA activities, rooted in a desire to contribute to societally beneficial outcomes" (Guston and Sarewitz 2002) and an awareness of the possibility of modulating the decisions accordingly. In the Netherlands, the CSG Center for Society and the Life Sciences (originally named the Center for Society and Genomics) was established to study and improve the relationship between society and the life sciences. The center's research aims at improving the way in which the life sciences meet the expectations and needs of society by mapping out the social, legal and ethical issues surrounding the life sciences and engaging researchers, policy makers and lay citizens in constructive dialogue on these issues. Also in the Netherlands, the philosophy departments of the three technical universities have established the 3TU. Centre for Ethics and Technology, which identifies "responsible innovation," "ethical parallel research," and "value sensitive design" as its distinguishing research methods. In terms of funding, the Dutch NanoNextNL consortium, consisting of more than 100 companies, universities, knowledge institutes and university medical centres, has available a total sum of 250 million euros for research into micro and nanotechnology, of which six million euros is dedicated to Technology Assessment. Within this NanoNextNL project, nanoscience and engineering PhD students are offered the opportunity for 3 months projects to study societal and/or governance aspects of their topic, and to write up the results as a chapter in their PhD thesis. This "PhD+" trajectory is supported by researchers in the TA theme.

In Italy, the Department of Philosophy and the Department of Comparative Law of Padua University, established the CIGA Center, with linkages to Veneto Nanotech: the European Center for the Sustainable Impact of NT (ECSIN). This multidisciplinary research center was established to initiate research projects and organize outreach activities on the ethical, social, and legal aspects involved into the development of nanotechnologies and other new technoscientific innovations, with the explicit mission to promote interdisciplinary exchange. Cooperation exists with other research centers, primarily in Eastern and Central Europe.

1.4 Approaches to Early Engagement

While interdisciplinary engagements existed before the rise of the policy mandates mentioned in Sect. 1.2, recently a number of new approaches have emerged that look more closely at research and innovation processes themselves. In addition to "downstream" outreach and public dialogue on the outcomes of research and "upstream" engagement with research objectives, attention has moved towards the "midstream" of research and innovation: the day-to-day research decisions that are at the heart of the innovation process itself.

In this volume, we provide an overview of recent approaches to early engagement. Our interest is specifically with approaches that carry out TA-like activities concurrently to research & development (R&D) and with an involvement of engineers and scientists. Moreover, we focus on approaches in which these TA-like activities have a normative element and aim at encouraging ethical reflection. We are aware that different authors might have different notions of “normativity” and “ethics” and might disagree about the exact aims of ethical reflection – yet to some extent the engagement approaches seem to be underwritten by the normative goal to enable “better technology in a better society” (Schot and Rip 1997). In the concluding chapter, we will reflect on the various normative engagements inherent in the different approaches. The third element shared by the approaches presented here is their focus on interdisciplinarity. At the very least they involve an engineering/scientific and a humanities/social science component, but often they bring together scholars from different disciplines including engineers, social scientists and ethicists.

On the basis of these three criteria we have selected five approaches that will be discussed in part two of the volume:

1. Constructive Technology Assessment (CTA). Characteristic for CTA is that it shifts the focus away from assessing impacts of new technologies to broadening design, development, and implementation processes, with an emphasis on dialogue among and early interaction with new actors. We will particularly focus on “insertion” which is aimed at making the co-evolution of technology and society more reflexive.
2. Value Sensitive Design (VSD): VSD refers to an approach to the design of technology that accounts for human values in a principled and systematic manner throughout the design process (Friedman and Kahn 2003).
3. Socio-Technical Integration Research (STIR): STIR is an approach aimed at stimulating awareness of the possibility of modulating research decisions. The focus is not on the nature of societal concerns, but rather on the nature of engineering decisions, and on the potential capacity of researchers to perform integration by “modulating” their decisions (Fisher and Mahajan 2006).
4. Network Approach for Moral Evaluation (NAME): NAME is an approach that has a threefold objective. The first is aimed at identifying moral issues in R&D networks; the second at ethical reflection and judgment on these issues, and the third on the distribution of responsibilities for addressing these issues. Crucial for successful application of this approach is its ability to take the plurality of the moral views in the network into account (Zwart et al. 2006; Van de Poel and Zwart 2010).
5. Political Technology Assessment (PTA). PTA aims to inform and contribute to opinion formation of politicians and policy makers. Cooperation with politicians or policy makers is an important way to actively involve them in the debate on emerging technologies. Although PTA analysts do not directly aim at getting embedded in the technical research itself, many PTA interventions could be covered under the label engagement because they try to engage the policy makers with the technical research and R&D (a.o., organizing on-site debates between policy makers, politicians, and technical researchers). Recent experiences with interdisciplinary engagements also indicate that the policy and political level

> should be included to make the engagement effective (as also mentioned in Chap. 3 in this volume). We have therefore chosen to include PTA as a separate method.

Although primarily discussed by American and European authors, the methods mentioned above have been applied to countries around the world. The STIR project, for example, reflects a global effort with lab studies being done in China, South-Korea, Spain, Norway, the Netherlands, the US, Canada, Belgium, and France. The VSD approach has recently been applied to the development of an information system in Rwanda (Yoo et al. 2013).

1.5 Broader Questions

Whereas part II of this volume presents each of these methods in further detail, the findings of the various interdisciplinary collaborations leave open a range of further questions: what are the likely roles that embedded researchers can assume, and how do the different roles play out in terms of broadening reflection (Calvert and Martin 2009)? At what stages of the R&D process is the explicit consideration of social and ethical concerns most likely to affect research decisions? What kinds of concerns can or should be addressed? How to strike a balance between collaboration and critique? And how to overcome power differences and avoid the risks of co-optation that inevitably adjoin the various forms of "embeddedness"? There are also principled questions concerning the kind of "normativity" or "ethics" involved, and methodological questions concerning the way how to shape the involvement of the intervening researchers. These points are related. Whereas TA traditionally belonged to the more descriptive sciences, applied ethicists have recently become more actively involved in this field. This raises questions as to the normative content of these approaches. What is ethical reflection, and what is the aim of ethical reflection? What is the role of the embedded researcher? What background should he or she have?

This range of broader questions is taken up in part III of the volume. Issues of collaboration are addressed in the chapters by Gorman et al. (Chap. 8) and by Pauwels (Chap. 11). Gorman and his co-authors conceptualize the interventions that are part of the various approaches discussed in part II as trading zones, where "people from different perspectives and agencies can work together to define a common goal in a way that would be acceptable to their core communities" (Gorman et al.; Chap. 8). These trading zones may enable scientists and engineers to work across apparently incommensurable barriers of language, culture and practice. Thinking in terms of trading zones may help finding the right institutional setting for early engagements and overcome some of the challenges that the respective methods have to deal with (continuity, commitment).

Another range of issues relates to the role of the social scientist. These are primarily discussed in the chapters by Calvert (Chap. 9) and by Pauwels (Chap. 11). These chapters illustrate the different roles an engaged social scientist can adopt, including critic, collaborator, guide, midwife and gadfly. As Calvert discusses in

Chap. 9, one of the most pressing issue relates to the distinction between social scientific researcher and scientific informant. In collaborative research, this distinction has become problematic. She introduces the notion of "paraethnographer" to refer to the former informant, a role that is now replaced with that of discussant. In these new collaborative relationships, the role of the social scientists engaged with scientific and technological fields could best be seen as that of a jester or trickster.

A third issue is the role of time and history, to which Van der Burg (Chap. 10) draws attention. Technologies often have a history which partly determines their further development and social impacts, an aspect that is relatively understudied by the approaches discussed in part II. A similar observation applies to the context of technology regulation. In a recent paper on the regulation of nano-products, Elen Stokes point out that new technologies are often deferred by existing regulatory regimes, which may "involve the reproduction of deeply ingrained traditions and assumptions which, under the weight of history, makes scrutiny extremely difficult" (Stokes 2012; 93). Not paying attention to this legacy may result in the application of ill-suited rules and standards.

1.6 Outline of the Volume

The volume is divided into three parts. Part I (Exploring the Terrain) provides a history of TA approaches and presents the different approaches that are currently applied. Following this introductory chapter, Armin Grunwald provides an overview of developments in the field of Technology Assessment; it describes the increasing attention for early engagement with new technology and the reasons behind it and discusses how the field of TA has traditionally tried to deal with various challenges.

In Part II, five different approaches are discussed in more detail, viz. Constructive Technology Assessment (CTA), Value Sensitive Design (VSD), Socio-Technical Integration Research (STIR), Network Approach for Moral Evaluation (NAME), and Political Technology Assessment (PTA).

In Chap. 3, Arie Rip and Douglas Robinson discuss CTA, more in particular the insertion of a CTA actor who moves actively within and across the "multi-layered landscape" of technological design and development, observing and intervening in "soft" ways. CTA actors may also host workshops, linking multiple stakeholders to develop sociotechnical scenarios that anticipate possible future courses of technologies. The CTA actor is more than a visitor; he or she is "fitting and stretching," as Rip and Robinson call it.

Chapter 4 is dedicated to Value Sensitive Design (VSD). Batya Friedman, Peter Kahn, Alan Borning, and Alina Pommeranz focus on the ways in which well-established methods in VSD such as direct and indirect stakeholder analyses, Futures Workshops, value scenarios, and Envisioning Cards might be used at different stages during engineering innovation to foreground human values in the engineering process. This chapter discusses new work on multi-lifespan information system design that enlarges the scope and timescale of traditional information system design to engage significant societal problems that are unlikely to be solved within a single human lifespan.

In Chap. 5, Erik Fisher and Daan Schuurbiers discuss the possibility for more reflexive participation by scientists and engineers in the internal governance of technology development. The chapter reviews various historical attempts to govern technoscience and introduces the concept of midstream modulation, through which scientists and engineers, ideally in concert with others, bring societal considerations to bear on laboratory practice. It discusses and evaluates a number of applications of socio-technical integration research.

In Chap. 6, Ibo van de Poel and Neelke Doorn present the Network Approach for Moral Evaluation (NAME). This approach takes the engineers and scientists involved in R&D as entry point for discerning and discussing ethical issues and is to be carried out parallel to the R&D trajectory. The approach consists of two main parts. The first is aimed at discerning ethical issues in R&D networks; the second at ethical reflection and judgment on these issues. The approach is illustrated with an example and some limitations and challenges for the further development and application of the approach are discussed.

Chapter 7 on Political Technology Assessment (PTA), by Rinie van Est, concludes part II of this volume. Van Est describes PTA as boundary work, which requires a trustworthy identity and a helping hand from actors at the other side of the border. This trustworthy identity is primarily an issue of having a good academic record. As Van Est indicates, TA has its roots in academia and this scientific nature is an important part of TA's public identity and political legitimacy. The second requirement for successful Political TA is the ability to effectively communicate with political actors and to build up a long-term relationship of trust.

Part III of this volume provides reflections on the opportunities and challenges for early engagement in the laboratory. In Chap. 8, Mike Gorman, Antonio Calleja-López, Shannon Conley and Farzad Mahootian introduce the concept of "Trading Zone" to explain how actors from different perspectives and agencies can work together to define a common goal in a way that would be acceptable to their core communities. These concepts are illustrated with case studies of upstream and midstream modulation.

In Chap. 9, Jane Calvert discusses the interdisciplinary engagement approaches in relation to laboratory studies in Science and Technology Studies (STS). Calvert reflects on three experiences in interdisciplinary collaboration in the field of synthetic biology. She discusses methodological issues to which these interdisciplinary activities give rise. Based on anthropological literature, Calvert sketches collaboration as a research method in itself, requiring new ways of producing knowledge together with the epistemic partners. Although collaboration as a research method may sound ambitious, Calvert warns that it may as well force social scientists to be more modest and give up the claim of ultimate explainers of science.

Drawing on the relevant STS literature, Van der Burg reflects on the different methods and argues for the inclusion of a historical biography of the emergent technologies in Chap. 10. Many engagement studies focus on the future in which the technology will be implemented. However, as Van der Burg points out in her contribution, new technologies are partly dependent on their technological ancestors, and a study of technological ancestry may assist in articulating part of the values they incorporate.

Chapter 11, by Eleonore Pauwels, describes the contexts for cross-field collaborations within the life sciences and highlights relevant theoretical reflections on the concepts of "insertion," "modulation," and "trading zones." Pauwels discusses metaphors used by the synthetic biology community, combining data mining with observation and interviews. She also discusses elements of a manifesto that lists principles for collaboration among social scientists, ethicists, scientists and engineers.

In the concluding Chap. 12 of this volume, the editors reflect on the achievements of the respective methods and develop a tentative framework for Comprehensive Technology Engagement, identifying the major questions and challenges to advancing interdisciplinary engagements at early stages of technological development.

References

Berloznik, R., & Van Langenhove, L. (1998). Integration of technology assessment in R&D management practices. *Technological Forecasting and Social Change, 58*, 23–33.

Bhatti, Y. A., Khilji, S. E., & Basu, R. (2013). Frugal innovation. In S. Khilji & C. Rowley (Eds.), *Globalization, change and learning in South Asia*. Oxford: Chandos Publishing.

Calvert, J., & Martin, P. (2009). The role of social scientists in synthetic biology. *EMBO Reports, 10*, 201–204.

Collingridge, D. (1980). *The social control of technology*. New York: St. Martin's Press.

Connell, B. R., & Sanford, J. A. (1999). Research implications of universal design. In E. Steinfeld & G. S. Danford (Eds.), *Enabling environments: Measuring the impact of environment on disability and rehabilitation*. New York: Kluwer.

Consoli, L. (2008). The intertwining of ethics and methodology in science and engineering: A virtue-ethical approach. *Interdisciplinary Science Reviews, 33*, 234–243.

Doorn, N. (2012). Exploring responsibility rationales in research and development (R&D). *Science, Technology & Human Values, 37*, 180–209.

Doubleday, R. (2007). The laboratory revisited: Academic science and the responsible governance of nanotechnology. *NanoEthics, 1*, 167–176.

European Commission (2007). *Work programme 2007*, *capacities*, *part 5*, *Science in society*, *C*(*2007*) *563*. Brussels: Office for Official Publications of the European Communities.

Fisher, E. (2007). Ethnographic invention: Probing the capacity of laboratory decisions. *NanoEthics, 1*, 155–165.

Fisher, E. (2011). Editorial overview: Public science and technology scholars: Engaging whom? *Science and Engineering Ethics, 17*, 607–620.

Fisher, E., & Mahajan, R. L. (2006). Midstream modulation of nanotechnology research in an academic laboratory. In *ASME International Mechanical Engineering Congress and Exposition* (*IMECE2006*) Nov 5–10, 2006, Chicago, Ill, USA (pp. 1–7).

Friedman, B., & Kahn, P. H., Jr. (2003). Human values, ethics and design. In J. Jacko & A. Sears (Eds.), *Handbook of human-computer interaction*. Mahwah: Lawrence Erlbaum Associates.

Gorman, M. E., Groves, J. F., & Shrager, J. (2004). Societal dimensions of nanotechnology as a trading zone: Results from a pilot project. In D. Baird, A. Nordmann, & J. Schummer (Eds.), *Discovering at the nanoscale* (pp. 63–73). Amsterdam: IOS.

Guston, D. H., & Sarewitz, D. (2002). Real-time technology assessment. *Technology in Society, 24*, 93–109.

Knorr Cetina, K. (1981). *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. Oxford: Pergamon Press.

Latour, B., & Woolgar, S. (1979). *Laboratory life: The construction of scientific facts*. Berverly Hills: Sage.

Mitcham, C. (2003). Co-responsibility for research integrity. *Science and Engineering Ethics, 9*, 273–290.

Nieusma, D. (2004). Alternative design scholarship: Working towards appropriate design. *Design Issues, 20*(3), 13–24.

NWO. (2008). *Responsible innovation: Description of thematic programme*. The Hague: Netherlands Organisation for Scientific Research.

Robinson, D. K. R. (2009). Co-evolutionary scenarios: An application to prospecting futures of the responsible development of nanotechnology. *Technological Forecasting and Social Change, 76*, 1222–1239.

Sarewitz, D. (2005). This won't hurt a bit: Assessing and governing rapidly advancing technologies in a democracy. In M. Rodemeyer, D. Sarewitz, & J. Wilsdon (Eds.), *The future of technology assessment* (pp. 14–21). Washington, DC: Woodrow Wilson International Center for Scholars.

Schot, J. W., & Rip, A. (1997). The past and future of constructive technology assessment. *Technological Forecasting and Social Change, 54*, 251–268.

Schuurbiers, D. (2011). What happens in the lab: Applying midstream modulation to enhance critical reflection in the laboratory. *Science and Engineering Ethics, 17*, 769–788.

Snow, C. P. (1959). *The two cultures and the scientific revolution*. Cambridge: Cambridge University Press.

Stegmaier, P. (2009). The rock'n'roll of knowledge co-production; Science and society series on convergence research. *EMBO Reports, 10*, 114–119.

Stokes, E. (2012). Nanotechnology and the products of inherited regulation. *Journal of Law and Society, 39*, 93–112.

Van de Poel, I. R., & Van Gorp, A. C. (2006). The need for ethical reflection in engineering design: The relevance of type of design and design hierarchy. *Science, Technology & Human Values, 31*, 333–360.

Van de Poel, I. R., & Zwart, S. D. (2010). Reflective equilibrium in R&D networks. *Science, Technology & Human Values, 35*, 174–199.

Van der Burg, S. (2009). Imagining the future of photoacoustic mammography. *Science and Engineering Ethics, 15*, 97–110.

Yoo, D., Lake, M., Nilsen, T., Utter, M. E., Alsdorf, R., Bizimana, T., Nathan, L. P., Ring, M., Utter, E. J., Utter, R. F., & Friedman, B. (2013). Envisioning across generations: A multi-lifespan information system for international justice in Rwanda. In *Proceedings of CHI 2013*. New York: ACM Press.

Ziman, J. M. (1998). Why must scientists become more ethically sensitive than they used to be? *Science, 282*, 1813–1814.

Zwart, S. D., Van de Poel, I. R., Van Mil, H., & Brumsen, M. (2006). A network approach for distinguishing ethical issues in research and development. *Science and Engineering Ethics, 12*, 663–684.

Chapter 2
Technology Assessment and Approaches to Early Engagement

Armin Grunwald and Matthias Achternbosch

Abstract Technology Assessment (TA) emerged in the 1970s as a research-based policy-advising activity. During the 1980s, TA discovered technology development at the lab level as a subject of interest, reflection, and intervention. Since that time, TA as orientation for shaping new technology and innovation has been part of the overall TA portfolio. TA concepts and approaches to early engagement have been developed in different frameworks, e.g. as Constructive TA. In the last 10 years, a new wave of early engagement in TA occurred mainly in the field of new and emerging technologies such as nanotechnology, enhancement technologies, and synthetic biology. This wave led to many activities involving TA in early stages of development. In this chapter, we describe the most relevant approaches in TA aiming at early engagement. A deeper look will be presented into conceptual backgrounds that are specifically relevant to early engagement, such as concepts of technology determinism, social constructivism, and co-evolution. As a case study, we discuss an ongoing activity at the Karlsruhe Institute of Technology (KIT) where a major technological development in the field of new cement is accompanied by systems analysis and innovation research from its very beginning. This case allows specifically for discussing chances and opportunities of early engagement but also pitfalls and obstacles.

A. Grunwald (✉) • M. Achternbosch
Institute for Technology Assessment and Systems Analysis,
Karlsruhe Institute of Technology, P.O. Box 36 40,
Karlsruhe 76021, Germany
e-mail: armin.grunwald@kit.edu; matthias.achternbosch@kit.edu

N. Doorn et al. (eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*,
Philosophy of Engineering and Technology 16, DOI 10.1007/978-94-007-7844-3_2,

2.1 Introduction and Overview

Technology Assessment (TA) emerged in the 1970s as a research-based policy-advising activity. In the first period of TA, technology was regarded to follow its own dynamics (technology determinism) with the consequence that the main task of TA was seen in functions of early-warning of risks and early recognition of opportunities. The objective was to enable political actors to undertake measures to, for example, compensate or prevent anticipated negative impacts of technology. The spheres of technology development at the lab and the issue of shaping technology at the level of products and systems were not addressed at all because of the deterministic attitude concerning technology and the focus on policy advice.

This situation changed in the 1980s. Following the emerging social constructivist paradigm, slogans such as "social shaping of technology" were coined and the approach of Constructive Technology Assessment (CTA) was developed (Rip and Robinson; Chap. 3). TA discovered technology development at the lab level as a subject of interest, reflection, and intervention. Since that time, TA as orientation for shaping new technology and innovation has been part of the overall TA portfolio. This portfolio covers the whole spectrum from the political (e.g. parliamentarian) level far away from the lab up to concrete intervention in engineering, design and development at the level of R&D programmes and concrete projects at the lab (Grunwald 2009a).

An early observation was that, in order to be able to contribute to shaping technology *ex ante* instead of only analysing impacts and consequences of its use *ex post*, TA should be involved *in early stages* of technology development. The concern that TA might be too late in order to do its job has accompanied the development of TA from its beginnings. This has motivated concepts and approaches to early engagement – however, the possibility of early engagement seems to be threatened by the so-called Control Dilemma (Collingridge 1980; see Sect. 2.3 of this chapter) stating that shaping technology will fail anyway: in early stages of development because of lack of knowledge, and in later stages because then there will be no more room for shaping.

In the last 10 years, a new wave of early engagement (sometimes called "upstream engagement") has been observed in TA. It occurred mainly in the field of new and emerging technologies (NEST) such as nanotechnology, nanobiotechnology, enhancement technologies, and synthetic biology. These fields of science and technology development show a strong "enabling character", probably allowing for manifold applications in different areas which are extremely difficult to anticipate. This situation makes it necessary – from a TA perspective – to perform an *accompanying TA process* reflecting on the ethical, social, legal, and economic issues at stake (Rip and Swierstra 2007; Grunwald 2010a).

In this chapter, we describe the most relevant approaches in TA and in neighbouring research activities aiming at early engagement. To this end, we start with a general introduction to TA by mentioning some issues of its history, objectives, and recent developments (Sect. 2.2). Then we present a deeper look into conceptual backgrounds that are specifically relevant to early engagement, referring to the debates around technology determinism, social constructivism, and co-evolution (Sect. 2.3). This look prepares the ground for presenting TA approaches aimed at early engagement such as accompanying systems analysis, the approach of the German Association of Engineers, and the Vision Assessment (Sect. 2.4). As a case study, we discuss an ongoing activity at the Karlsruhe Institute of Technology (KIT) where a major technological development in the field of new cement is accompanied by systems analysis and innovation research from its very beginning (Sect. 2.5). The last section includes some reflections on challenges, limits, and obstacles of the TA approaches developed so far.

2.2 Technology Assessment[1]

TA arose from specific historical circumstances in the 1960s and 1970s far away from any lab context. Activities and concerns in the US political system, in particular in the US Congress, culminated in the creation of the Office of Technology Assessment (OTA) in 1972 (Bimber 1996). The concrete background consisted in the asymmetrical access to technically and politically relevant information between the legislative and executive bodies of the United States. From this point of view, the aim of legislative TA serving the Parliament was to restore parity. This very specific origin of TA found a lot of successors in Europe which succeeded in establishing a lively network (European Parliamentary Technology Assessment EPTA).

Parallel to this specific development in the political system, radical intellectual changes were taking place. The optimistic belief in scientific and technical progress, which had predominated in the post-Second World War period, came under pressure. Broad segments of Western society were deeply unsettled by the “Limits of Growth” which addressed the limitedness of natural resources. Furthermore, problems with unintended side effects, in particular with respect to the natural environment, and new ethical questions led to societal conflicts on the legitimacy of technology. A fundamental dispute on how to deal with science and technology emerged. Far beyond supporting parity between executive and legislative forces in democracy, TA was then expected to contribute to new forms of societal orientation and legitimisation of science and technology. This constellation led to a complex and multi-dimensional set of objectives and rationales of TA.

Nowadays, the term “Technology Assessment” is a widely used designation of the systematic approaches and methods to investigate the conditions for and the consequences of technology and to denote their societal evaluation. According to an existing definition, TA is a scientific, interactive, and communicative process which aims to contribute to the formation of public and political opinion on societal aspects of science and technology (Decker and Ladikas 2004). TA thus provides knowledge,

---

[1] This Section builds on earlier and more comprehensive descriptions of Technology Assessment given by one of the authors (Grunwald 2009a, 2010a).

orientation, and procedures on how to cope with challenges at the interface between technology and society on both directions: TA explores and assesses possible impacts and consequences of technology in a prospective manner on the one hand, and helps to introduce society's expectations and needs concerning new technology into the relevant decision-making processes, on the other.

The focus of TA dwells on the perspective of *unintended side effects* like accidents, negative environmental impacts, ethical problems, and unintended social consequences (Bechmann et al. 2007) and on assessing these in relation to expected benefits and innovation potential. TA has been set up to enable assessment procedures making use of scientific knowledge, ethical orientation, and participatory processes as well (Grunwald 2009a). The mission of TA is, thus, to contribute to "a better technology in a better society" (Rip et al. 1995) in various dimensions and following various objectives in detail.

The first objective of TA was, as indicated above, to support political decision-making mainly in the fields of regulation and research funding. The classical institutions and procedures of democracy should be provided with knowledge and orientation in order to make "better decisions". Reflexivity should be added to political bodies and their decision-making processes. This type of TA is still active and expanding in the fields of parliamentary technology assessment (Cruz-Castro and Sanz-Menendez 2004) and in many forms of advising governmental bodies and authorities.

A second objective is directly related to democracy or, in a more radical sense, aims to prevent a possibly emerging technocracy. From the 1960s on there have been concerns that the scientific and technological advance could threaten the functioning of democracy because only few experts were capable of really understanding the complex technologies (Habermas 1970). The technocracy hypothesis was born painting a picture of a future society where experts would make the decisions with respect to their own value systems by coming up with the "one best solution" for any situation of choice, and with politics restricted to implementing the results. Against this background, one of the many origins of TA is to counteract and to enable and empower society to take active roles in democratic deliberation on science and technology (Schomberg 1999).

A third objective of TA is related to the ways in which society deals with conflicts over new technology. Experiences of severe technology conflicts and of legitimacy deficits have accompanied many Western countries in the past decades. There was (and partially still is) little acceptance of particular political decisions on technology such as on nuclear power and nuclear waste disposal sites in some countries. Also doubts about their legitimacy combined with suspicions of technocratic decision-making, fuelled the emergence of severe conflicts. These developments motivated TA to think about procedures of conflict prevention and resolution, in particular including participatory approaches (Joss and Belucci 2002).

The fourth and in the context of this paper most relevant objective concerns the level of technology itself rather than the ways society deals with technology issues. The idea of shaping technology according to social values was born in the framework of Social Constructivism (Sect. 2.3). If this would succeed, so the hope, problems of rejection or non-acceptance would no longer occur at all, and a better technology, serving the demands of people and of society in a better way, could be reached. This is the very intention of Constructive Technology Assessment (Rip and Robinson; Chap. 3).

An international community established itself around the concept of TA and its various dimensions and diverse objectives, using different concepts and methodologies. Part of this community works in institutions explicitly devoted to TA (e.g., to provide advice to Parliaments, cf., for instance, the EPTA network, www.eptanetwork.org), part of it is organised in networks (cf. Netzwerk Technikfolgenabschätzung, www.netzwerk-ta.net), part is describing its work as systems analysis, in particular in the fields of material flow analysis, energy balancing, and life cycle assessment LCA, and another part converges on the fringes of disciplinary organisations and conferences, such as in sections of sociological or philosophical organisations, or in the STS Community, e.g., under the auspices of EASST (the European Association for the Study of Science and Technology), and of many IEEE (Institute of Electrical and Electronics Engineers) activities relating to the social implications of technology.

2.3 Technology Determinism, Social Constructivism, and Co-evolution

TA needs models of technology development and of the processes of governance influencing research and development in order to be able to fulfil its tasks in the ways intended. These models then influence the way how TA is to be conceptualised, how it is embedded in governance, how it is actually performed, and to which addressees its results shall be customised. To give a hypothetical illustration: if one believed in a technology development model ascribing all power to industry and international companies, TA as policy advice would simply be nonsense.

There are two grand and mutually contradicting narratives underlying different forms of TA: the *Technology Determinism* paradigm, on the one hand, claims that technology cannot be influenced by but determines society. The *Social Constructivism* paradigm, on the other, states that technology can be “constructed’ according to social values. Over the past decade, a third and intermediary paradigm has become the dominant narrative: the model of a *co-evolution* of technology and society. These lines of thought will be described briefly to allow for a better understanding of the options of choice and their implications at the conceptual level.

Scientific progress and technological advance have been the most powerful driving forces of and in society for decades. It is plausible to raise the question whether the dynamics of science and technology has purely internal roots and origins, whether this dynamics could be influenced from an external perspective, e.g. by politics, or not, and in which way society should act according to the answers given to the first two questions. Technology Determinism assumes a strong inherent dynamics of science and technology at place and gives a negative answer to the question whether technology is designable and controllable by society. The course of technology development over time is regarded as a result of that internal and inherent dynamics. Society and policy-makers should, therefore, not aim at steering or shaping technology because this would not be possible in principle. They only could, following this line of thought, prepare themselves for the new technologies that would inevitably come, and try to deal with their impacts and consequences in a most socially compatible way. Technology itself is, in this perspective, no subject for societal and political influence at all – only the way society deals with the impacts of technology could be subject to political measures.

This paradigm was dominant in the 1970s and the beginning of the 1980s. It had decisive impact on the early concepts of TA. If, according to Technology Determinism, technology could not be influenced by society but society had to prepare itself for coming technologies and their impacts (opportunities as well as risks), TA was mainly seen as a means of forecasting and predicting impacts and consequences of technology in order to enhance the opportunities of early societal action and adaptation. In particular, the idea of the "early warning" function of TA was coined against this background. The lab level and the development processes of technology were no serious subjects of interest of TA, simply because of the dominating paradigm mentioned (Grunwald 2009a, 2010a).

Social sciences' research on technology, however, proved naïve technology determinism false and motivated, fuelled by Social Constructivism (Bijker et al. 1987), approaches of social shaping of technology (SST) (Yoshinaka et al. 2003). It was shown that technology is really being "made" and influenced by many groups and institutions in society. The development of technology should, following these ideas, be perceived as the result of societal processes of meaning giving, negotiation, and decision-making. Theories of technology development were established which highlighted the importance of decision-making processes involving many actors at all stages of development and which showed that the resulting decisions depend on values and interests of those actors (Bijker and Law 1994).

The idea of "shaping technology" became, based on this particular paradigm, dominant in the 1990s (Bijker and Law 1994) and motivated new TA concepts, in particular the emergence of Constructive TA (Rip et al. 1995). CTA has lobbied for the early and broad participation of societal actors, including key economic players, and for the establishment of a learning society experimenting with technology. In the normative respect, CTA builds on a basis of deliberative democracy in which a liberal picture of the state highlights self-organising processes in the marketplace.

Social Constructivism has in the meantime been criticised for being too optimistic with regard to the malleability of technology. Critics pointed to path dependencies, to economic forces, and to irreversibilities in technology development setting limitations to shaping approaches. The idea of a co-evolution of technology and society takes this criticism into account and considers both sides as mutually influencing each other and as being closely interlinked with each other (Rip 2007). Following this idea, neither naïve shaping of technology according to social values is possible, nor is society helplessly damned to adapt itself to a self-dynamic technology.

Instead, the situation is complex, and though possibilities for society influencing technology development are seen, it is considered ambitious to realise them.

This development refers in a specific sense to the "Control Dilemma" (Collingridge 1980) which was already briefly mentioned in the introduction. Collingridge looked at technologies in an early stage of development and at mature technologies as well. Regarding the latter, he stated that there could only be a minimal societal influence. For economic reasons it would not be possible to change or modify mature and market-ready technological products or systems to a significant extent. In the best case, minor adaptations might be expected in case of societal pressure. Referring to technologies in an early stage of development, Collingridge problematised the feasibility of targeted societal influence on technology according to societal values, expectations of benefits and the avoidance of risk, due to a severe lack of knowledge about future products and their impacts. The dilemma therefore reads: shaping technologies with regard to societal expectations in an early stage of development must fail because of lack of knowledge, and, shaping technologies in a mature state is not possible because of economic forces. Once the impacts of a technology are relatively well-known, the chances of influencing this technology will significantly decrease. Thus, shaping technology with regard to societal expectations and values would not be possible at all.

The dichotomy expressed in the Control Dilemma is, however, artificial if understood in an extreme way. The question of whether TA should start early and should be prospective or only start when reliable statements about consequences are available poses a false dilemma. The issue here is not an either-or one but the differentiation of TA in line with the problem at hand and with the validity of the available knowledge of the consequences. TA differs conceptually and methodologically depending on whether it is concerned with measurable consequences of technology or with more prospective ones. That means TA should be conceptualised as research and assessment *accompanying* the technology development and using concepts and methods appropriate to the stage of development of the technology under consideration. What "appropriate" means here refers to the governance of the respective field to which TA is to contribute.

TA carried out in early stages of development may look different in the cases of NEST (new and emerging science and technology), where ethical and philosophical questions are in the centre of interest (e.g. Grunwald 2010b), in the field of transforming large infrastructures such as the energy supply system where something like "transition management" in the framework of reflexive governance (Voss et al. 2006) is needed, and in the field of developing new materials and processes where early systems analysis including Life Cycle Assessment is required to help shaping the development towards sustainable development (see the case study in Sect. 2.5). In all of these fields, TA adds specific prospective knowledge and reflexivity to the perspectives of natural scientists, managers, developers, and engineers involved.

Opening up the lab in this context means bringing additional perspectives and TA knowledge to the lab, aiming at supporting decision-making there. This could include, for instance, increasing the awareness of lab researchers with respect to possibly involved ethical issues, organising debates on the responsibilities of researchers in the overall governance of science, bringing knowledge about innovation and diffusion processes to the lab level, and enhancing the consciousness of lab researchers with regard to possible societal implications and consequences of their work. The resulting increase of knowledge and reflexivity could have an influence on ongoing design decisions concerning experiments and development processes but also on the agenda setting processes at the lab level.

2.4 TA Approaches to Early Engagement

The first and still most frequently cited TA approach to early engagement is the already mentioned CTA approach (Rip et al. 1995) initially based on Social Constructivism and later on following the co-evolution picture. In this section, we will not introduce CTA because the Chapter by Rip and Robinson (Chap. 3) is dedicated to it. Instead, we will present the accompanying systems analysis approach (4.1) which is frequently used and mostly not subsumed under the TA label, the concept proposed and implemented by the Association of German Engineers (4.2), the Responsible Innovation approach as a recent development (4.3), and the Vision Assessment looking at visionary communication in the context of NEST (4.4).

2.4.1 Accompanying Systems Analysis

Basic task of an accompanying systems analysis is to assess opportunities, potentials and risks of a new technology in an early stage of development despite existing uncertainties in order to provide decision-makers in politics, science and economy with first information. Such an analysis is not restricted to aspects of engineering and natural science, e.g. estimating conditions for a large-scale technical realisation and possible ecological impacts, but also considers the political, societal and economic framework requirements of the innovation. Besides the opportunities and potentials, the (economic and societal) benefit of the invention is assessed.

The systems analysis starts with a description and analysis of the existing "landscape" (the added value chain and its societal context) where the innovation will be implemented and used in future. Besides an analysis of the current political, economic and technical conditions, this includes a characterisation of the pertaining industry, in particular of its structure and the regulatory framework conditions. Furthermore, stakeholders and lobbies have to be identified which would be affected by including the innovation in the industrial and societal metabolism.

Moreover, technological questions on the conditions for a large-scale technical realisation have to be answered. This comprises the (cumulative) energy expenditure and material flows, taking into account the entire process chain (including the upstream chain) from cradle to grave as well as the availability of raw materials, the cost of their supply, and associated possible ecological effects which have to be identified and assessed. These issues are studied using methods like material flow analysis and energy balancing based on thermodynamic and thermochemical approaches including realistic efficiency estimations. In some cases, the Life Cycle Analysis (LCA) tool can be used (cp. Schepelmann et al. 2009).

Results can be used to advance the development of the technology, focusing on its large-scale technical realisation, and for giving early insights into sustainability effects. The work is an iterative process depending on the current level of the development and new findings. In addition, to foster technical development, the results of the systems analysis are used for a comprehensive assessment of the complete production process, including sustainability aspects.

At the forefront of economic considerations are first estimates of the anticipated capital and operating costs of the new technology. A systems analysis can also include extended economic investigations taking into account issues such as the determination of CO_2 mitigation costs, “hidden extras”, “societal extras”, and problems of discounting. The sustainability assessment usually relies on adequate indicators covering economic, ecological and social aspects to evaluate the technological development.

Positive material properties of a new “socially wanted” product do not automatically guarantee a successful launch. Therefore, also the innovation process has to be investigated with a special focus on fostering development and disclosure of inhibitory factors. Marketing strategies for a product launch, however, are not in the focus of the TA investigation. Instead, the results of the analyses are used to provide the actors involved with supportive knowledge for implementation.

As shown above, a relatively tight involvement of systems analytical investigations in technology development at the lab is, on the one hand, necessary for the systems analytical part, in particular because this constellation allows direct access to relevant data and an in-depth analysis of the innovation process. Close cooperation with systems analysis should, on the other, also be of great benefit for the development project because the results of systems analysis could be used to directly inform developers about possible obstacles and could give hints how to meet sustainability objectives by optimising the design (Poel 2009). Nevertheless, awareness must be raised of possible unwanted effects such as decreasing independence of systems analytical TA (see Sect. 2.5).

2.4.2 Technology Evaluation

The Association of German Engineers (VDI, Verein Deutscher Ingenieure) has been dealing with challenges of technology to society since the 1960s. Many VDI publications address issues such as technology and society, responsibility of engineers and a code of conduct.

The most prominent outcome of these activities is the VDI Guideline No. 3780 (VDI 1991, also available in English), which has become quite well-known, at least in the German-speaking countries. It is intended to provide a “Guide to Technology Assessment According to Individual and Social Ethical Aspects”. For engineers and in industry, assessments are to a certain extent part of their daily work. Evaluations play a central role, for instance whenever a line of technology is judged to be promising or to lead to a dead end; whenever the chances for future products are assessed; whenever a choice between competing materials is made; or whenever a new production method is introduced to a company. Though evaluation may be commonplace in engineering practice, the new thing about this guideline for societal technological evaluation is its scope, which also includes the societally relevant dimensions of impacts as well as technical and economic factors. Technological evaluation should be conducted in line with socially acknowledged values. Eight central values forming the VDI “Value Octagon” have been identified: functional reliability, economic efficiency, prosperity, safety, health, environmental quality, personality development, and social quality (VDI 1991). These values are thought to influence technical action and fall under the premise that it “should be the objective of all technical action […] to secure and to improve human possibilities in life” (VDI 1991, p. 7).

According to the VDI Guideline, the identified values shall be considered in processes of technology development, in particular in technology design (see Poel 2009). They shall virtually be *built into* the technology. Engineers or scientists should, on the basis of their knowledge and abilities, push the development of technology in the “right” direction by observing these values and avoiding undesirable developments. If this exceeds their authority or competence, engineers should take part in the corresponding procedures of technology evaluation. This mode of operation is rather close to Value Sensitive Design (see Chap. 4 in this volume). However, VDI did not pay much attention to how to make this approach work. Although the approach is well integrated in the education of engineers at many technical universities in Germany, it did not have much impact on practical development yet.

As the Guideline addresses directly the actions and decisions of engineers, it is relevant to research and development also in the lab, in publicly funded science as well as in industry. “Opening up the lab” would not be necessary as an extra effort because, following the Guideline, engineers and researchers at the lab level should act according to the values mentioned in the Guideline. However, theory and practice seem to differ considerably in this respect.

2.4.3 Responsible Research and Innovation (RRI)

The ideas of “responsible research” in scientific and technological advance and of “responsible innovation” in the field of new products, services and systems have been discussed for some years now with increasing intensity (Siune et al. 2009). The postulate of Responsible Research and Innovation (RRI) adds explicit ethical reflection to Technology Assessment (TA) and Science, Technology and Society (STS) studies and includes all of them into integrative approaches to shaping technology and innovation (Schomberg 2012). Responsible innovation brings together TA with its experience in assessment procedures, actor involvement, foresight and ethical evaluation, in particular under the framework of responsibility, and also builds on the body of knowledge about R&D and innovation processes provided by STS studies and STIS studies (Science, Technology, Innovation and Society) (Grunwald 2012).

Science institutions, including research funding agencies, have started taking a pro-active role in promoting integrative research and development (see the Responsible Innovation programme of the Dutch National Science Foundation NWO as an example). Thus, the governance of science and of R&D processes is changing, opening up new possibilities and opportunities for involving new actors and new types of reflection. In particular, RRI aims at intervening R&D processes in early stages of development:

> Responsible development of nanotechnology can be characterized as the balancing of efforts to maximize the technology's positive contributions and minimize its negative consequences. Thus, responsible development involves an examination both of applications and of potential implications. It implies a commitment to develop and use technology to help meet the most pressing human and societal needs, while making every reasonable effort to anticipate and mitigate adverse implications or unintended consequences. (National Research Council 2006, p. 73)

The emergence of RRI reflects the diagnosis that available approaches to shaping science and technology still do not meet all of the far-ranging expectations towards technology governance and achieving a "better technology in a better society" (Rip et al. 1995). The hope behind the RRI movement is that new – or further developed – approaches could add considerably to existing approaches such as TA and engineering ethics. Indeed, compared to earlier approaches such as SST or CTA there are shifts of accentuation and new focuses of emphasis (Schomberg 2012; Grunwald 2012):

- "Shaping innovation" complements or even replaces the former slogan "shaping technology" which characterised the social constructivist approach to technology. This shift reflects the insight that it is not technology *as such* which influences society and therefore should be shaped according to society's needs, expectation and values, but it is *innovation* by which technology and society interact as has been pointed out by many STS studies.
- There is a closer look on societal contexts of new technology and science. RRI can be regarded as a further step towards taking the demand-pull perspective and social values in shaping technology and innovation more serious.
- Instead of distant *observation* following classical paradigms of science there is now a clear indication for *intervention* into the development and innovation process: RRI projects shall "make a difference" not only in terms of research but also as interventions into the "real world".

- The spectrum of stakeholders to be involved in participatory processes and dialogue is to be broadened further because of new forms of science and technology governance (Siune et al. 2009; Schomberg 2012).
- Following the above-mentioned issues, RRI can be regarded as a radicalisation of the well-known post-normal science (Funtowicz and Ravetz 1993), being even closer to social practice and being prepared for intervention and for taking responsibility for this intervention.

The concrete realisation within the Responsible Innovation programme of the Dutch Science Foundation makes clear that “opening up the lab” is part of the agenda. Ethicists, TA researchers and STS scholars shall cooperate with natural scientists, engineers, developers, and managers in order to come up with modified and, hopefully, more “responsible” products, systems and services. It is still too early to assess whether the ambitious goals have been reached, and what conditions should be fulfilled to support fulfilling the expectations.

2.4.4 Vision Assessment

Quite often visions and metaphors mark the revolutionary advance of science in general and act as an important factor in societal debates, in particular in NEST (Grunwald 2007 for the case of Human Enhancement). Available studies have shown that futuristic visions are ambivalent: they may cause fascination as well as concern and fear. The main argument for requiring early engagement of TA in the form of a vision assessment is the importance of visions in actual debates, that is, both in the debate on the opportunities afforded by scientific and technological progress and in ongoing risk debates (Grunwald 2007).

Vision assessment is a new TA tool that is not directed at the assessment of technologies but at the assessment of visions which are communicated across the many interfaces between technology and society (Grin and Grunwald 2000). Vision assessment can be analytically divided into (see Grunwald 2009b):

- vision *analysis* – which is itself subdivided into a *substantial* aspect (what is the content of the respective vision?) and a pragmatic aspect (how is it used in concrete communication?),
- vision *evaluation* (how could the content of the vision be evaluated and judged?), and
- vision *management* (how should the people and groups affected deal with the visions?).

The general aim is to provide transparent disclosure of the relationship between knowledge and values, knowledge and the lack of it, and the evaluation of these relationships and their implications. In particular, vision assessment should allow the various and partly divergent normative aspects of visions of the future to directly confront each other in order to improve transparency and clarity of their contents, premises, and meanings.

The lab level is indirectly addressed by vision assessment. Techno-visionary communication often has impacts on the agenda of science and research, it may attract young people to the respective research fields, it may influence research funding, and it may create public awareness and political support. Researchers at the lab can contribute to developing visions but also to scrutinising and assessing them. Accordingly, in spite of the fact that vision assessment mainly addresses public debates on new and emerging science and technology researchers operating at the lab must be involved in processes of vision management.

2.4.5 Real-Time Technology Assessment and New Approaches

Technology Assessment is a field of research and engagement which requires continuous adaptation to new developments and changing boundary conditions as well as learning procedures. The history of TA can be told as continued exploration of and experimentation with new approaches.

One of the most recent conceptual developments in TA is the so-called Real-time Technology Assessment (Guston and Sarewitz 2002), which is a social technology with the goal of redesigning knowledge production and application to make design and other choices in research and development more explicit, informed, transparent, accountable, and participatory. It combines problem-oriented empirical research and dynamics of technology with research on and engagement with values accompanying these developments. It was nanotechnology which gave rise to the development of this approach.

Another recent development in TA is related to the notion of a “third generation TA” (Yoshizawa 2012). This development identifies a first (expert-based, parliamentary-centred) and a second (involving selected citizens, parliamentary-related) generation of TA. Third generation TA is, in contrast, not necessarily based in an established organisation, “but rather in a flexible distribution network of existing intellectual and human resources, facilitating active engagement of lay public as well as intermediate actors between experts and technology end-users” (Yoshizawa 2012). There are some relations to the idea of Responsible Innovation (see above), but in this case everything is still ongoing.

2.5 Case Study: Systems Analysis for Developing New Cement

In this section, we describe the approaches of accompanying systems analysis and innovation research based on a case study in which we discuss an ongoing technology development at Karlsruhe Institute of Technology (KIT), a project which has become a beacon of the KIT.

Cement is a mass construction material with large world-wide growth rates, whose production is energy-intensive and connected with high CO_2 emissions. Cement, mostly Ordinary Portland Cement (OPC), is mainly used for concrete – next to water the most commonly used material worldwide. According to our assessments, the global cement production emitted in 2008 a total of about 2.5 billion tons CO_2 (Achternbosch et al. 2011a). This corresponds to 8 % of the global anthropogenic CO_2 emissions. The cement industry has adopted many measures to reduce these emissions, such as improving techniques or using secondary materials. However, it can be shown that these measures are not sufficient to stabilise or even lower the amount of future CO_2 emissions. Carbon Capture and Storage (CCS) is currently discussed controversially as a "tool" for reducing CO_2 emissions, but it is likely that this process is very energy-intensive, uneconomic, and possible risks are difficult to assess (Luhmann 2009). Instead of focusing on this end-of-pipe technology, which has not even been realised, the development of "low-CO_2 cements" that are produced with substantially lower CO_2 emissions would be a more sustainable course provided these innovative cements do have the potential to replace conventional cement.

At the Karlsruhe Institute of Technology (KIT; www.kit.edu), cement and concrete have been topics of research for many years. Since 1997, the workgroup "Construction Materials" (BSG) of the Institute of Technical Chemistry (ITC) has been investigating the cement chemistry related to hardened cement stone (calcium silicate hydrates). A major impulse for the development of efficient and novel binders came with the analysis of material flows in cement production carried out by the Institute for Technology Assessment and Systems Analysis (ITAS) of KIT (Achternbosch et al. 2005) in cooperation with the BSG. The research was funded by the Federal Environment Agency from 2000 to 2003. In the year 2006, the activities of four BSG researchers resulted in the development of a new class of low-CO_2 binders, which form the basis of Celitement®.

In the same year, ITAS informed the Executive Board of Directors about this remarkable development and recommended its support and funding. High potential was seen in developing a new cementitious binder, whose production is probably associated with much lower CO_2 emissions than that of conventional cement. ITAS was convinced that the research results could be transformed to industrial scale and that the new cement, in addition to the cleaner production issues, could show improved structural properties – compared to OPC. In particular, due to current and forthcoming climate policy regulations these new binders could have a good chance to enter the market and to revolutionise the construction industry.

The Board followed these recommendations. Since that date, the cement project has received great attention and importance. According to a directive issued by the Board in 2006, the development was treated as strictly confidential for obvious reasons, and up to the issuance of related patents it was not allowed to publish anything, not even the fact that this development was already on the way. This statement was an excellent starting position to protect the interests of the KIT with regard to transferring this technology to industry.

Since 2007, several patents have been issued for Celitement®, concerning both the material composition and the principle of the production process. The (binder) system of Celitement® is similar to that of the conventional OPC. However, there are large differences in the manufacturing process, the potential of energy efficiency, and the amount of CO_2 emissions associated with the production. In order to promote the development of the binder and the process design for commercial use, the Celitement® GmbH was founded in 2009 as a spin-off of the Schwenk Group (industry), the Karlsruhe Institute of Technology KIT, and the originators of the invention.

In the last 5 years, the project "Celitement®" has become a beacon of KIT, and since that time it has been accompanied by systems analysis conducted at ITAS. In the case of a successful market introduction, the Celitement® invention would have great economic potential. Thus it is quite understandable that further research in this field need to be handled with great care by all persons involved: KIT (Executive Board), inventors (BSG), accompanying systems analysis (ITAS), and industry (Schwenk Group). Therefore, all project participants were sworn to secrecy; for ITAS this means that results of the work in the years 2007–2009 could be used only internally and were not allowed to be published. Even the fact that ITAS carries out research to evaluate this project was not revealed until February 2009, when the information was published by KIT Press Release 014/2009 and the website of the spin-off Celitement® GmbH (www.celitement.com).

The work of ITAS in the context of the innovative binder covers various areas. Studies on the energy consumption for the production of Celitement® and the associated CO_2 emissions for evaluating the CO_2 reduction potential of the process are included. The findings are made available to the inventors to foster the development process; on the other hand they are used by ITAS for a systems analytical assessment of the entire process chain. The results are compared with the corresponding data for the production of conventional cements.

Another task of ITAS is to examine the opportunities and potentials of products based on Celitement® for a successful market entry. Closely connected with this work is the analysis of the innovation process. This could be described as a "life analysis", because this time it is not an analysis "afterwards" but an analysis which can react immediately on changes in "parameters", changing basic and edge conditions. To what extent it will be possible to publish the results of this work is difficult to estimate because they are closely linked with the interests of the cement industry, especially Celitement's shareholders.

In the summer of 2011, on the premises of the KIT North Campus, a pilot plant went into operation, which will accelerate the process of developing Celitement® from laboratory scale to technical readiness for the industrial production process. With this pilot plant it is now possible to produce larger amounts of Celitement®, which must be available to evaluate its material properties in detail. The data provided by the pilot plant enable ITAS to make more reliable estimates of energy use and CO_2 savings resulting from the production of Celitement.

The work and resulting recommendations of ITAS are seen only as internal help in decision-making by the other project participants. The ITAS project group, however, has never been actively involved in any of these processes because of the strong economic orientation of this project. The shareholders of the Celitement® GmbH – the inventors, the Schwenk Group and the KIT Board – alone make strategic decisions and take care about their implementation. The pursuit of economic interests with the premise of a fast market launch requires development of strategies and their implementation, which are not ITAS tasks.

Since 2009, ITAS has focused on topics that offer greater academic freedom, but which are nevertheless of crucial importance for the overall project. Parts of this work focus on studies of current and possible future actions of the global cement industry to reduce CO_2 emissions. First results were already published in Achternbosch et al. (2011a) after being "signed off" by the other project participants (ensuring that the confidentiality obligation was complied with). The report focuses on issues of demand, production capacity, and the raw material situation, but also looks at the CO_2 emissions released in the manufacturing process and the changing framework requirements due to the international climate regime.

Another task is extensive research on the development of low-CO_2 cements that are currently discussed in the media as possible alternatives to conventional cement (e.g. Novacem, Calera). These are analysed for their potential as mass building materials and their possibilities for reducing CO_2 emissions. The results of this work can be published without prior consultation with the shareholders of the Celitement® GmbH. The very first publication focused on Novacem (Achternbosch et al. 2011b).

It is already apparent from this brief account that the ITAS project group operates in a field of tension:

- ITAS sees itself as an independent and neutral research and TA institution. This is an important condition of our work, especially in the context of our mandate for policy advice. From the outside, this is even an indispensable prerequisite for the credibility of an institute for technology assessment. It is quite understandable that, from the external view (outside), a very tight involvement in projects with economic interests would put the neutrality of ITAS in question. This is one reason that topics such as public relations, promotion and marketing are not in our responsibility when practising supporting research on technical developments. As a consequence, ITAS is rarely mentioned in press releases or media appearances of Celitement. On the other hand, a "tighter" integration has positive effects on our work, because this allows an in-depth analysis of the innovation process, which can react on actual experiences and information.
- The systems analytical work, particularly on alternative cements, has brought us a lot of attention in the "cement community". We are advisors for associations, the industry partner of the Celitement® GmbH and research institutions of universities and are in contact with the Federal Association of the cement industry.
- The position of the cement industry is often ambivalent towards our studies: If the results do not affect their business and their interests, our work is acknowledged with great appreciation. Sometimes the industry draws important lessons from our work (advisory role), because a lot of information that we provide are based on studies which cannot be carried out in that depth by one

of their own. And in addition, we often confront the cement industry with quite different perspectives.

- It is remarkable that the new low-CO_2 cements (Novacem, Calera, Celitement®) which have gained the attention of the media and the public were not developed in the research departments of the cement industry, but almost exclusively in the "non-industry-oriented" university environment or in corresponding major research institutions. The further development of these inventions is done by start-up companies. The studies ITAS is currently working on do not affect the interests of the cement industry, of university research institutes or major research centres connected with the building industry. From the perspective of the established industries, the results for low-CO_2 cements, which are not always favourable, confirm their strategy – namely their adherence to developments that closely lean on conventional cements.

The construction industry is rather a conservative branch where traditionally only incremental innovations can be implemented. The cement industry looks back to an era of 150 years of successful implementation of Portland cement, a cement system which up to now seems unchangeable in the minds of experts of the industry and institutes for Materials Research of the building industry. Editorial boards of professional journals of Materials Research in the building industry are mostly composed of these experts. These experts are often also authorities in political motivated committees. If systems analytical articles with another system view are submitted for publication in these journals, constraints for publication cannot be excluded due to this conservative view. If rejected, the desired recipients cannot be reached.

Our case study reveals that the systems analysis accompanying the Celitement® project in close relation with lab research is embedded in the field of conflict (technology determinism vs. social constructivism) elucidated in Chap. 3. This is particularly perceptible in respect to the aspect of governance. On the one hand, the attendance of TA is clearly wanted and promoted. On the other, however, the old distrust against TA scientists as technology laggards is ingrained, leading to sometimes complex processes of negotiation and to restrictions concerning publications.

2.6 Conclusion

Technology Assessment looks back on a history of now more than four decades. According to different and partially heterogeneous expectations of what TA should deliver, and because the expectations vary over time, TA has developed several approaches to meet different challenges and match different contexts. Besides the more policy-oriented approaches, several TA concepts deal with the challenge to support technology development in direct contact with developers at the lab level. The background theory of Social Constructivism gives rise to some optimism about the possibility of shaping technology.

However, there are also limitations and obstacles, as well as possible non-intended side-effects if TA works closely with lab research. It is important to point out that the model of TA research intimately linked with technical R&D also harbours problems. Its independence might be threatened, especially if the necessary distance to the technical developments and those working on them is lost. The case study (Sect. 2.5) is an illustrative example of this problem but also shows how to tackle them successfully and continue cooperation between lab development and TA in the case of conflict. To the same extent as TA would become part of the development process and identify itself with the technical success, the suspicion might emerge that possible positive results were "purchased" or that it was nothing but an accelerator in the process of innovation. The credibility of TA – which is essential for it to do its job – would be endangered.

A second critical issue is conflict and freedom of research. In innovation research and development, usually strong economic interests are part of the game. These could lead to conflicts as was also shown in the Celitement® case (Sect. 2.5). If TA came up with unexpected and unfavourable results for the innovation under consideration, e.g. concerning the competitiveness or sustainability indicators, voices might come up to suppress these results. Frequently, free publication is restricted by confidentiality agreements (see the case study in Sect. 2.5) which would allow preventing the publication of unwanted TA results. In this case, the task of TA would be to convince the partners that negative results for the innovation considered should not be suppressed, but taken seriously and used to re-design and improve it. Indeed, negative or unexpected results can often be interpreted as recommendation for change which could improve the chances of the innovation under development at the marketplace. Anyway, simultaneously sustaining the independence of TA and its relevance to R&D in concrete research and innovation processes requires balancing the distance of an observer to the neighbourhood of an involved person, which is an ambitious and delicate task.

References

Achternbosch, M., Bräutigam, K.-R., Hartlieb, N., Kupsch, C., Richers, U., & Stemmermann, P. (2005). Impact of the use of waste on trace element concentrations in cement and concrete. *Waste Management and Research, 23*(2005), 328–337.

Achternbosch, M., Kupsch, C., Nieke, E., & Sardemann, G. (2011a). Klimaschonende Produktion von Zement: eine Utopie? *GAIA, 20*(1), 31–40.

Achternbosch, M., Kupsch, Chr., Nieke, E., & Sardemann, G. (2011b). *Sind "Green Cements" die Zukunft? Erste systemanalytische Abschätzungen zu innovativen Bindemitteln. Teil 1: Novacem*. Karlsruhe: KIT Scientific Publishing. (KIT Scientific Reports 7589).

Bechmann, G., Decker, M., Fiedeler, U., & Krings, B.-J. (2007). Technology assessment in a complex world. *International Journal on Foresight and Innovation Policy, 3*, 6–27.

Bijker, W. E., & Law, J. (Eds.). (1994). *Shaping technology/building society*. Cambridge, MA: MIT Press.

Bijker, W. E., Hughes, T. P., & Pinch, T. J. (Eds.). (1987). *The social construction of technological systems. New directions in the sociology and history of technological systems*. Cambridge, MA: MIT Press.

Bimber, B. A. (1996). *The politics of expertise in congress: The rise and fall of the office of technology assessment*. New York: State University of New York Press.

Collingridge, D. (1980). *The social control of technology*. New York: St. Martin's Press.

Cruz-Castro, L., & Sanz-Menendez, L. (2004). Politics and institutions: European parliamentary technology assessment. *Technological Forecasting and Social Change, 27*, 79–96.

Decker, M., & Ladikas, M. (Eds.). (2004). *Bridges between science, society and policy. Technology assessment – methods and impacts*. Berlin: Springer.

Funtowicz, S., & Ravetz, J. (1993). The emergence of post-normal science. In R. von Schomberg (Ed.), *Science, politics and morality*. London: Sage.

Grin, J., & Grunwald, A. (Eds.). (2000). *Vision assessment: Shaping technology in 21st century society*. Berlin/Heidelberg/New York: Springer.

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

Grunwald, A. (2009a). Technology assessment: Concepts and methods. In A. Meijers (Ed.), *Philosophy of technology and engineering sciences* (Vol. 9, pp. 1103–1146). Amsterdam: Elsevier.

Grunwald, A. (2009b). Vision assessment supporting the governance of knowledge – the case of futuristic nanotechnology. In G. Bechmann, V. Gorokhov, & N. Stehr (Eds.), *The social integration of science. Institutional and epistemological aspects of the transformation of knowledge in modern society* (pp. 147–170). Berlin: Edition Sigma.

Grunwald, A. (2010a). *Technikfolgenabschätzung – eine Einführung*. Berlin: Edition Sigma.

Grunwald, A. (2010b). From speculative nanoethics to explorative philosophy of nanotechnology. *NanoEthics, 4*(2), 91–101.

Grunwald, A. (2012). Technology assessment for responsible innovation. In *Proceedings of the responsible innovation conference*, The Hague, Apr 2011 (to appear).

Guston, D. H., & Sarewitz, D. (2002). Real-time technology assessment. *Technology in Culture, 24*, 93–109.

Habermas, J. (1970). *Toward a rational society*. Boston: Beacon Press. First publication: Habermas, J. (Ed.). (1968). Technik und Wissenschaft als Ideologie. Frankfurt: Suhrkamp.

Joss, S., & Belucci, S. (Eds.). (2002). *Participatory technology assessment – European perspectives*. London: University of Westminster.

Luhmann, H. J. (2009). CO_2-Abscheidung und -Lagerung bei Kohlekraftwerken: kein Beitrag zur Lösung des Klimaproblems. *GAIA, 18*(4), 294–299.

National Research Council. (2006). *A matter of size: Triennial review of the national nanotechnology initiative*. Washington, DC: National Academies Press.

Rip, A. (2007). Die Verzahnung von technologischen und sozialen Determinismen und die Ambivalenzen von Handlungsträgerschaft im 'Constructive Technology Assessment'. In U. Dolata & R. Werle (Eds.), *Gesellschaft und die Macht der Technik. Sozioökonomischer und institutioneller Wandel durch Technisierung* (pp. 83–106). Frankfurt/New York: Campus Verlag.

Rip, A., & Swierstra, T. (2007). Nano-ethics as NEST-ethics: Patterns of moral argumentation about new and emerging science and technology. *NanoEthics, 1*, 3–20.

Rip, A., Misa, T., & Schot, J. (Eds.). (1995). *Managing technology in society*. London: Pinter Publishers.

Schepelmann, P., Ritthoff, M., Jeswani, H., Azapagic, A., & Suomalainen, K. (2009). *Options for deepening and broadening LCA*. CALCAS – Co-ordination Action for innovation in Life-Cycle Analysis for Sustainability. Brussels et al. European Commission.

Siune, K., Markus, E., Calloni, M., Felt, U., Gorski, A., Grunwald, A., Rip, A., Semir, V. de, & Wyatt, S. (2009). *Challenging futures of science in society* (Report of the MASIS Expert Group). Brussels.

Van de Poel, I. (2009). Values in engineering design. In A. Meijers (Ed.), *Philosophy of technology and engineering sciences* (Vol. 9, pp. 973–1006). Amsterdam: Elsevier.

VDI – Verein Deutscher Ingenieure (1991). Richtlinie 3780 Technikbewertung, Begriffe und Grundlagen. Düsseldorf. Available also in English at www.vdi.de

Von Schomberg, R. (Ed.). (1999). *Democratizing technology. Theory and practice of a deliberative technology policy*. Hengelo: ICHPA.

Von Schomberg, R. (2012). Prospects for technology assessment in the 21st century: The quest for the “right” impacts of science and technology. An outlook towards a framework for responsible research and innovation. To appear In M. Dusseldorp, et al. (Eds.), *Technikfolgen abschätzen lehren*. Wiesbaden: Springer VS.

Voss, J.-P., Bauknecht, D., & Kemp, R. (Eds.). (2006). *Reflexive governance for sustainable development*. Cheltenham: Edward Elgar.

Yoshinaka, Y., Clausen, C., & Hansen, A. (2003). The social shaping of technology: A new space for politics? In A. Grunwald (Ed.), *Technikgestaltung zwischen Wunsch oder Wirklichkeit* (pp. 117–138). Berlin: Springer.

Yoshizawa, G. (2012). *Third generation of technology assessment*. Paper presented at the annual meeting of the 4S Annual Meeting – Abstract and Session Submissions, Komaba I Campus, Tokyo, Japan. http://www.allacademic.com/meta/p422007_index.html. Download 15 July 2012.

Part II
Approaches to Early Engagement

Chapter 3
Constructive Technology Assessment and the Methodology of Insertion

Arie Rip and Douglas K.R. Robinson

Abstract Constructive Technology Assessment (CTA) started out (in the Netherlands in the late 1980s) as an attempt to broaden technology developments by including more aspects and more actors, and has been further positioned as a way to overcome the institutionalised division of labour between promotion and control of technology. For newly emerging technologies like nanotechnology, which live on promises, CTA has to address uncertain futures. It does so by analysing dynamics and emerging irreversibilities in a technology domain, identifying "endogenous futures" and creating socio-technical scenarios exploring what could happen. Such scenarios are a platform for interaction between stakeholders in strategy-articulation workshops. Organizing such workshops by CTA agents constitutes a soft intervention in ongoing developments, and contributes to make ongoing co-evolution of science, technology and society more reflexive. The CTA analyst inserts herself in ongoing developments in the domain that is being addressed, to identify what is at stake. This is not just data collection, but already interaction, as a knowledgeable visitor. Such a role has to be earned, for example by offering useful (but also critical) insights based on circulation in the domain and social-science analysis. This constitutes a methodology of inquiry-in-interaction, which increases reflexivity of the developments. It is an essential part of the CTA enterprise.

A. Rip (✉)
Department of Science, Technology and Policy Studies, School
of Management and Governance, University of Twente,
P.O. Box 217, Enschede 7500AE, The Netherlands
e-mail: a.rip@utwente.nl

D.K.R. Robinson
TEQNODE Limited, 282 Rue Saint Jacques, Paris 75005, France

IFRIS-LATTS, Université de Paris-Est, Paris, France
e-mail: douglas.robinson@teqnode.com

N. Doorn et al. (eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*,
Philosophy of Engineering and Technology 16, DOI 10.1007/978-94-007-7844-3_3,

3.1 Introduction

The two key elements of Constructive Technology Assessment (CTA), broadening technology development by including more aspects and involving more actors, and doing so on the basis of an understanding of the dynamics of technology development and its embedding in society, were identified in the mid/late 1980s in the Netherlands (Schot and Rip 1997). It was part of a larger perspective, laid down in the government's Policy Memorandum on Integration of Science and Technology in Society (Ministerie van Onderwijs en Wetenschappen 1984). On the basis of the Policy Memorandum, a Netherlands Organization for Technology Assessment (now Rathenau Institute) was established in 1986. One of its projects was to develop the approach of Constructive Technology Assessment (Daey Ouwens et al. 1987). In the Ministry of Education and Sciences and in the Netherlands Organization for Technology Assessment, perspectives and expertise from Science, Technology and Society studies played an important role. The further development of CTA occurred in STS studies, linked to evolutionary economics of technological change (Rip et al. 1995), and in the evaluation of attempts to broaden technology development, as in social experiments with electric vehicles (Hoogma 2000; Hoogma et al. 2002). The CTA approach was taken up in studies in Canada, the UK, Australia, Denmark and Sweden. And it was positioned as part of an overall move towards more reflexive co-evolution of science, technology and society (Rip 2002).

Newly emerging technologies like nanotechnology, with their promises but also raising concerns about possible negative impacts, are a challenge for the CTA approach because the envisioned broadening of technology development must now be about possible future developments rather than current practices. Such a challenge had been recognized before, and could then be addressed systematically from the early 2000s onwards when the Dutch national R&D program NanoNed, on nanoscience and nanotechnology, wanted to have a Technology Assessment (TA) component, and made funding available for PhD students and postdocs. The findings of this TA NanoNed program are the basis for this chapter, located in the larger picture of reflexive co-evolution of science, technology and society.

CTA is a "soft" intervention, and studies and reports are an input, not the main result. For emerging technologies, two key components of a CTA activity are (1) the building of sociotechnical scenarios of possible technological developments and the vicissitudes of their embedding in society (based on extensive document study and field work) and (2) the organizing and orchestration of workshops with a broad variety of stakeholders. The scenarios help to structure the discussion in the workshops (Robinson 2010) and stimulate learning about possible strategies (Parandian 2012). Therefore it is important to have scenarios of high quality and relevance, and which can be seen as legitimate by workshop participants.

Compared with other approaches as discussed in this volume, CTA activities take into account what happens on a variety of "work floors": research laboratories, conferences, workshops, agenda setting and planning meetings, roadmapping events, public debates anticipating on issues related to technology developments. A corollary is that the CTA actor has to move about, observe and actively circulate in locations where actors are shaping the emerging paths of nanotechnology and how it will become embedded in society. We will call this 'insertion' by the CTA actor, to emphasize it is not just a practical matter of collecting data, but also part of the methodology of CTA, combining diagnosis of dynamics and some soft intervention.

3.2 The Enterprise of CTA: Goals and Practices

While explicit goals for CTA were specified already in the 1980s, the actual approaches were also shaped by opportunities and circumstances that arose following its inception. Based on the experiences, there was further articulation of goals. This section is an attempt to take stock, by looking at overall goals, how these are linked to more concrete objectives, particularly for the case of emerging technologies, and what sort of concrete activities and methodologies are now in place.

CTA sees itself as part of the overall undertaking of TA, starting in the late 1960s. The background of this undertaking can be formulated, in retrospect, as a 'philosophy' of TA (cf. Rip 2001a):

> Reduce the (human) costs of learning by trial-and-error -- which characterized much of our handling of technology in society –, and do so by anticipating future developments and their impacts, and by accommodating these insights in decision making and implementation.

This is not easy because early signalling may not get a hearing – particularly if it is early warning (cf. Harremoës 2001). And it is not limited to commissioned TA studies. It is a societal learning process, in which many actors participate. Actually, over the years, TA has moved in the direction of societal debate and agenda-building, at least with Rathenau Institute and some other European TA offices (Delvenne 2011).

Within TA, some of the specifics of CTA derive from a diagnosis of how the handling of technology in society has evolved: the separation of "promotion" and "control" of technology in our societies, which emerged in the nineteenth century and are still with us (Rip et al. 1995). It is a heritage of the industrial revolution of the eighteenth and nineteenth centuries, where technology development became a separate activity, carried by engineers and located in firms and public or semi-public research institutes. Culturally, a mandate to do so emerged: new technologies could then be developed as such, because they could be positioned as contributing to progress of society, and therefore to be accepted, almost by definition. Institutionally, an indication of the separation between "promotion" and "control" is the division of labour between government ministries, some promoting the development of new technologies and innovation, while other ministries consider impacts and regulation. TA emerged within this regime of handling technology in society, and was institutionalized at the "control" side of the division of labour. An important argument was (and is) the asymmetry between technology development actors and society at large, with the latter coming in at a late stage, and little information about the technology. The asymmetry is structural, but TA would offer information and considerations to the "control" side, and reduce the asymmetry.[1] CTA wants to compensate for the asymmetry in TA approaches, by focusing on technology development.[2]

Building on this diagnosis, CTA aims to bridge the gap between innovation and the consideration of social aspects which inform attempts at "control", and in doing so, broaden technology development and its embedding in society. It is "constructive" TA because it aims to be part of the construction of new technologies and their embedding in society. This was the starting point of the enterprise of Constructive TA (Daey Ouwens et al. 1987). These aims can then be taken as objectives for the design and execution of CTA activities. They require analysis of dynamics of technology development and its embedding in society, and the ways it is influenced/shaped – insights which can be translated into leverage for change. They are input into the preparation for concrete CTA activities like "bridging" workshops with stakeholders in a technology domain. They are also building blocks for a theory of CTA (Rip 1992).

The rationale for pursuing these objectives stems from larger goals and perspectives, as was clear in how we developed a diagnosis of what is the case now in handling technology in society, with the implication that it should be improved. By now, a number of overlapping goals have been put forward. Taking an evolutionary perspective, the division of labour between "promotion" and "control" of technology in society is part of how technology and society co-evolved. One can then take a step back, and consider ongoing co-evolution of science, technology and society, and in particular, how it is becoming more reflexive, for example through technology policy, technology foresight and technology assessment (Rip 2002). Thus, one can work towards improving reflexivity of the co-evolution, in various ways – this implies some modulation of the co-evolution. This qualifies as a background goal for CTA and is linked to learning (cf. also Grin and Van de Graaf 1996). It has been emphasized in the studies in the TA NanoNed program (e.g. Robinson 2010; Parandian 2012). Then, constructive in CTA refers to its being part of the construction of increased reflexivity in science, technology and society.[3]

Broadening technology development and increasing reflexivity serve a purpose. To be explicit about this, Schot and Rip (1997) emphasized an overall goal served by CTA, of a better technology in a better society. It is important to keep such a substantive goal visible, in general but also because the CTA objective of including more actors is often taken as advocating more participation, and thus refer to a goal

---

[1] This then led technology developers to see TA as "technology harassment".

[2] We note that there is another tradition of TA, in firms and research institutes, where technological options are assessed as to eventual performance and production possibilities and costs. This can be called "technical" TA, to distinguish it from the "public" TA that we discussed here (Rip 2001a). When broader considerations would be taken into account, "technical" TA would become "sociotechnical" TA, and the tools of CTA (see below) could be used by the firms and research institutes, or by consultancies that are commissioned to do "sociotechnical" TA.

[3] Note that 'reflexivity' here refers to institutions and approaches in society and sectors in society, not to individuals becoming more reflective – even while that is part of overall reflexivity.

of democratization of technological development (Genus 2006; cf. also Callon et al. 2001 for an intermediate position). Of course, no one has a monopoly on goals for CTA. The point is that recognition of a goal has implications for what are appropriate CTA activities. The activities we describe in this chapter are appropriate to the overlapping goals we have outlined, so it is inappropriate to criticize them as being insufficiently democratic.

3.2.1 Signs of Change

An increase in reflexivity of co-evolution of science, technology and society is visible in the recent policy discourse about responsible development of new technology, and responsible innovation. There are now some attempts to implement this, especially in the domain of nanotechnologies. One example is the Code of Conduct for Responsible Nanosciences and Nanotechnologies Research (European Commission 2008), which can now be referred to in the Member States of the European Union. There is overlap with CTA objectives, in the sense that responsible development is a way to bridge promotion and control, by internalizing control at the side of technology development. This can still keep a focus on promotion, when 'responsible' is only modifying 'development'. When 'responsible' is emphasized the development itself might be queried, up to the possibility of stopping it.[4]

Thus, there are signs that the institutional separation of technology development and attempts at control (because of projected societal impact), is being bridged. At least, there are pressures to bridge and various attempts at handling these pressures. Of course, there were such pressures before, as when TA was proposed and started to become institutionalized in the 1970s. What is new is that anticipation on societal impacts is now seen as being also a responsibility of technology developers (see also Gustin and Sarewitz 2002).

While the dichotomies (innovation vs. responsible, technology developers vs. users) remain visible, there are interactions and mixed approaches, and the situation evolves further. The domain of nanoscience and nanotechnologies turns out to be a site for experimentation and learning – including controversy. There is widespread uncertainty about impacts and risks, while there are also proposals for regulation, and NGOs which advocate a precautionary approach. There is additional uncertainty about consumer and citizen reactions to new nanotechnology-enabled products and processes, and innovators can fear for barriers to public acceptance and possibly a public backlash if something would go wrong. All this is to be expected.

---

[4] A well-known precedent is the temporary moratorium on recombinant DNA research, after the 1974 Asilomar meeting. The present call for a moratorium on nano-particle development comes from critical outsiders, not from nanoscientists. A mixed case (early 2012) is the voluntary stop (for 60 days) of bird flu virus research, after the US National Science Advisory Board on Biosecurity had required a virology research group in Erasmus University Rotterdam to take out details in their pending publication in *Science*, because of the risk of misuse.

What is new is that innovation actors are asked by societal actors to account for what they do. This will set articulation processes in motion.[5] When some stabilization occurs, there will be *de facto* governance, i.e. steering and shaping of action that has some legitimity, even if there is no formal authoritative basis as in law and regulation (Rip 2010b). Up to a modification of the division of labour, with responsible innovation becoming the responsibility of innovation actors, in interaction with various societal actors.

The experimentation and mutual learning that occurs in and around nanotechnology is now taken up for other emerging technologies like synthetic biology and ambitious technological ventures like geo-engineering. Thus, one can take learning in sectors and in society as a further overall goal, and formulate stimulation of such learning as a broad objective for CTA.[6] For new technologies, the point has been made that responsibilities are distributed, just like technological development itself (Von Schomberg 2007). The simple contrast between technology developers and users is inapplicable then. Interaction and mutual learning become important to overcome mismatches and fragmentation, in innovation as well as in 'distributed responsible development'. New 'divisions of moral labour' have to be invented, and one can see various actors exploring (even if reluctantly) possibilities (Rip and Shelley-Egan 2010).

3.2.2 Transforming Objectives into Activities

In the move from objectives to concrete activities, particularly for doing CTA about new technologies, some further conceptualizations are introduced – in effect, more building blocks for a theory of CTA.

Our diagnosis of a gap between promotion and control of technology at the societal level, and as we phrased it in the TA NanoNed program, the gap between innovation and ELSA in a sociotechnical domain or sector,[7] can be detailed further, to the level of interactions, using Garud and Ahlstrom (1997). They distinguish "insiders" (i.e. developers/promoters) and "outsiders" (i.e. users/regulators) and show that their evaluations of technology are structurally different because of this difference in position. They also consider situations where insiders and outsiders interact, to some extent, calling these situations 'bridging events'. One of the

---

[5] Perspectives, expectations, preferences and positions of various actors/stakeholders will be articulated, i.e. become more explicit, further specified and linked to arguments, findings and values, in interaction and this may lead to scrutiny and assessment.

[6] This is particularly important when the focus is on embedding of technology in society (including further sociotechnical development). This is how Hoogma and Schot evaluated social experiments with electric vehicles (Hoogma 2000; Hoogma et al. 2002, see also Schot and Rip 1997).

[7] Ethical, Legal and Societal Aspects, the "Aspects" are sometimes referred to as Issues (then the acronym becomes ELSI).

examples they study are hearings conducted by a regulatory agency like the US Food and Drug Administration.

Their terminology of insiders and outsiders captures one aspect of the positions with respect to technology development, but assumes these positions are given. However, a firm developing technology for new products or processes of its own, may also be a user of products supplied by another firm and then position itself as an outsider, e.g. requiring quality assurance. When Garud and Ahlstrom (1997) discuss the difference in perspective between insiders and outsiders, they speak of "enactment" and "selection" cycles, respectively, in which the two function. "Enactment", a term from symbolic interactionism, here refers to technology developers and promoters working to realize their goal and vision, "enacting" their project. Thus, a functional terminology is possible, of "enactors" who realize the technology and identify with the project of doing so, and "comparative selectors" who can consider different options to select from and do formal or informal versions of cost-risk-benefit assessment (Rip 2006).[8] Garud and Ahlstrom show how enactors focus on their projections (i.e. informal scenarios) for further development of the technology and its embedding in society, and thus see society as a constellation of possible barriers which have to be overcome. If questions are raised about the technology, such an enactor perspective will immediately see them as indications of potential barriers, even when the questions are mainly inquiry rather than criticism. The response of the enactor then is to emphasize the promise of the new technology – with the corollary that the commentators, if still reluctant, are positioned as being against progress. If this happens in the public domain, it will incite further, and possibly more critical, responses (Swierstra and Rip 2007).

One concrete implication of this diagnosis of the two positions and related perspectives is that CTA workshops must have 'enactors' as well as 'comparative selectors' as participants, so as to function as bridging events, where participants can (in Garud and Ahlstrom's felicitous phrase) probe each other's realities. With the right mix of participants, what happens in these CTA workshops will reflect dynamics in the wider world, so they will be like a micro-cosmos. The workshop is also a protected space, where participants have the opportunity to consider alternatives and the possibility of modifying their strategies and eventual interactions in the real world without there being immediate repercussions.[9] Still, the wider world has

---

[8] The term "enactor" can be used for all cases where a project is pursued, and identification occurs so that the world is seen in terms of whether it helps or hinders the project. An actor can be enactor in one case, and comparative selector in another case. An interesting example is the NGO Greenpeace, almost by definition an outsider/comparative selector. But Greenpeace Germany, at one moment, pushed for an environmentally-friendly fridge, and collaborated with scientists and a firm to realize it (Van de Poel 1998: 84–97). So it became an enactor, for the time being.

[9] This is often a novel possibility for participants. Moving beyond their own interests and perspectives comes easier to some than others, but it is recognized as a possibility in post-workshop interviews with participants (Parandian 2012). The set-up of a CTA workshop has to facilitate and stimulate this, by making sure various actor perspectives are visible, and possible developments in the real world are considered, for example with the help of sociotechnical scenarios.

its own dynamics, and these are important for eventual uptake and effect of the CTA exercise.[10]

There is a further implication, given that we decided to develop sociotechnical scenarios as an input into the CTA workshops. Scenarios speak to an enactor perspective, in their projection of further development of a new technology. But we introduce twists, showing unexpected shifts (for enactors) and repercussions. Stakeholders representing comparative selectors, from potential users to regulators and NGOs, will be present in the workshops. Thus, in the interactions, different perspectives as visible in the scenarios will come alive because their protagonists are present. This will work out well only if the scenarios reflect what is at stake in the worlds of the participants, otherwise they will be disregarded as irrelevant. At the same time, the scenarios must offer challenges to participants' understanding of the situation. This is where social-science insights (from innovation studies, from STS, and more generally) will have to come in, to improve the quality of mutual probing in the workshops.[11]

In general, analysis and diagnosis of developments are necessary steps to prepare a CTA exercise and orchestrate it productively. One has to know about the forces at play in the technology domain and the evolving relationships (or lack of relationships) between stakeholders. A key point for understanding what happens as well as the eventual construction of scenarios is that "entanglements" occur, existing and emerging mutual dependencies which guide and thus limit interactions and strategic choices (Rip 2010c). This shapes the way new technologies (in our cases, nanotechnologies) will materialize. In other words, the future is predicated on these patterns and dynamics: an "endogenous future" (Rip and Te Kulve 2008; Robinson 2009). The scenarios develop the endogenous future into a number of possible futures, each starting with certain interventions and interactions and then exploring responses, repercussions, and eventual outcomes.

For example, in the case of possible nanotechnology applications in food packaging, studied by Te Kulve (2011), there is reluctance with the producers and retailers to invest in it because of uncertainty about consumer acceptance, combined with uncertainty about eventual regulation of the products. The mutual dependencies have the form of a waiting game (Parandian et al. 2012), and if nothing happens, the waiting game will continue (thus, an endogenous future). Given this diagnosis, one can imagine that interventions occur attempting to break through the waiting game.

---

[10] Marris et al. (2008) have shown this for an Interactive TA exercise about field tests of genetically modified vines in France. Their point is reinforced by what happened subsequently: productive co-construction of the design of the field tests between local stakeholders and researchers, and 5 years later, August 2010, the destruction of the test fields by critics of GMO. In LMC et al. (2010), the story is told from the perspective of the actors involved in the co-construction.

[11] Scenarios add substance to the interactions, which is necessary because they are not just about participation and empowerment (which are sometimes taken as goals for CTA, cf. earlier comments on democracy). To serve the change aim of CTA, they must be seen as relevant as well as challenging to the participants. Quite some effort has to be put into the creation of robust sociotechnical scenarios. Thus, they become a product in their own right, which can be put to further use, also by participants.

This was the starting point for the construction of three scenarios. In scenario 1, "Only a little nano", collaborations between academic and industrial researchers are sought and supported, but that leads to niche applications only. The big promise of nanotechnology is backgrounded. In scenario 2, "Regulation helps", the concerns about health and safety aspects cast a shadow over the developments, and small companies move away from working on nano-applications, also because regulation might be strict (and thus make product development expensive). The big incumbents welcome regulation because it reduces the uncertainties, and they proceed – cautiously. In scenario 3, "Thresholds are passed", some institutional entrepreneurs recognizing the barriers set up a consortium for product development and persuade consumer organizations and risk research institutes to participate, arguing that this is a way for them to have some influence on the shape of future technology. This creates legitimacy and further support becomes available for strategic research topics like nano-enabled improvement of barrier properties of paper and plastic packaging. Pharmaceutical companies then become interested as well.

3.2.3 Choices to Be Made

As is clear from this example, in constructing scenarios choices must be made about what to focus on, and what not. These choices can be discussed in the workshop, and alternatives may be considered. In general, the need to make choices in setting up the CTA activity is a challenge (and a task) for the CTA analyst, especially for emerging technologies like nanotechnology which live on promises: Which expectations are to be taken into account as more realistic and/or more important? What is seen as important also depends, of course, on the position from which such expectations are voiced, e.g. by an enactor or a comparative selector. The CTA analyst can build on her knowledge of the domain and its dynamics, including expectations and investments in the different worlds in which a new technology option is being developed and will be embedded. But the challenge remains.

The challenge can be brought out (even if in a somewhat simplified manner) by considering the hype-disappointment cycle, as introduced by Gartner Inc. Figure 3.1 shows the cycle, as well as different options for projecting a future state of the world. The realistic option (the eventual "plateau of productivity") is also the most uncertain one, while relying on present promises may risk becoming victim of inflated expectations.

The risk is real, and not only in funding applications and other resource mobilisation activities, where exaggerated promises are expected, and discounted. In discussions and activities exploring potential futures of a technology and its ethical and societal impacts, there is a tendency to go for the big impacts, so as to justify the effort to anticipate. It is all too easy then to extrapolate from current promises and end up in brave new worlds where human enhancement or interventionist ambient intelligence creates interesting ethical dilemmas. Nordmann and Rip (2009) have criticized such "speculative" ethics of new technologies as disconnected from

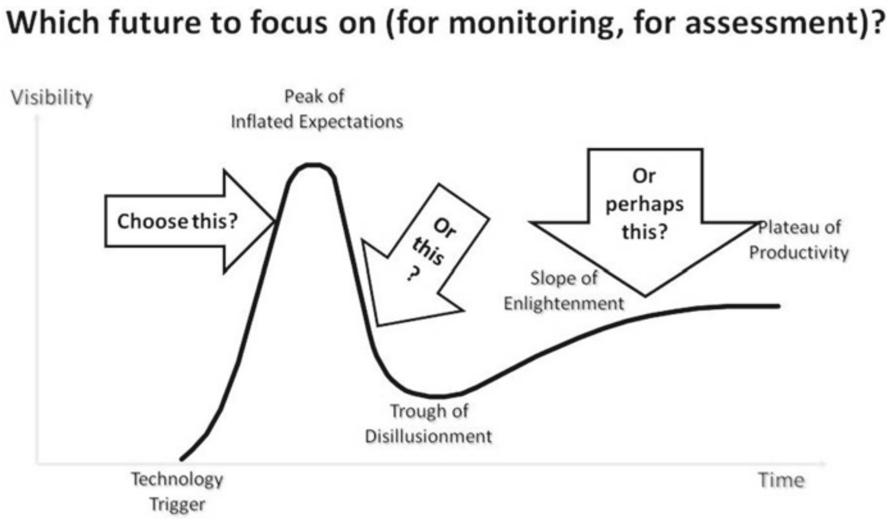


Fig. 3.1 Gartner Group's hype-disappointment cycle (Versions of the hype-cycle were presented by Gartner Group since at least 1999, see Fenn 1999)

ongoing activities and the choices, ethical and otherwise, that have to be made there. Our sociotechnical scenarios, building on endogenous futures, start from the other side. There is still speculation and imagination, of course, but it is not free-floating.[12]

For actors articulating their strategies the question of hype is a recurrent concern. Interaction with other relevant actors is important to reduce uncertainty, and in fact, the CTA workshops offer an opportunity to do so, and are appreciated for it. This was clear in the domain of Organic Large Area Electronics, studied by Parandian (2012). In one of his scenarios, he actually used the phenomenon of hype and disappointment, for nano-enabled RFID applications for security. This induced extended consideration of the value of government measures to realize the promises of a new technology.

So far, we have presented the CTA activities as doing a good job. And indeed they do, but some reflection is in order. CTA for new technologies aims to broaden design and development, at an early stage. Thus, it has an upstream bias: better outcomes result from doing better at an earlier stage. It is a bias, because it is the overall co-production process that leads to eventual outcomes, there is no determinism. But it

---

[12] The emphasis on choices in ongoing developments is also important to counter the opposite position, that there is no way to predict future impacts of a technology, so better give up on technology assessment and other attempts at anticipation and feedback. This "hard truth" was pushed by Nathan Rosenberg in an OECD workshop on Social Sciences and Innovation (Tokyo, 2000), but it overlooks how present dynamics shape opportunities and constraints for future developments, and are thus a basis for anticipation and feedback (Rip 2001b). The further point is that anticipations need not be correct to be useful in guiding action – think of self-negating prophecies.

is an unavoidable bias if one wants to address new technologies – which are by definition still at an early stage.

Upstream public engagement (in the UK and elsewhere) has the same bias, but in contrast with CTA it focuses on actors with little or no agency. They may well remain empty exercises, even if the views and discussions reported might be taken up by policy makers when they see fit. CTA addresses stakeholders, and does more than just soliciting views from stakeholders. There are orchestrated bridging events, and there must be something at stake, for the participants and developments in the domain or sector. Looking back at the almost twenty CTA workshops we organized in the TA NanoNed program, we see that the less successful ones indeed suffered from there being little at stake (Robinson 2010).

3.3 A Methodology of Insertion

The aims of CTA to broaden technological design and development and make it more reflexive, imply an action-orientation of CTA. CTA agents are change agents, but softly, through support and attempts at opening up, rather than pushing. If there is pushing, it is a push for more reflexivity (cf. Schot and Rip 1997). Theoretically and practically, this relates to the rationale of making the co-evolution of science, technology and society more reflexive (so there will be some modulation of the co-evolution).

What happens in practice is that a CTA exercise, like the strategy-articulation workshops we discussed, is inserted in ongoing developments and interactions, often with support of one or more of the actors involved, for example the EU Network of Excellence Frontiers,[13] which is important to create some legitimacy for the exercise. In preparing the exercise, the organizer (CTA analyst/agent) moves about in the relevant worlds, finding out about "entanglements", forces at play, and stakes, and using those insights to prepare for the workshop and orchestrate it. When moving about, it is the CTA analyst (as a social scientist) who inserts herself in these worlds. But in doing so, she leaves traces and thus creates small changes: the CTA analyst is already a CTA agent.

Becoming an agent in this way is not just a circumstance that requires some methodological reflection. It is actually a methodology in its own right, a methodology of insertion. Our recognizing it as a methodology emerged gradually over time. It started with the notion that the analyst moving about makes patterns in the co-evolution of technology and society visible, and thus creates some reflexivity. We learned by doing, also building on some general insights. Robinson (2010) devoted a chapter in his PhD thesis to describe his "insertions" and their outcomes, from the perspective of a methodology in the making.

---

[13] This network of nanotechnology research institutions focused on the development of nanotechnology instrumentation and approaches for the life sciences (see Robinson 2010).

Table 3.1 Multi-layered landscape of insertion

The **top layer has** broad activities related to public policy, regulation and societal debate. This includes overall institutions, arrangements and authorities in our society.

The **middle layer** is located in collectives of actors, relevant institutions and networks that are directly involved in nanotechnology development through coordination and agenda setting.

The **bottom layer** represents ongoing practices and projects (often shaped by enactment cycles). For nanotechnology these may occur in publicly funded research laboratories, universities, and large or small firms.

The recent interest in "integration" or "immersion" of social scientists and humanities scholars in the work on the lab floor can be seen as having a similar thrust, and has sometimes been developed as a methodology.[14] The important difference is that "insertion", as we use the term here, happens at a variety of "work floors", it happens in a multi-layered landscape and addresses the layers explicitly. Table 3.1 indicates the layers.

In the lab floor studies, the bottom layer is what is focused on, but the other layers are still there, and shape what happens on the lab floor.

What does the methodology of insertion consist of? We will indicate steps, but what we mostly do is report on our learning by doing, offering some evaluations and further perspectives. The first step is 'moving about' in the world of nanotechnology. In particular, visiting locations of nanotechnology R&D, conferences and other meetings, and tracing anticipatory coordinating activities like roadmaps and European Technology Platform meetings where nanotechnology developments are being shaped. Interactions occur, and the CTA analyst & agent-to-be should be willing to enter into the substance of the developments and concerns so as to be a legitimate partner.[15] The CTA analyst must be recognized as a knowledgeable visitor, and this constitutes the second step of the methodology, the actual 'insertion' in the world of nanotechnology. Insertion is the process of becoming a temporary member of the field, a legitimate visitor. But the inserted CTA analyst should not go native, and make sure she is recognized as a visitor and not a full member.

Moving about helps to capture what is going on, and thus to target, tailor and embed CTA exercises. CTA exercises must embed themselves, and thus fit to evolving circumstances in order to be accepted as legitimate/plausible. But there must also be some stretching of these circumstances so as to broaden enactment processes and stimulate reflexive learning. In other words, the visitor moving about is doing more than sightseeing. Fitting and stretching requires deep knowledge of dynamics and contexts. Along with the rapidly evolving developments in and around nanotechnology such knowledge can only be garnered by insertion. This is more than an anthropologist, also a visitor by definition, would do. The CTA analyst

---

[14] In particular in the Socio-Technical Integration Research (STIR) project, funded by the US National Science Foundation and led by Erik Fisher (Arizona State University). See Schuurbiers and Fisher (2009).

[15] So this is more than participant observation, or anthropologists alternating between insider and outsider positions.

moving about in the nano-world is also formulating diagnoses about what is happening and could happen.

Insertion into the world of nanotechnology development requires the active circulation of the analyst in locations were actors are shaping the emerging paths of nanotechnology R&D. This includes research laboratories, conferences, workshops, agenda setting meetings, roadmapping events, and public debates anticipating on issues related to technology developments. As a knowledgeable visitor, and based on her diagnoses of the situation, the CTA analyst can actively probe views and interactions, so as to find out about the forces at play. This will be done in preparation for a CTA exercise, but the insertion can continue over the course of a few years, so that changes over time can be traced. This is what Robinson did, within the European Network of Excellence Frontiers, and more broadly. His role evolved from 'foreigner' to 'regular': his activities became gradually accepted, visible, and in some circles, legitimate.

Important in these activities were aggregation of what was happening in the nanoworld, and analyzing it, creating an overall picture, and presenting it if only in conversation with members of the nanoworld. This functions as an entry ticket ("see, I am inserted and knowledgeable") and a way of getting feedback. But there will be the danger of being positioned as part of the nanoworld, so being pressed to go native, or positioned in a service role to the nanoworld which limits the freedom of movement of the analyst. Thus, there is further requirement: play a distinct role in the nanoworld and make sure it is seen as distinct. This role of a (welcome) visitor can be highlighted by moving in and moving out of the nanoworld. The possibility to refer to own social-science publications which could be helpful to nanoscientists and nanotechnologists, (for example, Robinson and Propp 2008) turned out to be a good way to create legitimity. Given the vicissitudes of insertion, including working under time pressure, there will be lots of contingencies. So there will be no simple recipes.

As to overall changes, there is a clear difference between 2004 when the CTA projects started and nano-scientists looked dubiously at the intruders, and the present situation in which social scientists and other non-technical actors are welcome in the nano-world. In the particular case of Robinson, his pro-active service role was recognized, i.e. that such non-technical actors could be of some help (in indicating innovation dynamics and contributing to roadmapping, for example). The main drivers of acceptance were the pressures on the nano-world, as visible in the concerns about risk and in the call for responsible development. Listening to the knowledgeable visitor, and accepting CTA exercises, were ways to address these pressures.

Are outcomes in terms of CTA goals visible? Of course, it is too early to see better technology in a better society (and if so, it would not be attributable to CTA exercises). But one may see increased reflexivity in co-evolution. This relates to anticipatory coordination. In the world of nanotechnology, there is an interest in anticipation and coordination so as to choose right directions. Actual and potential stakeholders are attempting to shape emerging nanotechnology developments, in different fora and with a variety of strategies. CTA exercises are part of this move, and they create further openings. As they do this, they become recognized and accepted. There is some institutionalization of scenario/strategy workshops (Robinson 2010; Parandian 2012).

Insertion is an integral part of the CTA activities, and necessary to make them effective. It is not a means to achieve CTA goals directly, even it does contribute.[16] It is reconnoitering the lay of the land and probing the dynamics. On that basis, circumstances (like CTA workshops) can be created that stimulate actors to reflect, act and interact in ways that might achieve the CTA agent's objectives.

A key element in achieving these objectives is making visible what was invisible to actors,[17] not by explaining (although that might occur), but in interaction with actors (that's also where the scenarios come in). As it is experimenting in real-world interactions, there is an interesting link with Lindblom's (1990) plea for inquiry rather than a search for truth as such, in relation to change. People probe the world (probe into situations, into other actor's perspectives, into problems and possible solutions) in order to change it, and this constitutes inquiry. The resulting insights can be formulated as such, somewhat independent of proposed actions. Social scientists also probe the world, whether they have a change perspective or not. Lindblom emphasizes that there is no epistemological difference between probing by citizens, by government functionaries and by social scientists. However, as he notes, the latter may well have more honed and articulated probing skills. When one scales down the scope of Lindblom's argument from society in general to the world of nanotechnology development, it constitutes a justification of the 'insertion' approach. It is probing by the social scientist, but also stimulates probing by the actors themselves.[18]

3.4 Concluding Thoughts

For new technologies, most concrete activities are at the R&D stage, rather than product development and uptake in society. Firms and research institutes are important locations, but given the open-ended promises for new technologies like nanotechnology, academic research institutions are important as well. This introduces additional dynamics, related to "opening up the laboratory", as the title of this volume phrases it.

In a sense, scientists (even the technoscientists that abound in nanotechnology) are outsiders to society, because they live in protected places (Rip 2010a). They are insiders in their own world of science, and strongly feel like insiders, up to

---

[16] Social scientists moving about in the world of a scientific specialty or domain will set the members of that world thinking about what is happening, and about patterns that enable or constrain. This is relevant for the overall CTA goal of increasing reflexivity of co-evolution of technology and society. Moving about in the nano-worlds may have such an effect, but it was not an explicit aim that structured the moving about.

[17] A sort of sociological enlightenment in the small, cf. Rip and Groen (2001).

[18] Phrased in this way, there is overlap with participatory research approaches (cf. Bergold and Thomas 2012). There, the social scientists have the higher status, while in our case, nano-scientists and policy makers tend to relegate the social scientists to a service role. Thus, building trust will have a different complexion.

patrolling and protecting the boundaries of their world.[19] Bridging the gap between the inside world of science and the outside world now occurs in various ways, proactively or because of outside pressures.

Social scientists and humanities scholars are outsiders to that world of science, in particular to the protected place of the lab where the work of science is done. They can visit, even become accommodated to some extent – perhaps as "social scientist in residence". Social scientists visiting a lab, occasionally staying there for some time, shift out of their own world. Anthropologists and ethnographers (of science) have been doing that all along, but with another purpose, to gather data rather than changing the world they study. Their presence would increase reflexivity of the actors, however, whether they wanted that or not. Our methodology of insertion is explicit about this.[20]

CTA has a larger scope, and addresses embedding in society, if only through anticipation. The dynamics will be more complex: there are now different overlapping worlds, different perspectives, and actors at the collective level (ranging from branch organizations to government agencies), with some collective responsibility. And there larger and long-term developments, in particular the traditional division of labour between promotion and control, which is now questioned, as in the discourse of responsible research and innovation.

Concretely, in the world of nanotechnology, CTA exercises are welcomed (and funded) by the technology developers and technology promoters, who see them as necessary to anticipate on societal embedding, and meeting possible reactions from various societal actors. Co-evolution of technology and society goes on anyway, but anticipations are becoming more important, so that the co-evolution will be more reflexive – even if enactors will work from their concentric perspective.

If co-evolution becomes reflexive, and actors absorb CTA activities in their practices, will CTA agents become superfluous? Not yet, and probably never. One reason is that CTA agents can circulate across locations, and observe and analyse what happens at the collective level, which will be more difficult for regular actors. Another reason is that these visiting "knowledgeable" strangers irritate existing ways of working and thus create openings for learning and further evolution of how we handle new technologies in our society.

References

Bergold, J., & Thomas, S. (2012). Participatory research methods: A methodological approach in motion. *Forum: Qualitative Social Research, 13*(1), 1–13.

Callon, M., Lascoumes, P., & Barthe, Y. (2001). *Agir dans un monde incertain. Essai sur la démocratie technique*. Paris: éd. Seuil.

---

[19] There is a functional argument: scientists should live in protected spaces, at least to some extent, in order to be productive (Rip 2010a).

[20] There are normative issues involved, which can refer to the background goals of CTA, but have also an experimental component, finding out about the issues by doing and learning (cf. also Laurent and Van Oudheusden 2013).

Daey Ouwens, C., van Hoogstraten, P., Jelsma, J., Prakke, F., & Rip, A. (1987). *Constructief Technologisch Aspectenonderzoek. Een Verkenning*. Den Haag: Staatsuitgeverij (NOTA Voorstudie 4).

Delvenne, P. (2011). *Science, technologie et innovation sur le chemin de la réflexivité. Enjeus et dynamiques du Technology Assessment parlementaire*. Louvain-La-Neuve: Harmattan-Academia.

European Commission (2008, February 7). *Commission recommendation on a code of conduct for responsible nanosciences and nanotechnologies research* (C(2008)424 final) Available at http://ec.europa.eu/nanotechnology/pdf/nanocode-rec_pe0894c_en.pdf

Fenn, J. (1999). *When to leap on the hype cycle: Research note*. Stamford: Gartner Group. www.cata.ca/files/PDF/Resource_Centres/hightech/reports/indepstudies/Whentoleaponthehypecycle.pdf

Garud, R., & Ahlstrom, D. (1997). Technology assessment: A socio-cognitive perspective. *Journal of Engineering and Technology Management, 14*, 25–48.

Genus, A. (2006). Rethinking constructive technology assessment as democratic, reflective, discourse. *Technology Forecasting and Social Change, 73*, 13–26.

Grin, J., & Van de Graaf, H. (1996). Technology assessment as learning. *Science Technology & Human Values, 21*, 72–99.

Guston, D. H., & Sarewitz, D. (2002). Real-time technology assessment. *Technology and Society, 24*, 93–109.

Harremoës, P. (Ed.). (2001). *Late lessons from early warnings: The precautionary principle 1896–2000*. Copenhagen: European Environmental Agency.

Hoogma, R. (2000). *Exploiting technological niches: Strategies for experimental introduction of electric vehicles*. PhD dissertation, Twente University Press, Enschede.

Hoogma, R., Kemp, R., Schot, J., & Truffer, B. (2002). *Experimenting for sustainable transport: The approach of strategic niche management*. London: Spon Press.

Laurent, B., & van Oudheusden, M. (2013). Experimental normativity. Shifting and deepening engagement in public participation in science and technology. Manuscript, Paris: CSI, Ecole des Mines.

Lindblom, C. E. (1990). *Inquiry and change. The troubled attempt to understand and shape society*. New Haven: Yale University Press.

LMC, et al. (2010). (Local Monitoring Committee, Olivier Lemaire, Anne Moneyron, Jean E. Masson), “Interactive Technology Assessment” and beyond: The Field Trial of Genetically Modified Grapevines at INRA-Colmar. *PLoS Biology, 8*(11), separately paginated 1–7.

Marris, C., Rip, A., & Joly, P.-B. (2008). Interactive technology assessment in the real world: Dual dynamics in an iTA exercise on genetically modified vines’. *Science, Technology & Human Values, 33*(1), 77–100.

Ministerie van Onderwijs en Wetenschappen. (1983–1984). Integratie van Wetenschap en Technologie in de Samenleving. Beleidsnota, ‘s-Gravenhage: Tweede Kamer, 18 421, nrs. 1–2. (Policy Memorandum: Integration of Science and Technology in Society)

Nordmann, A., & Rip, A. (2009). Mind the gap revisited. *Nature Nanotechnology, 4*(May 2009), 273–274.

Parandian, A. (2012). *Constructive TA of newly emerging technologies*. Stimulating learning by anticipation through bridging events. PhD thesis, Technical University Delft, defended 12 Mar 2012.

Parandian, A., Rip, A., & Te Kulve, H. (2012). Dual dynamics of promises and waiting games around emerging nanotechnologies. *Technology Analysis & Strategic Management, 24*(6), 565–582.

Rip, A. (1992). Between innovation and evaluation: Sociology of technology applied to technology policy and technology assessment. *RISESST, 2*, 39–68.

Rip, A. (2001a). Technology assessment. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (Vol. 23, pp. 15512–15515). Oxford: Oxford University Press.

Rip, A. (2001b). Assessing the impacts of innovation: New developments in technology assessment. In *OECD proceedings, social sciences and innovation* (pp. 197–213). Paris: OECD.

Rip, A. (2002, June 7). *Co-evolution of science, technology and society*. Expert review for the Bundesministerium Bildung and Forschung's Förderinitiative 'Politik, Wissenschaft und Gesellschaft' (Science Policy Studies), managed by the Berlin-Brandenburgische Akademie der Wissenschaften. Enschede: University of Twente.

Rip, A. (2006). Futures of ELSA. *EMBO Reports, 10*(7), 666–670.

Rip, A. (2010a). Protected spaces of science: Their emergence and further evolution in a changing world. In M. Carrier & A. Nordmann (Eds.), *Science in the context of application: Methodological change, conceptual transformation, cultural reorientation* (pp. 197–220). Dordrecht: Springer.

Rip, A. (2010b). De facto governance of nanotechnologies. In M. Goodwin, B.-J. Koops, & R. Leenes (Eds.), *Dimensions of technology regulation* (pp. 285–308). Nijmegen: Wolf Legal Publishers.

Rip, A. (2010c). Processes of Entanglement, in Madeleine Akrich, Yannick Barthe, Fabian Muniesa et Philippe Mustar (réd.), Débordements. Mélanges offerts à Michel Callon (pp. 381–392). Paris: Transvalor – Presses des Mines.

Rip, A., & Groen, A. (2001). Many visible hands. In R. Coombs, K. Green, V. Walsh, & A. Richards (Eds.), *Technology and the market. Demands, users and innovation* (pp. 12–37). Cheltenham: Edward Elgar.

Rip, A., & Shelley-Egan, C. (2010). Positions and responsibilities in the "real" world of nanotechnology. In R. von Schomberg & S. Davies (Eds.), *Understanding public debate on nanotechnologies. Options for framing public policy* (pp. 31–38). Brussels: Commission of the European Communities.

Rip, A., & Te Kulve, H. (2008). Constructive technology assessment and sociotechnical scenarios. In E. Fisher, C. Selin, & J. M. Wetmore (Eds.), *The yearbook of nanotechnology in society* (Presenting futures, Vol. I, pp. 49–70). Berlin etc.: Springer.

Rip, A., Misa, T. J., & Schot, J. (Eds.). (1995). *Managing technology in society. The approach of constructive technology assessment*. London/New York: Pinter Publishers.

Robinson, D. K. R. (2009). Co-evolutionary scenarios: An application to prospecting futures of the responsible development of nanotechnology. *Technological Forecasting and Social Change, 76*, 1222–1239.

Robinson, D. K. R. (2010). *Constructive technology assessment of emerging nanotechnologies. Experiments in interactions*. Enschede: University of Twente. PhD thesis, defended 25 Nov 2010.

Robinson, D. K. R., & Propp, T. (2008). Multi-path mapping as a tool for reflexive alignment in emerging S&T. *Technological Forecasting and Social Change, 75*(2008), 517–538.

Schot, J., & Rip, A. (1997). The past and future of constructive technology assessment. *Technological Forecasting and Social Change, 54*(1997), 251–268.

Schuurbiers, D., & Fisher, E. (2009). Lab-scale intervention. *EMBO Reports, 10*(5), 424–427.

Swierstra, T., & Rip, A. (2007). Nano-ethics as NEST-ethics: Patterns of moral argumentation about new and emerging science and technology. *NanoEthics, 1*(2007), 3–20.

Te Kulve, H. (2011). *Anticipatory interventions in the co-evolution of nanotechnology and society*. PhD thesis, University of Twente, defended 21 Apr 2011.

Van de Poel, I. R. (1998). *Changing technologies A comparative study of eight processes of transformation of technological regimes*. Enschede: Twente University Press.

Von Schomberg, R. (2007). *From the ethics of technology towards and ethics of knowledge policy*. Working document of the Service of the European Commission. http://ec.europa.eu/research/science-society/pdf/ethicsofknowledgepolicy_en.pdf

Chapter 4
Value Sensitive Design and Information Systems

Batya Friedman, Peter H. Kahn Jr., and Alan Borning

Abstract Value Sensitive Design is a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process. It employs an integrative and iterative tripartite methodology, consisting of conceptual, empirical, and technical investigations. We explicate Value Sensitive Design by drawing on three case studies. The first study concerns information and control of web browser cookies, implicating the value of informed consent. The second study concerns using high-definition plasma displays in an office environment to provide a "window" to the outside world, implicating the values of physical and psychological well-being and privacy in public spaces. The third study concerns an integrated land use, transportation, and environmental simulation system to support public deliberation and

---

From *Human-Computer Interaction and Management Information Systems: Foundations Advances in Management Information Systems,* Volume 5 (Advances in Management Information Systems), ed. Ping Zhang & Dennis Galletta (Armonk, NY: M.E. Sharpe, 2006), pp. 348–372. Copyright © 2006 by M.E. Sharpe, Inc. Reprinted with permission.

With an addendum by Alina Huldtgren, PhD
Duesseldorf University of Applied Sciences
Josef-Gockeln-Str. 9, 40474 Duesseldorf, Germany
e-mail: alina.huldtgren@fh-duesseldorf.de

B. Friedman (✉)
Information School, University of Washington, Box 352840, Seattle, WA 98195-2840, USA
e-mail: batya@uw.edu

P.H. Kahn Jr.
Department of Psychology, University of Washington,
Box 351525, Seattle, WA 98195-1525, USA
e-mail: pkahn@uw.edu

A. Borning
Department of Computer Science and Engineering, University of Washington,
Box 352350, Seattle, WA 98195-2350, USA
e-mail: borning@cs.washington.edu

N. Doorn et al. (eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*, Philosophy of Engineering and Technology 16, DOI 10.1007/978-94-007-7844-3_4,

debate on major land use and transportation decisions, implicating the values of fairness, accountability, and support for the democratic process, as well as a highly diverse range of values that might be held by different stakeholders, such as environmental sustainability, opportunities for business expansion, or walkable neighborhoods. We conclude with direct and practical suggestions for how to engage in Value Sensitive Design.

The original version of this chapter is published by M.E. Sharpe (www.mesharpe.com). This chapter contains a reprint of the original paper with an additional commentary.

4.1 Introduction

There is a longstanding interest in designing information and computational systems that support enduring human values. Researchers have focused, for example, on the value of *privacy* (Ackerman and Cranor 1999; Agre and Rotenberg 1998; Fuchs 1999; Jancke et al. 2001; Palen and Grudin 2003; Tang 1997), *ownership* and *property* (Lipinski and Britz 2000), *physical welfare* (Leveson 1991), *freedom from bias* (Friedman and Nissenbaum 1996), *universal usability* (Shneiderman 1999, 2000; Thomas 1997), *autonomy* (Suchman 1994; Winograd 1994), *informed consent* (Millett et al. 2001), and *trust* (Fogg and Tseng 1999; Palen and Grudin 2003; Riegelsberger and Sasse 2002; Rocco 1998; Zheng et al. 2001). Still, there is a need for an overarching theoretical and methodological framework with which to handle the value dimensions of design work.

Value Sensitive Design is one effort to provide such a framework (e.g., Friedman (1997a), Friedman and Kahn (2003), Friedman and Nissenbaum (1996), Hagman et al. (2003), Nissenbaum (1998), Tang (1997), and Thomas (1997)). Our goal in this paper is to provide an account of Value Sensitive Design, with enough detail for other researchers and designers to critically examine and systematically build on this approach.

We begin by sketching the key features of Value Sensitive Design, and then describe its integrative tripartite methodology, which involves conceptual, empirical, and technical investigations, employed iteratively. Then we explicate Value Sensitive Design by drawing on three case studies. One involves cookies and informed consent in web browsers; the second involves HDTV display technology in an office environment; the third involves user interactions and interface for an integrated land use, transportation, and environmental simulation. We conclude with direct and practical suggestions for how to engage in Value Sensitive Design.

4.2 What Is Value Sensitive Design?

Value Sensitive Design is a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process.

4.2.1 What Is a Value?

In a narrow sense, the word “value” refers simply to the economic worth of an object. For example, the value of a computer could be said to be 2,000 dollars. However, in the work described here, we use a broader meaning of the term wherein a value refers to what a person or group of people consider important in life.[1] In this sense, people find many things of value, both lofty and mundane: their children, friendship, morning tea, education, art, a walk in the woods, nice manners, good science, a wise leader, clean air.

This broader framing of values has a long history. Since the time of Plato, for example, the content of value-oriented discourse has ranged widely, emphasizing “the good, the end, the right, obligation, virtue, moral judgment, aesthetic judgment, the beautiful, truth, and validity” (Frankena 1972, p. 229). Sometimes ethics has been subsumed within a theory of values, and other times conversely, with ethical values viewed as just one component of ethics more generally. Either way, it is usually agreed (Moore 1903/1978) that values should not be conflated with facts (the “fact/value distinction”) especially insofar as facts do not logically entail value. In other words, “is” does not imply “ought” (the naturalistic fallacy). In this way, values cannot be motivated only by an empirical account of the external world, but depend substantively on the interests and desires of human beings within a cultural milieu. In Table 4.1 in Sect. 4.2.2, we provide a list of human values with ethical import that are often implicated in system design, along with working definitions and references to the literature.

4.2.2 Related Approaches to Values and System Design

In the 1950s, during the early periods of computerization, cyberneticist Norbert Wiener (1953/1985) argued that technology could help make us better human beings, and create a more just society. But for it to do so, he argued, we have to take control of the technology. We have to reject the “worshiping [of] the new gadgets which are our own creation as if they were our masters” (p. 678). Similarly, a few decades later, computer scientist Joseph Weizenbaum (1972) wrote:

> What is wrong, I think, is that we have permitted technological metaphors…and technique itself to so thoroughly pervade our thought processes that we have finally abdicated to technology the very duty to formulate questions…Where a simple man might ask: “Do we need these things?”, technology asks “what electronic wizardry will make them safe?” Where a simple man will ask “is it good?”, technology asks “will it work?” (pp. 611–612).

---

[1] The Oxford English Dictionary definition of this sense of value is: “the principles or standards of a person or society, the personal or societal judgement of what is valuable and important in life.” (Simpson and Weiner 1989).

Table 4.1 Human values (with Ethical Import) often implicated in system design

Human value	Definition	Sample literature
Human welfare	Refers to people's physical, material, and psychological well-being	Leveson (1991), Friedman et al. (2003), Neumann (1995), Turiel (1983, 1998)
Ownership and property	Refers to a right to possess an object (or information), use it, manage it, derive income from it, and bequeath it	Becker (1977), Friedman (1997b), Herskovits (1952), Lipinski and Britz (2000)
Privacy	Refers to a claim, an entitlement, or a right of an individual to determine what information about himself or herself can be communicated to others	Agre and Rotenberg (1998), Bellotti (1998), Boyle et al. (2000), Friedman (1997b), Fuchs (1999), Jancke et al. (2001), Palen and Dourish (2003), Nissenbaum (1998), Phillips (1998), Schoeman (1984), Svensson et al. (2001)
Freedom from bias	Refers to systematic unfairness perpetrated on individuals or groups, including pre-existing social bias, technical bias, and emergent social bias	Friedman and Nissenbaum (1996), cf. Nass and Gong (2000), Reeves and Nass (1996)
Universal usability	Refers to making all people successful users of information technology	Aberg and Shahmehri (2001), Shneiderman (1999, 2000), Cooper and Rejmer (2001), Jacko et al. (1999), Stephanidis (2001)
Trust	Refers to expectations that exist between people who can experience good will, extend good will toward others, feel vulnerable, and experience betrayal	Baier (1986), Camp (2000), Dieberger et al. (2001), Egger (2000), Fogg and Tseng (1999), Friedman et al. (2000a), Kahn and Turiel (1988), Mayer et al. (1995), Olson and Olson (2000), Nissenbaum (2001), Rocco (1998)
Autonomy	Refers to people's ability to decide, plan, and act in ways that they believe will help them to achieve their goals	Friedman and Nissenbaum (1997), Hill (1991), Isaacs et al. (1996), Suchman (1994), Winograd (1994)
Informed consent	Refers to garnering people's agreement, encompassing criteria of disclosure and comprehension (for "informed") and voluntariness, competence, and agreement (for "consent")	Faden and Beauchamp (1986), Friedman et al. (2000b), The Belmont Report (1978)
Accountability	Refers to the properties that ensures that the actions of a person, people, or institution may be traced uniquely to the person, people, or institution	Friedman and Kahn (1992), Friedman and Millet (1995), Reeves and Nass (1996)
Courtesy	Refers to treating people with politeness and consideration	Bennett and Delatree (1978), Wynne and Ryan (1993)

(continued)

Table 4.1 (continued)

Human value	Definition	Sample literature
Identity	Refers to people's understanding of who they are over time, embracing both continuity and discontinuity over time	Bers et al. (2001), Rosenberg (1997), Schiano and White (1998), Turkle (1996)
Calmness	Refers to a peaceful and composed psychological state	Friedman and Kahn (2003), Weiser and Brown (1997)
Environmental sustainability	Refers to sustaining ecosystems such that they meet the needs of the present without compromising future generations	United Nations (1992), World Commission on Environment and Development (1987), Hart (1999), Moldan et al. (1997), Northwest Environment Watch (2002)

More recently, supporting human values through system design has emerged within at least four important approaches. *Computer Ethics* advances our understanding of key values that lie at the intersection of computer technology and human lives, e.g., Bynum (1985), Johnson and Miller (1997), and Nissenbaum (1999). *Social Informatics* has been successful in providing socio-technical analyses of deployed technologies, e.g., Johnson (2000), Kling et al. (1998), Kling and Star (1998), Orlikowski and Iacono (2001) and Sawyer and Rosenbaum (2000). *Computer Supported Cooperative Work* (CSCW) has been successful in the design of new technologies to help people collaborate effectively in the workplace, e.g., Fuchs (1999), Galegher et al. (1990), Olson and Teasley (1996), and Grudin (1988). Finally, *Participatory Design* substantively embeds democratic values into its practice, e.g., Bjerknes and Bratteteig (1995), Bødker (1990), Carroll and Rosson (2006), Ehn (1989), Greenbaum and Kyng (1991), and Kyng and Mathiassen (1997). (See Friedman and Kahn (2003) for a review of each of these approaches.)

4.3 The Tripartite Methodology: Conceptual, Empirical, and Technical Investigations

Think of an oil painting by Monet or Cézanne. From a distance it looks whole; but up close you can see many layers of paint upon paint. Some paints have been applied with careful brushstrokes, others perhaps energetically with a palate knife or fingertips, conveying outlines or regions of color. The diverse techniques are employed one on top of the other, repeatedly, and in response to what has been laid down earlier. Together they create an artifact that could not have been generated by a single technique in isolation of the others. So, too, with Value Sensitive Design. An artifact or design emerges through iterations upon a process that is more than the sum of its parts. Nonetheless, the parts provide us with a good place to start. Value Sensitive Design builds on an iterative methodology that integrates conceptual, empirical, and technical investigations; thus, as a step toward conveying Value Sensitive Design, we describe each investigation separately.

4.3.1 Conceptual Investigations

Who are the direct and indirect stakeholders affected by the design at hand? How are both classes of stakeholders affected? What values are implicated? How should we engage in trade-offs among competing values in the design, implementation, and use of information systems (e.g., autonomy vs. security, or anonymity vs. trust)? Should moral values (e.g., a right to privacy) have greater weight than, or even trump, non-moral values (e.g., aesthetic preferences)? Value Sensitive Design takes up these questions under the rubric of conceptual investigations.

In addition, careful working conceptualizations of specific values clarify fundamental issues raised by the project at hand, and provide a basis for comparing results across research teams. For example, in their analysis of trust in online system design, Friedman et al. (2000a), drawing on Baier (1986), first offer a philosophically informed working conceptualization of trust. They propose that people trust when they are vulnerable to harm from others, yet believe those others would not harm them even though they could. In turn, trust depends on people's ability to make three types of assessments. One is about the harms they might incur. The second is about the good will others possess toward them that would keep those others from doing them harm. The third involves whether or not harms that do occur lie outside the parameters of the trust relationship. From such conceptualizations, Friedman et al. were able to define clearly what they meant by trust online. This definition is in some cases different from what other researchers have meant by the term – for example, the Computer Science and Telecommunications Board, in their thoughtful publication *Trust in Cyberspace* (Schneider 1999), adopted the terms "trust" and "trustworthy" to describe systems that perform as expected along the dimensions of correctness, security, reliability, safety, and survivability. Such a definition, which equates "trust" with expectations for machine performance, differs markedly from one that says trust is fundamentally a relationship between people (sometimes mediated by machines).

4.3.2 Empirical Investigations

Conceptual investigations can only go so far. Depending on the questions at hand, many analyses will need to be informed by empirical investigations of the human context in which the technical artifact is situated. Empirical investigations are also often needed to evaluate the success of a particular design. Empirical investigations can be applied to any human activity that can be observed, measured, or documented. Thus, the entire range of quantitative and qualitative methods used in social science research is potentially applicable here, including observations, interviews, surveys, experimental manipulations, collection of relevant documents, and measurements of user behavior and human physiology.

Empirical investigations can focus, for example, on questions such as: How do stakeholders apprehend individual values in the interactive context? How do they prioritize competing values in design trade-offs? How do they prioritize individual values and usability considerations? Are there differences between espoused practice (what people say) compared with actual practice (what people do)? Moreover, because the development of new technologies affects groups as well as individuals, questions emerge of how organizations appropriate value considerations in the design process. For example, regarding value considerations, what are organizations' motivations, methods of training and dissemination, reward structures, and economic incentives?

4.3.3 Technical Investigations

As discussed in Sect. 4.5 (Value Sensitive Design's Constellation of Features), Value Sensitive Design adopts the position that technologies in general, and information and computer technologies in particular, provide value suitabilities that follow from properties of the technology. That is, a given technology is more suitable for certain activities and more readily supports certain values while rendering other activities and values more difficult to realize.

In one form, technical investigations focus on how existing technological properties and underlying mechanisms support or hinder human values. For example, some video-based collaborative work systems provide blurred views of office settings, while other systems provide clear images that reveal detailed information about who is present and what they are doing. Thus the two designs differentially adjudicate the value trade-off between an individual's *privacy* and the group's *awareness* of individual members' presence and activities.

In the second form, technical investigations involve the proactive design of systems to support values identified in the conceptual investigation. For example, Fuchs (1999) developed a notification service for a collaborative work system in which the underlying technical mechanisms implement a value hierarchy whereby an individual's desire for privacy overrides other group members' desires for awareness.

At times, technical investigations – particularly of the first form – may seem similar to empirical investigations insofar as both involve technological and empirical activity. However, they differ markedly on their unit of analysis. Technical investigations focus on the technology itself. Empirical investigations focus on the individuals, groups, or larger social systems that configure, use, or are otherwise affected by the technology.

4.4 Value Sensitive Design in Practice: Three Case Studies

To illustrate Value Sensitive Design's integrative and iterative tripartite methodology, we draw on three case studies with real world applications, one completed and two under way. Each case study represents a unique design space.

4.4.1 Cookies and Informed Consent in Web Browsers

Informed consent provides a critical protection for privacy, and supports other human values such as autonomy and trust. Yet currently there is a mismatch between industry practice and the public's interest. According to a recent report from the Federal Trade Commission (2000), for example, 59 % of Web sites that collect personal identifying information neither inform Internet users that they are collecting such information nor seek the user's consent. Yet, according to a Harris poll (2000), 88 % of users want sites to garner their consent in such situations.

Against this backdrop, Friedman, Felten, and their colleagues (Friedman et al. 2002, 2000b; Millett et al. 2001) sought to design web-based interactions that support informed consent in a web browser through the development of new technical mechanisms for cookie management. This project was an early proof-of-concept project for Value Sensitive Design, which we use here to illustrate several key features of the methodology.

4.4.1.1 Conceptualizing the Value

One part of a conceptual investigation entails a philosophically informed analysis of the central value constructs. Accordingly, Friedman et al. began their project with a conceptual investigation of informed consent itself. They drew on diverse literature, such as the Belmont Report, which delineates ethical principles and guidelines for the protection of human subjects (Belmont Report 1978; Faden and Beauchamp 1986), to develop criteria for informed consent in online interactions. In brief, the idea of "informed" encompasses disclosure and comprehension. *Disclosure* refers to providing accurate information about the benefits and harms that might reasonably be expected from the action under consideration. *Comprehension* refers to the individual's accurate interpretation of *what* is being disclosed. In turn, the idea of "consent" encompasses voluntariness, comprehension, and agreement. *Voluntariness* refers to ensuring that the action is not controlled or coerced. *Competence* refers to possessing the mental, emotional and physical capabilities needed to be capable of giving informed consent. *Agreement* refers to a reasonably clear opportunity to accept or decline to participate. Moreover, agreement should be ongoing, that is, the individual should be able to withdraw from the interaction at any time. See Friedman et al. (2000b) for an expanded discussion of these five criteria.

4.4.1.2 Using a Conceptual Investigation to Analyze Existing Technical Mechanisms

With a conceptualization for informed consent online in hand, Friedman et al. conducted a retrospective analysis (one form of a technical investigation) of how the cookie and web-browser technology embedded in Netscape Navigator and Internet

Explorer changed with respect to informed consent over a 5-year period, beginning in 1995. Specifically, they used the criteria of disclosure, comprehension, voluntariness, competence, and agreement to evaluate how well each browser in each stage of its development supported the users' experience of informed consent. Through this retrospective analysis, they found that while cookie technology had improved over time regarding informed consent (e.g., increased visibility of cookies, increased options for accepting or declining cookies, and access to information about cookie content), as of 1999 some startling problems remained. For example: (a) While browsers disclosed to users some information about cookies, they still did not disclose the right sort of information – that is, information about the potential harms and benefits from setting a particular cookie. (b) In Internet Explorer, the burden to accept or decline all third party cookies still fell to the user, placing undue burden on the user to decline each third party cookie one at a time. (c) Users' out-of-the-box experience of cookies (i.e., the default setting) was no different in 1999 than it was in 1995: to accept all cookies. That is, the novice user installed a browser that accepted all cookies and disclosed nothing about that activity to the user. (d) Neither browser alerted a user when a site wished to use a cookie and for what purpose, as opposed to when a site wished to store a cookie.

4.4.1.3 The Iteration and Integration of Conceptual, Technical, and Empirical Investigations

Based on the results from these conceptual and technical investigations, Friedman et al. then iteratively used the results to guide a second technical investigation: a redesign of the Mozilla browser (the open-source code for Netscape Navigator). Specifically, they developed three new types of mechanisms: (a) peripheral awareness of cookies; (b) just-in-time information about individual cookies and cookies in general; and (c) just-in-time management of cookies (see Fig. 4.1). In the process of their technical work, Friedman et al. conducted formative evaluations (empirical investigations) which led to a further design criterion, minimal distraction, which refers to meeting the above criteria for informed consent without unduly diverting the user from the task at hand. Two situations are of concern here. First, if users are overwhelmed with queries to consent to participate in events with minor benefits and risks, they may become numbed to the informed consent process by the time participation in an event with significant benefits and risks is at hand. Thus, the user's participation in that event may not receive the careful attention that is warranted. Second, if the overall distraction to obtain informed consent becomes so great as to be perceived to be an intolerable nuisance, users are likely to disengage from the informed consent process in its entirety and accept or decline participation by rote. Thus undue distraction can single-handedly undermine informed consent. In this way, the iterative results of the above empirical investigations not only shaped and then validated the technical work, but impacted the initial conceptual investigation by adding to the model of informed consent the criterion of minimal distraction.

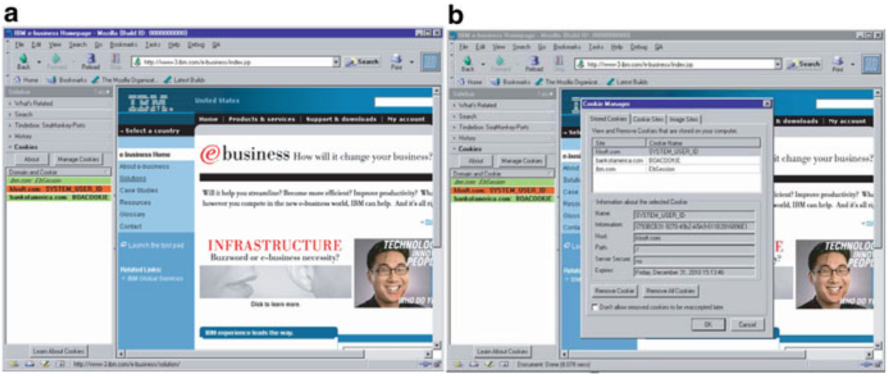


Fig. 4.1 Screen shot (**a**) of the Mozilla implementation shows the peripheral awareness of cookies interface (at the *left*) in the context of browsing the web. Each time a cookie is set, a color-coded entry for that cookie appears in the sidebar. Third party cookies are *red*; others are *green*. At the user's discretion, he or she can click on any entry to bring up the Mozilla cookie manager for that cookie. Screen shot (**b**) after the user has clicked on an entry to bring up the just-in-time cookie management tool (in the *center*) for a particular cookie

Thus, this project illustrates the iterative and integrative nature of Value Sensitive Design, and provides a proof-of-concept for Value Sensitive Design in the context of mainstream Internet software.

4.4.2 Room with a View: Using Plasma Displays in Interior Offices

Janice is in her office, writing a report. She's trying to conceptualize the report's higher-level structure, but her ideas won't quite take form. Then she looks up from her desk and rests her eyes on the fountain and plaza area outside her building. She notices the water bursting upward, and that a small group of people are gathering by the water's edge. She rests her eyes on the surrounding pool of calm water. Her eyes then lift toward the clouds and the streaking sunshine. Twenty seconds later she returns to her writing task at hand, slightly refreshed, and with an idea taking shape.

What's particularly novel about this workplace scenario is that Janice works in an interior office. Instead of a real window looking out onto the plaza, Janice has a large screen video plasma display that continuously displays the local outdoor scene in real-time. Realistic? Beneficial? This design space is currently being researched by Kahn, Friedman, and their colleagues, using the framework of Value Sensitive Design.

In Kahn et al.'s initial conceptual investigation of this design space, they drew on the psychological literature that suggests that interaction with real nature can garner physiological and psychological benefits. For example, in one study, Ulrich (1984) found that post-operative recovery improved when patients were assigned to a room with a view of a natural setting (a small stand of deciduous trees) versus a view of a brown brick wall. More generally, studies have shown that even minimal connection

Fig. 4.2 Plasma Display Technology Studies. From the standpoint of illustrating Value Sensitive Design, we would like to emphasize three ideas. (**a**) The watcher, (**b**) The HDTV camera, (**c**) The watched

with nature – such as looking at a natural landscape – can reduce immediate and long-term stress, reduce sickness of prisoners, and calm patients before and during surgery. (See Beck and Katcher (1996), Kahn (1999), and Ulrich (1993) for reviews.) Thus Kahn et al. hypothesized that an "augmented window" of nature could render benefits in a work environment in terms of the human values of physical health, emotional well-being, and creativity.

To investigate this question in a laboratory context, Kahn et al. are comparing the short-term benefits of working in an office with a view out the window of a beautiful nature scene versus an identical view (in real time) shown on a large video plasma display that covers the window in the same office (Fig. 4.2a). In this latter condition, they employed a High Definition TV (HDTV) camera (Fig. 4.2b) to capture real-time local images. The control condition involved a blank covering over the window. Their measures entailed (a) physiological data (heart rate), (b) performance data (on cognitive and creativity tasks), (c) video data that captured each subject's eye gaze on a second-by-second level, and time synchronized with the physiological equipment, so that analyses can determine whether physiological benefits accrued immediately following an eye gaze onto the plasma screen, and (d) social-cognitive data (based on a 50-min interview with each subject at the conclusion of the experimental condition wherein they garnered each subject's reasoned perspective on the experience). Preliminary results show the following trends. First, participants looked out the plasma screen just as frequently as they did the real window, and more frequently than they stared at the blank wall. In this sense, the plasma-display window was functioning like a real window. But, when participants gazed for 30 s or more, the real window provided greater physiological recovery from low-level stress as compared to the plasma display window.

4.4.2.1 Multiple Empirical Methods

Under the rubric of empirical investigations, Value Sensitive Design supports and encourages multiple empirical methods to be used in concert to address the question at hand. As noted above, for example, this study employed physiological data (heart rate), two types of performance data (on cognitive and creativity tasks), behavioral data (eye gaze), and reasoning data (the social-cognitive interview). From a value-oriented perspective, multiple psychological measures increase the veracity of most accounts of technology in use.

4.4.2.2 Direct and Indirect Stakeholders

In their initial conceptual investigation of the values implicated in this study, Kahn et al. sought to identify not only direct but also indirect stakeholders affected by such display technology. At that early point, it became clear to the researchers that an important class of indirect stakeholders (and their respective values) needed to be included: namely, the individuals who, by virtue of walking through the fountain scene, unknowingly had their images displayed on the video plasma display in the "inside" office (Fig. 4.2c). In other words, if this application of projection technology were to come into widespread use (as web cams and surveillance cameras have begun to) then it would potentially encroach on the privacy of individuals in public spaces – an issue that has been receiving increasing attention in the field of computer ethics and public discourse (Nissenbaum 1998). Thus, in addition to the experimental laboratory study, Kahn et al. initiated two additional but complementary empirical investigations with indirect stakeholders: (a) a survey of 750 people walking through the public plaza, and (b) in-depth social cognitive interviews with 30 individuals walking through the public plaza (Friedman et al. 2006). Both investigations focused on indirect stakeholders' judgments of privacy in public space, and in particular having their real-time images captured and displayed on plasma screens in nearby and distant offices. The importance of such indirect stakeholder investigations is being borne out by the results. For example, significant gender differences were found in their survey data: more women than men expressed concern about the invasion of privacy through web cameras in public places. This finding held whether their image was to be displayed locally or in another city (Tokyo), or viewed by one person, thousands, or millions. One implication of this finding is that future technical designs and implementations of such display technologies need to be responsive to ways in which men and women might perceive potential harms differently.

4.4.2.3 Coordinated Empirical Investigations

Once Kahn et al. identified an important group of indirect stakeholders, and decided to undertake empirical investigations with this group, they then coordinated these empirical investigations with the initial (direct stakeholder) study. Specifically, a subset of identical questions were asked of both the direct stakeholders ("The Watchers") and indirect stakeholders ("The Watched"). Results show some interesting differences. For example, more men in The Watched condition expressed concerns that people's images might be displayed locally, nationally, or internationally than men in The Plasma Display Watcher condition. No differences were found between women in The Watcher Plasma Display Condition and women in the Watched condition. Thus, the Value Sensitive Design methodology helps to bring to the forefront values that matter not only to the direct stakeholders of a technology (such as physical health, emotional well-being, and creativity), but to the indirect stakeholders (such as privacy, informed consent, trust, and physical safety). Moreover, from the standpoint of Value Sensitive Design, the above study highlights how investigations of indirect stakeholders can be woven into the core structure of the experimental design with direct stakeholders.

4.4.2.4 Multiplicity of and Potential Conflicts Among Human Values

Value Sensitive Design can help researchers uncover the multiplicity of and potential conflicts among human values implicated in technological implementations. In the above design space, for example, values of physical health, emotional well-being, and creativity appear to partially conflict with other values of privacy, civil rights, trust, and security.

4.4.2.5 Technical Investigations

Conceptual and empirical investigations can help to shape future technological investigations, particularly in terms of how nature (as a source of information) can be embedded in the design of display technologies to further human well-being. One obvious design space involves buildings. For example, if Kahn et al.'s empirical results continue to emerge in line with their initial results, then one possible design guideline is as follows: we need to design buildings with nature in mind, and within view. In other words, we cannot with psychological impunity digitize nature and display the digitized version as a substitute for the real thing (and worse, then destroy the original). At the same time, it is possible that technological representations of nature can garner some psychological benefits, especially when (as in an inside office) direct access to nature is otherwise unavailable. Other less obvious design spaces involve, for example, airplanes. In recent discussions with Boeing Corporation, for example, we were told that for economic reasons engineers might like to construct airplanes without passenger windows. After all, windows cost more to build and decrease fuel efficiency. At stake, however, is the importance of windows in the human experience of flying.

In short, this case study highlights how Value Sensitive Design can help researchers employ multiple psychological methods, across several studies, with direct and indirect stakeholders, to investigate (and ultimately support) a multiplicity of human values impacted by deploying a cutting-edge information technology.

4.4.3 *UrbanSim: Integrated Land Use, Transportation, and Environmental Simulation*

In many regions in the United States (and globally), there is increasing concern about pollution, traffic jams, resource consumption, loss of open space, loss of coherent community, lack of sustainability, and unchecked sprawl. Elected officials, planners, and citizens in urban areas grapple with these difficult issues as they develop and evaluate alternatives for such decisions as building a new rail line or freeway, establishing an urban growth boundary, or changing incentives or taxes. These decisions interact in complex ways, and, in particular, transportation and land use decisions interact strongly with each other. There are both legal and common sense reasons to try to understand the long-term consequences of these interactions and decisions. Unfortunately, the need for this understanding far outstrips the capability of the analytic tools used in current practice.

In response to this need, Waddell, Borning, and their colleagues have been developing UrbanSim, a large simulation package for predicting patterns of urban development for periods of 20 years or more, under different possible scenarios (Waddell 2002; Noth et al. 2003; Waddell et al. 2003; Borning et al. 2008). Its primary purpose is to provide urban planners and other stakeholders with tools to aid in more informed decision-making, with a secondary goal to support further democratization of the planning process. When provided with different scenarios – packages of possible policies and investments – UrbanSim models the resulting patterns of urban growth and redevelopment, of transportation usage, and of resource consumption and other environmental impacts.

As of early 2007, UrbanSim has been applied (either experimentally or in some cases transitioning to operational use) in the metropolitan regions in the U.S. around Detroit, El Paso, Eugene/Springfield, Oregon (Fig. 4.3), Honolulu, Houston, Salt Lake City, and Seattle; and internationally in Amsterdam, Paris, Tel Aviv, and Zurich. Additional projects have been launched in Burlington, Durham, Phoenix, and San Francisco, and internationally in Melbourne, Australia. Value Sensitive Design has played a central role in the ongoing design and implementation of interactions around UrbanSim indicators. UrbanSim illustrates important aspects of Value Sensitive Design in addition to those described in the previous two case studies:

4.4.3.1 Distinguishing Explicitly Supported Values from Stakeholder Values

In their conceptual investigations, Borning et al. (2005) distinguished between explicitly supported values (i.e., ones that they explicitly want to embed in the simulation) and stakeholder values (i.e., ones that are important to some but not necessarily all of the stakeholders). Next, Borning et al. committed to three specific moral values to be supported explicitly. One is fairness, and more specifically freedom from bias. The simulation should not discriminate unfairly against any group of stakeholders, or privilege one mode of transportation or policy over another.

Fig. 4.3 Results from UrbanSim for Eugene/Springfield, Oregon, forecasting land use patterns over a 14-year period. These results arise from the simulated interactions among demographic change, economic change, real estate development, transportation, and other actors and processes in the urban environment. Map (**a**) shows the employment density in 1980 (number of jobs located in each 150×150 m grid cell). *Darker shade* indicates higher density. Map (**b**) shows the predicted change from 1980 to 1994 (where *darker shade* indicates a greater change), and map (**c**) the predicted employment density in 1994. In a historical validation of the model, this result was then compared with the actual 1994 employment, with a 0.917 correlation over a 1-cell radius

A second is accountability. Insofar as possible, stakeholders should be able to confirm that their values are reflected in the simulation, evaluate and judge its validity, and develop an appropriate level of confidence in its output. The third is democracy. The simulation should support the democratic process in the context of land use, transportation, and environmental planning. In turn, as part of supporting the democratic process, Borning et al. decided that the model should not a priori favor or rule out any given set of stakeholder values, but instead, should allow different stakeholders to articulate the values that are most important to them, and evaluate the alternatives in light of these values.

4.4.3.2 Handling Widely Divergent and Potentially Conflicting Stakeholder Values

From the standpoint of conceptual investigations, UrbanSim as a design space poses tremendous challenges. The research team cannot focus on a few key values, as occurred in the Web Browser project (e.g., the value of informed consent), or the Room with a View project (e.g., the values of privacy in public spaces, and physical and psychological well-being). Rather, disputing stakeholders bring to the table widely divergent values about environmental, political, moral, and personal issues. Examples of stakeholder values are environmental sustainability, walkable neighborhoods, space for business expansion, affordable housing, freight mobility, minimal government intervention, minimal commute time, open space preservation, property rights, and environmental justice. How does one characterize the wide-ranging and deeply held values of diverse stakeholders, both present and future? Moreover, how does one prioritize the values implicated in the decisions? And how can one move from values to measurable outputs from the simulation to allow stakeholders to compare alternative scenarios?

As part of addressing these questions, the research group implemented a web-based interface that groups indicators into three broad value categories pertaining to the domain of urban development (economic, environmental, and social), and more specific value categories under that. To allow stakeholders to evaluate alternative urban futures, the interface provides a large collection of *indicators*: variables that distill some attribute of interest about the results (Gallopin 1997). (Examples of indicators are the number of acres of rural land converted to urban use each year, the degree of poverty segregation, or the mode share between autos and transit.) These categories and indicators draw on a variety of sources, including empirical research on people's environmental concepts and values (Kahn 1999; Kahn and Kellert 2002), community-based indicator projects (Palmer 1998; Hart 1999), and the policy literature. Stakeholders can then use the interface to select indicators that speak to values that are important to them from among these categories.

This interface illustrates the interplay among conceptual, technical, and empirical investigations. The indicators are chosen to speak to different stakeholder values – responding to our distinction between explicitly supported values and stakeholder values in the initial conceptual investigation. The value categories are rooted empirically in both human psychology and policy studies, not just philosophy – and then embodied in a technical artifact (the web-based interface), which is in turn evaluated empirically.

4.4.3.3 Legitimation

As we continued our work on VSD and UrbanSim, in our conceptual investigations we identified *legitimation* as a key instrumental value (Borning et al. 2005; Davis 2006). UrbanSim's legitimacy is crucial for its effective use in the planning process: stakeholders who do not see its use as legitimate may disengage from its use, or if they remain in the process, may never accept the analyses that it informs. Our conceptualization of legitimation draws primarily on the work of Jürgen Habermas (1979, 1984). Since the legitimacy of an urban planning process depends on a huge number of factors – most of which are outside of UrbanSim's scope – we concern ourselves with the *legitimation potential* of the modeling system. Again following Habermas, *communicative action* plays a key role in legitimation potential. The implicit validity claims of an utterance in a communicative act lead to a set of testable design goals for the system regarding comprehensibility, validity, transparency, and freedom from bias, which we then used in structuring our empirical investigations.

4.4.3.4 Technical Choices Driven by Initial and Emergent Value Considerations

Most of the technical choices in the design of the UrbanSim software are in response to the need to generate indicators and other evaluation measures that respond to different strongly-held stakeholder values. For example, for some stakeholders, walkable, pedestrian-friendly neighborhoods are very important. But being able to model walking as a transportation mode makes difficult demands on the underlying simulation, requiring a finer-grained spatial scale than is needed for modeling automobile transportation alone. In turn, being able to answer questions about walking as a transportation mode is important for two explicitly supported values: fairness (not to privilege one transportation mode over another), and democracy (being able to answer questions about a value that is important to a significant number of stakeholders). As a second example of technical choices being driven by value considerations, UrbanSim's software architecture is designed to support rapid evolution in response to changed or additional requirements. For instance, the software architecture decouples the individual component models as much as possible, allowing them to evolve and new ones to be added in a modular fashion. Further, the architecture separates out the computation of indicator values from the models, making it easy to write new indicators as needed, rather than embedding the indicator code in the component models themselves. For similar reasons, the UrbanSim team uses the YP agile software development methodology (Freeman-Benson and Borning 2003), which allows the system to evolve and respond quickly to emerging stakeholder values and policy considerations.

4.4.3.5 Designing for Credibility, Openness, and Accountability

The credibility of the system is of great importance, particularly when the system is being used in a politically charged situation and is thus the subject of intense scrutiny. The research group has undertaken a variety of activities to help foster credibility, including using behaviorally transparent simulation techniques (i.e., simulating agents in the urban environment, such as households, businesses, and real estate developers, rather than using some more abstract and opaque simulation technique), and performing sensitivity analyses (Franklin et al. 2002) and a historical validation. In the historical validation, for example, the group started the model with 1980 data from Eugene/Springfield, simulated through 1994, and compared the simulation output with what actually happened. One of these comparisons is shown in Fig. 4.3. In addition, our techniques for fostering openness and accountability are also intended to support credibility. These include using Open Source software (releasing the source code along with the executable), writing the code in as clear and understandable a fashion as possible, using a rigorous and extensive testing methodology, and complementing the Open Source software with an Open Process that makes the state of our development visible to anyone interested. For example, in our laboratory, a battery of tests is run whenever a new version of the software is committed to the source code repository. A traffic light (a real one) is activated by the testing regime – green means that the system has passed all tests, yellow means testing is under way, and red means that a test has failed. There is also a virtual traffic light, mirroring the physical one, visible on the web (www.urbansim.org/fireman). Similarly, the bug reports, feature requests, and plans are all on the UrbanSim project website as well. Details of this Open Process approach may be found in Freeman-Benson and Borning (2003).

For interactions around indicators, one project has been carefully documenting the available indicators and their limitations (Borning et al. 2005), including using "live documentation" that directly includes the source code used to compute the indicator values, and tests of that source code. Another project has involved partnering with different community organizations to produce "Indicator Perspectives" that provide different views on which indicators are most important, and how they should be evaluated (Schwartzman and Borning 2007). Finally, Janet Davis's Ph.D. dissertation (2006) describes the design and implementation of "personal indicators", which help users answer the question "how will this policy affect me and my family?", in addition to the more region-level results from the existing indicator sets.

Thus, in summary, Borning et al. are using Value Sensitive Design to investigate how a technology – an integrated land use, transportation, and environmental computer simulation – affects human values on both the individual and organizational levels; and how human values can continue to drive the technical investigations, including refining the simulation, data, and interaction model. Finally, employing Value Sensitive Design in a project of this scope serves to validate its use for complex, large-scale systems.

4.5 Value Sensitive Design's Constellation of Features

Value Sensitive Design shares and adopts many interests and techniques from related approaches to values and system design – computer ethics, social informatics, CSCW, and Participatory Design – as discussed in Sect. 4.2.2. However, Value Sensitive Design itself brings forward a unique constellation of eight features.

First, Value Sensitive Design seeks to be proactive: to influence the design of technology early in and throughout the design process.

Second, Value Sensitive Design enlarges the arena in which values arise to include not only the work place (as traditionally in the field of CSCW), but also education, the home, commerce, online communities, and public life.

Third, Value Sensitive Design contributes a unique methodology that employs conceptual, empirical, and technical investigations, applied iteratively and integratively (see Sect. 4.3).

Fourth, Value Sensitive Design enlarges the scope of human values beyond those of cooperation (CSCW) and participation and democracy (Participatory Design) to include all values, especially those with moral import. By moral, we refer to issues that pertain to fairness, justice, human welfare and virtue, encompassing within moral philosophical theory deontology (Dworkin 1978; Gewirth 1978; Kant 1785/1964; Rawls 1971), consequentialism ((Smart and Williams 1973); see Scheffler (1982) for an analysis), and virtue (Foot 1978; MacIntyre 1984; Campbell and Christopher 1996). Value Sensitive Design also accounts for conventions (e.g., standardization of protocols) and personal values (e.g., color preferences within a graphical user interface).

Fifth, Value Sensitive Design distinguishes between usability and human values with ethical import. Usability refers to characteristics of a system that make it work in a functional sense, including that it is easy to use, easy to learn, consistent, and recovers easily from errors (Adler and Winograd 1992; Norman 1988; Nielsen 1993). However, not all highly usable systems support ethical values. Nielsen (1993), for example, asks us to imagine a computer system that checks for fraudulent applications of people who are applying for unemployment benefits by asking applicants numerous personal questions, and then checking for inconsistencies in their responses. Nielsen's point is that even if the system receives high usability scores some people may not find the system socially acceptable, based on the moral value of privacy.

Sixth, Value Sensitive Design identifies and takes seriously two classes of stakeholders: direct and indirect. Direct stakeholders refer to parties – individuals or organizations – who interact directly with the computer system or its output. Indirect stakeholders refer to all other parties who are affected by the use of the system. Often, indirect stakeholders are ignored in the design process. For example, computerized medical records systems have often been designed with many of the direct stakeholders in mind (e.g., insurance companies, hospitals, doctors, and nurses), but with too little regard for the values, such as the value of privacy, of a rather important group of indirect stakeholders: the patients.

Seventh, Value Sensitive Design is an interactional theory: values are viewed neither as inscribed into technology (an endogenous theory), nor as simply transmitted by social forces (an exogenous theory). Rather, the interactional position holds that while the features or properties that people design into technologies more readily support certain values and hinder others, the technology's actual use depends on the goals of the people interacting with it. A screwdriver, after all, is well-suited for turning screws, and is also amenable to use as a poker, pry bar, nail set, cutting device, and tool to dig up weeds, but functions poorly as a ladle, pillow, or wheel. Similarly, an online calendar system that displays individuals' scheduled events in detail readily supports accountability within an organization but makes privacy difficult. Moreover, through human interaction, technology itself changes over time. On occasion, such changes (as emphasized in the exogenous position) can mean the societal rejection of a technology, or that its acceptance is delayed. But more often it entails an iterative process whereby technologies are first invented, and then redesigned based on user interactions, which then are reintroduced to users, further interactions occur, and further redesigns implemented. Typical software updates (e.g., of word processors, browsers, and operating systems) epitomize this iterative process.

Eighth, Value Sensitive Design builds from the psychological proposition that certain values are universally held, although how such values play out in a particular culture at a particular point in time can vary considerably (Kahn 1999; Turiel 1998, 2002). For example, even while living in an igloo, Inuits have conventions that ensure some forms of privacy; yet such forms of privacy are not maintained by separated rooms, as they are in most Western cultures. Generally, the more concretely (act-based) one conceptualizes a value, the more one will be led to recognizing cultural variation; conversely, the more abstractly one conceptualizes a value, the more one will be led to recognizing universals. Value Sensitive Design seeks to work both levels, the concrete and abstract, depending on the design problem at hand. Note that this is an empirical proposition, based on a large amount of psychological and anthropological data, not a philosophical one. We also make this claim only for certain values, not all – there are clearly some values that are culture-specific.

The three case studies presented in Sect. 4.5 illustrate the different features in this constellation. For example, UrbanSim illustrates the goal of being proactive and influencing the design of the technology early in and throughout the design process (Feature 1), and also involves enlarging the arena in which values arise to include urban planning and democratic participation in public decision-making (Feature 2). The cookies work is a good illustration of Value Sensitive Design's tripartite methodology (Feature 3): conceptual, technical, and empirical investigations, applied iteratively and integratively, were essential to the success of the project. Each of the three projects brings out a different set of human values (Feature 4): among others, informed consent for the cookies work; physical and psychological well-being and privacy in public spaces for Room with a View; and fairness, accountability, and democracy for UrbanSim, as well as the whole range of different sometimes competing stakeholder values. The cookies project illustrates the complex interaction between usability and human values (Feature 5): early versions of the system supported informed consent at the expense of usability, requiring additional work to develop a system that was both usable and provided reasonable support for informed consent. The Room with a View work considers and takes seriously both direct and indirect stakeholders (Feature 6): the occupants of the inside office ("The Watchers"), and passers-by in the plaza ("The Watched"). Value Sensitive Design's position that values are neither inscribed into technology nor simply transmitted by social forces (Feature 7) is illustrated by UrbanSim: the system by itself is certainly not neutral with respect to democratic process, but at the same time does not on its own ensure democratic decision-making on land use and transportation issues. Finally, the proposition that certain values are universally held, but play out in very different ways in different cultures and different times (Feature 8) is illustrated by the Room with a View project: the work is informed by a substantial body of work on the importance of privacy in all cultures (for example, the deep connection between privacy and self-identity), but concerns about privacy in public spaces play out in a specific way in the United States, and might do so quite differently in another cultural context.

We could draw out additional examples that illustrate Value Sensitive Design's constellation of features, both from the three case studies presented in Sect. 4.5, and in other projects; but hope that this short description demonstrates the unique contribution that Value Sensitive Design can make to the design of technology.

4.6 Practical Suggestions for Using Value Sensitive Design

One natural question with Value Sensitive Design is, "How exactly do I do it?" In this section we offer some practical suggestions.

4.6.1 *Start with a Value, Technology, or Context of Use*

Any of these three core aspects – a value, technology, or context of use – easily motivates Value Sensitive Design. We suggest starting with the aspect that is most central to your work and interests. In the case of Informed Consent and Cookies, for example, Friedman et al. began with a value of central interest (informed consent) and moved from that value to its implications for Web browser design. In the case of UrbanSim, Borning et al. began with a technology (urban simulation) and a context of use (the urban planning process); upon inspection of those two, values issues quickly came to the fore.

4.6.2 *Identify Direct and Indirect Stakeholders*

As part of the initial conceptual investigation, systematically identify direct and indirect stakeholders. Recall that direct stakeholders are those individuals who interact directly with the technology or with the technology's output. Indirect stakeholders are those individuals who are also impacted by the system, though they never interact directly with it. In addition, it is worthwhile to recognize the following:

- Within each of these two overarching categories of stakeholders, there may be several subgroups.
- A single individual may be a member of more than one stakeholder group or subgroup. For example, in the UrbanSim project, an individual who works as an urban planner and lives in the area is both a direct stakeholder (i.e., through his or her direct use of the simulation to evaluate proposed transportation plans) and an indirect stakeholder (i.e., by virtue of living in the community for which the transportation plans will be implemented).
- An organizational power structure is often orthogonal to the distinction between direct and indirect stakeholders. For example, there might be low-level employees who are either direct or indirect stakeholders and who don't have control over using the system (e.g., workers on an assembly line). Participatory Design has contributed a substantial body of analysis to these issues, as well as techniques for dealing with them, such as ways of equalizing power among groups with unequal power. (See the references cited in Sect. 4.2.2.)

4.6.3 *Identify Benefits and Harms for Each Stakeholder Group*

Having identified the key stakeholders, systematically identify the benefits and harms for each group. In doing so, we suggest attention to the following points:

- Indirect stakeholders will be benefited or harmed to varying degrees; and in some designs it is probably possible to claim every human as an indirect stakeholder of some sort. Thus, one rule of thumb in the conceptual investigation is to give priority to indirect stakeholders who are strongly affected, or to large groups that are somewhat affected.

- Attend to issues of technical, cognitive, and physical competency. For example, children or the elderly might have limited cognitive competency. In such a case, care must be taken to ensure that their interests are represented in the design process, either by representatives from the affected groups themselves or, if this is not possible, by advocates.
- Personas (Pruitt and Grudin 2003) are a popular technique that can be useful for identifying the benefits and harms to each stakeholder group. However, we note two caveats. First, personas have a tendency to lead to stereotypes because they require a list of "socially coherent" attributes to be associated with the "imagined individual." Second, while in the literature each persona represents a different user group, in Value Sensitive Design (as noted above) the same individual may be a member of more than one stakeholder group. Thus, in our practice, we have deviated from the typical use of personas that maps a single persona onto a single user group, to allow for a single persona to map onto to multiple stakeholder groups.

4.6.4 Map Benefits and Harms onto Corresponding Values

With a list of benefits and harms in hand, one is in a strong position to recognize corresponding values. Sometimes the mapping is one of identity. For example, a harm that is characterized as invasion of privacy maps onto the value of privacy. Other times the mapping is less direct if not multifaceted. For example, with the Room with a View study, it is possible that a direct stakeholder's mood is improved when working in an office with an augmented window (as compared with no window). Such a benefit potentially implicates not only the value of psychological welfare, but also creativity, productivity, and physical welfare (health), assuming there is a causal link between improved mood and these other factors.

In some cases, the corresponding values will be obvious, but not always. Table 4.1 in Sect. 4.2.2 provides a table of human values with ethical import often implicated in system design. This table may be useful in suggesting values that should be considered in the investigation.

4.6.5 Conduct a Conceptual Investigation of Key Values

Following the identification of key values in play, a conceptual investigation of each can follow. Here it is helpful to turn to the relevant literature. In particular, the philosophical ontological literature can help provide criteria for what a value is, and thereby how to assess it empirically. (For example, Sect. 4.4.1.1 described how existing literature helped provide criteria for the value of informed consent.)

4.6.6 Identify Potential Value Conflicts

Values often come into conflict. Thus, once key values have been identified and carefully defined, a next step entails examining potential conflicts. For the purposes of design, value conflicts should usually not be conceived of as “either/or” situations, but as constraints on the design space. Admittedly, at times designs that support one value directly hinder support for another. In those instances, a good deal of discussion among the stakeholders may be warranted to identify the space of workable solutions. Typical value conflicts include accountability vs. privacy, trust vs. security, environmental sustainability vs. economic development, privacy vs. security, and hierarchical control vs. democratization.

4.6.7 Integrate Value Considerations into One’s Organizational Structure

Ideally, Value Sensitive Design will work in concert with organizational objectives. Within a company, for example, designers would bring values into the forefront, and in the process generate increased revenue, employee satisfaction, customer loyalty, and other desirable outcomes for their companies. In turn, within a government agency, designers would both better support national and community values, and enhance the organization’s ability to achieve its objectives. In the real world, of course, human values (especially those with ethical import) may collide with economic objectives, power, and other factors. However, even in such situations, Value Sensitive Design should be able to make positive contributions, by showing alternate designs that better support enduring human values. For example, if a standards committee were considering adopting a protocol that raised serious privacy concerns, a Value Sensitive Design analysis and design might result in an alternate protocol that better addressed the issue of privacy while still retaining other needed properties. Citizens, advocacy groups, staff members, politicians, and others could then have a more effective argument against a claim that the proposed protocol was the only reasonable choice.

4.6.8 Human Values (with Ethical Import) Often Implicated in System Design

We stated earlier that while all values fall within its purview, Value Sensitive Design emphasizes values with ethical import. In Table 4.1 in Sect. 4.2.2, we present a list of frequently implicated values. This table is intended as a heuristic for suggesting values that should be considered in the investigation – it is definitely not intended as a complete list of human values that might be implicated.

Two caveats. First, not all of these values are fundamentally distinct from one another. Nonetheless, each value has its own language and conceptualizations within its respective field, and thus warrants separate treatment here. Second, as noted above, this list is not comprehensive. Perhaps no list could be, at least within the confines of a paper. Peacefulness, respect, compassion, love, warmth, creativity, humor, originality, vision, friendship, cooperation, collaboration, purposefulness, devotion, loyalty, diplomacy, kindness, musicality, harmony – the list of other possible moral and non-moral values could get very long very quickly. Our particular list comprises many of the values that hinge on the deontological and consequentialist moral orientations noted above: human welfare, ownership and property, privacy, freedom from bias, universal usability, trust, autonomy, informed consent, and accountability. In addition, we have chosen several other values related to system design: courtesy, identity, calmness, and environmental sustainability.

4.6.9 Heuristics for Interviewing Stakeholders

As part of an empirical investigation, it is useful to interview stakeholders, to better understand their judgments about a context of use, an existing technology, or a proposed design. A semi-structured interview often offers a good balance between addressing the questions of interest and gathering new and unexpected insights. In these interviews, the following heuristics can prove useful:

In probing stakeholders' reasons for their judgments, the simple question "Why?" can go a good distance. For example, seniors evaluating a ubiquitous computing video surveillance system might respond negatively to the system. When asked "Why?" a response might be: "I don't mind my family knowing that other people are visiting me, so they don't worry that I'm alone – I just don't want them to know who is visiting." The researcher can probe again: "Why don't you want them to know?" An answer might be: "I might have a new friend I don't want them to know about. It's not their business." Here the first "why" question elicits information about a value conflict (the family's desire to know about the senior's well-being and the senior's desire to control some information); the second "why" question elicits further information about the value of privacy for the senior.

Ask about values not only directly, but indirectly, based on formal criteria specified in the conceptual investigation. For example, suppose that you want to conduct an empirical investigation of people's reasoning and values about "X" (say, trust, privacy, or informed consent), and that you decided to employ an interview methodology. One option is to ask people directly about the topic. "What is X?" "How do you reason about X?" "Can you give me an example from your own life of when you encountered a problem that involved X?" There is some merit to this direct approach. Certainly it gives people the opportunity to define the problem in their own terms. But you may quickly discover that it comes up short. Perhaps the greatest problem is that people have concepts about many aspects of the topic on which they cannot directly reflect. Rather, you will usually be better served by employing an alternative approach. As is common in social cognitive research (see Kahn (1999), chap. 5, for a discussion of methods), you could interview people about a hypothetical situation, or a common everyday event in their lives, or a task that you have asked them to solve, or a behavior in which they have just engaged. But, no matter what you choose, the important point is *a priori* to conceptualize what the topic entails, if possible demarcating its boundaries through formal criteria, and at a minimum employing issues or tasks that engage people's reasoning about the topic under investigation.

4.6.10 Heuristics for Technical Investigations

When engaging in value-oriented technical investigations, the following heuristics can prove useful:

Technical mechanisms will often adjudicate multiple if not conflicting values, often in the form of design trade-offs. We have found it helpful to make explicit how a design trade-off maps onto a value conflict and differentially affects different groups of stakeholders. For example, the Room with a View study suggests real-time displays in interior offices may provide physiological benefits for those in the inside offices (the direct stakeholders), yet may impinge on the privacy and security of those walking through the outdoor scene (the indirect stakeholders), and especially women.

Unanticipated values and value conflicts often emerge after a system is developed and deployed. Thus, when possible, design flexibility into the underlying technical architecture so that it can be responsive to such emergent concerns. In UrbanSim, for example, Borning et al. used agile programming techniques to design an architecture that can more readily accommodate new indicators and models.

The control of information flow through underlying protocols – and the privacy concerns surrounding such control – is a strongly contested area. Ubiquitous computing, with sensors that collect and then disseminate information at large, has only intensified these concerns. We suggest that underlying protocols that release information should be able to be turned off (and in such a way that the stakeholders are confident they have been turned off).

4.7 Conclusion

There is a growing interest and challenge to address values in design. Our goal in this paper has been to provide enough detail about Value Sensitive Design so that other researchers and designers can critically examine, use, and extend this approach. Our hope is that this approach can contribute to a principled and comprehensive consideration of values in the design of information and computational systems.

Acknowledgements Value Sensitive Design has emerged over the past decade and benefited from discussions with many people. We would like particularly to acknowledge all the members of our respective research groups, along with Edward Felten, Jonathan Grudin, Sara Kiesler, Clifford Nass, Helen Nissenbaum, John Thomas, and Terry Winograd. This research was supported in part by NSF Awards IIS-9911185, IIS-0325035, EIA-0121326, and EIA-0090832.

Notes The Oxford English Dictionary definition of this sense of value is: "the principles or standards of a person or society, the personal or societal judgement of what is valuable and important in life" (Simpson and Wiener 1989).

References

Aberg, J., & Shahmehri, N. (2001). An empirical study of human web assistants: Implications for user support in Web information systems. In *Proceedings of the conference on human factors in computing systems (CHI 2000)* (pp. 404–411). New York: Association for Computing Machinery Press.

Ackerman, M. S., & Cranor, L. (1999). Privacy critics: UI components to safeguard users' privacy. In *Extended abstracts of CHI 1999* (pp. 258–259). New York: ACM Press.

Adler, P. S., & Winograd, T. (Eds.). (1992). *Usability: Turning technologies into tools*. Oxford: Oxford University Press.

Agre, P. E., & Rotenberg, M. (Eds.). (1998). *Technology and privacy: The new landscape*. Cambridge, MA: MIT Press.

Baier, A. (1986). Trust and antitrust. *Ethics, 92*, 231–260.

Beck, A., & Katcher, A. (1996). *Between pets and people*. West Lafayette: Purdue University Press.

Becker, L. C. (1977). *Property rights: Philosophical foundations*. London: Routledge & Kegan Paul.

Bellotti, V. (1998). Design for privacy in multimedia computing and communications environments. In P. E. Agre & M. Rotenberg (Eds.), *Technology and privacy: The new landscape* (pp. 63–98). Cambridge, MA: The MIT Press.

Bennet, W. J., & Delatree, E. J. (1978). Moral education in the schools. *The Public Interest, 50*, 81–98.

Bers, M. U., Gonzalez-Heydrich, J., & DeMaso, D. R. (2001). Identity construction environments: Supporting a virtual therapeutic community of pediatric patients undergoing dialysis. In *Proceedings of the conference of human factors in computing systems (CHI 2001)* (pp. 380–387). New York: Association for Computing Machinery.

Bjerknes, G., & Bratteteig, T. (1995). User participation and democracy: A discussion of Scandinavian research on system development. *Scandinavian Journal of Information Systems, 7*(1), 73–97.

Bødker, S. (1990). *Through the interface – A human activity approach to user interface design*. Hillsdale: Lawrence Erlbaum Associates.

Borning, A., Friedman, B., Davis, J., & Lin, P. (2005, January). Informing public deliberation: Value sensitive design of indicators for a large-scale urban simulation. In *ECSCW 2005* (pp. 449–468). Dordrecht: Springer.

Borning, A., Waddell, P., & Förster, R. (2008). UrbanSim: Using simulation to inform public deliberation and decision-making. In *Digital government* (pp. 439–464). New York: Springer.

Boyle, M., Edwards, C., & Greenberg, S. (2000). The effects of filtered video on awareness and privacy. In *Proceedings of conference on computer supported cooperative work (CSCW 2000)* (pp. 1–10). New York: Association for Computing Machinery.

Bynum, T. W. (Ed.). (1985). *Metaphilosophy, 16*(4), 263–377. [Entire issue.]

Camp, L. J. (2000). *Trust & risk in internet commerce*. Cambridge, MA: MIT Press.

Campbell, R. L., & Christopher, J. C. (1996). Moral development theory: A critique of its Kantian presuppositions. *Developmental Review, 16*, 1–47.

Carroll, J. M., & Rosson, M. B. (2006). Dimensions of participation in information system design. In P. Zhang & D. Galletta (Eds.), *Human-computer interaction and management information* *systems: Applications* (Advances in management information systems, pp. 337–354). Armonk: M.E. Sharpe.

Cooper, M., & Rejmer, P. (2001). Case study: Localization of an accessibility evaluation. In *Extended abstracts of the conference on human factors in computing systems (CHI 2001)* (pp. 141–142). New York: Association for Computing Machinery Press.

Davis, J. (2006, August). *Value sensitive design of interactions with UrbanSim indicators*. Ph.D. dissertation, Department of Computer Science & Engineering, University of Washington.

Dieberger, A., Hook, K., Svensson, M., & Lonnqvist, P. (2001). Social navigation research agenda. In *Extended abstracts of the conference on human factors in computing systems (CHI 2001)* (pp. 107–108). New York: Association of Computing Machinery Press.

Dworkin, R. (1978). *Taking rights seriously*. Cambridge, MA: Harvard University Press.

Egger, F. N. (2000). "Trust me, I'm an online vendor": Towards a model of trust for e-commerce system design. In *Extended abstracts of the conference of human factors in computing systems (CHI 2000)* (pp. 101–102). New York: Association for Computing Machinery.

Ehn, P. (1989). *Work-oriented design of computer artifacts*. Hillsdale: Lawrence Erlbaum Associates.

Faden, R., & Beauchamp, T. (1986). *A history and theory of informed consent*. New York: Oxford University Press.

Federal Trade Commission. (2000, May). *Privacy online: Fair information practices in the electronic marketplace*. A Report to Congress. Washington, DC: Federal Trade Commission.

Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. In *Proceedings of CHI 1999* (pp. 80–87). Cambridge, MA: ACM Press.

Foot, P. (1978). *Virtues and vices*. Berkeley/Los Angeles: University of California Press.

Frankena, W. (1972). Value and valuation. In P. Edwards (Ed.), *The encyclopedia of philosophy* (Vol. 7–8, pp. 409–410). New York: Macmillan.

Franklin, J., Waddell, P., & Britting, J. (2002, November 21–24). *Sensitivity analysis approach for an Integrated Land Development & Travel Demand Modeling System*. Presented at the Association of Collegiate Schools of Planning 44th annual conference, Baltimore. Preprint available from www.urbansim.org

Freeman-Benson, B. N., & Borning, A. (2003, June). YP and urban simulation: Applying an agile programming methodology in a politically tempestuous domain. In *Proceedings of the 2003 Agile Programming Conference*, Salt Lake City. Preprint available from www.urbansim.org

Friedman, B. (Ed.). (1997a). *Human values and the design of computer technology*. New York: Cambridge University Press.

Friedman, B. (1997b). Social judgments and technological innovation: Adolescents' understanding of property, privacy, and electronic information. *Computers in Human Behavior, 13*(3), 327–351.

Friedman, B., & Kahn, P. H., Jr. (1992). Human agency and responsible computing: Implications for computer system design. *Journal of Systems Software, 17*, 7–14.

Friedman, B., & Kahn, P. H., Jr. (2003). Human values, ethics, and design. In J. Jacko & A. Sears (Eds.), *The human-computer interaction handbook*. Mahwah: Lawrence Erlbaum Associates.

Friedman, B., & Millett, L. (1995). "It's the computer's fault" – Reasoning about computers as moral agents. In *Conference companion of the conference on human factors in computing systems (CHI 95)* (pp. 226–227). New York: Association for Computing Machinery Press.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems, 14*(3), 330–347.

Friedman, B., & Nissenbaum, H. (1997). Software agents and user autonomy. In *Proceedings of the first international conference on autonomous agents* (pp. 466–469). New York: Association for Computing Machinery Press.

Friedman, B., Kahn, P. H., Jr., & Howe, D. C. (2000a). Trust online. *Communications of the ACM, 43*(12), 34–40.

Friedman, B., Millett, L., & Felten, E. (2000b). *Informed consent online: A conceptual model and design principles*. University of Washington Computer Science & Engineering Technical Report 00-12-2.

Friedman, B., Howe, D. C., & Felten, E. (2002, January). Informed consent in the Mozilla browser: Implementing value-sensitive design. In *HICSS. Proceedings of the 35th annual Hawaii international conference on system sciences, 2002* (pp. 247–257). IEEE.

Friedman, B., Kahn, P. H., Jr., & Hagman, J. (2003). Hardware companions?: What online AIBO discussion forums reveal about the human-robotic relationship. In *Conference proceedings of CHI 2003* (pp. 273–280). New York: ACM Press.

Friedman, B., Kahn, P. H., Jr., Hagman, J., Severson, R. L., & Gill, B. (2006). The watcher and the watched: Social judgments about privacy in a public place. *Human-Computer Interaction, 21*(2), 235–272.

Fuchs, L. (1999). AREA: A cross-application notification service for groupware. In *Proceedings of ECSCW 1999* (pp. 61–80). Dordrecht: Kluwer.

Galegher, J., Kraut, R. E., & Egido, C. (Eds.). (1990). *Intellectual teamwork: Social and technological foundations of cooperative work*. Hillsdale: Lawrence Erlbaum Associates.

Gallopin, G. C. (1997). Indicators and their use: Information for decision-making. In B. Moldan, S. Billharz, & R. Matravers (Eds.), *Sustainability indicators: A report on the project on indicators of sustainable development*. Chichester: Wiley.

Gewirth, A. (1978). *Reason and morality*. Chicago: University of Chicago Press.

Greenbaum, J., & Kyng, M. (Eds.). (1991). *Design at work: Cooperative design of computer systems*. Hillsdale: Lawrence Erlbaum Associates.

Grudin, J. (1988). Why CSCW applications fail: Problems in the design and evaluation of organizational interfaces. In *Proceedings of the conference on computer supported cooperative work (CSCW '88)* (pp. 85–93). New York: Association for Computing Machinery Press.

Habermas, J. (1979). *Communication and the evolution of society* (trans: McCarthy, T.). Boston: Beacon Press.

Habermas, J. (1984). *The theory of communicative action,* Vol 1. (trans: McCarthy, T.). Boston: Beacon Press.

Hagman, J., Hendrickson, A., & Whitty, A. (2003). What's in a barcode: Informed consent and machine scannable driver licenses. In *CHI 2003 extended abstracts of the conference on human factors in computing system* (pp. 912–913). New York: ACM Press.

Harris Poll/Business Week. (2000). A growing threat. http://www.buisnessweek.come/2000/00_12/b3673010.htm

Hart, M. (1999). *Guide to sustainable community indicators*. Hart Environmental Data, PO Box 361, North Andover, MA 01845, second edition.

Herskovits, M. J. (1952). *Economic anthropology: A study of comparative economics*. New York: Knopf.

Hill, T. E., Jr. (1991). *Autonomy and self-respect*. Cambridge: Cambridge University Press.

Isaacs, E. A., Tang, J. C., & Morris, T. (1996). Piazza: A desktop environment supporting impromptu and planned interactions. In *Proceedings of the conference on computer supported cooperative work (CSCW 96)* (pp. 315–324). New York: Association for Computing Machinery Press.

Jacko, J. A., Dixon, M. A., Rosa, R. H., Jr., Scott, I. U., & Pappas, C. J. (1999). Visual profiles: A critical component of universal access. In *Proceedings of the conference on human factors in computing systems (CHI 99)* (pp. 330–337). New York: Association for Computing Machinery Press.

Jancke, G., Venolia, G. D., Grudin, J., Cadiz, J. J., & Gupta, A. (2001). Linking public spaces: Technical and social issues. In *Proceedings of CHI 2001* (pp. 530–537). New York: ACM Press.

Johnson, E. H. (2000). Getting beyond the simple assumptions of organization impact [social informatics]. *Bulletin of the American Society for Information Science, 26*(3), 18–19.

Johnson, D. G., & Miller, K. (1997). Ethical issues for computer scientists and engineers. In A. B. Tucker, Jr. (Ed.-in-Chief), *The computer science and engineering handbook* (pp. 16–26). Boca Raton: CRC Press.

Kahn, P. H., Jr. (1999). *The human relationship with nature: Development and culture*. Cambridge, MA: MIT Press.

Kahn, P. H., Jr., & Kellert, S. R. (Eds.). (2002). *Children and nature: Psychological, sociocultural, and evolutionary investigations*. Cambridge, MA: MIT Press.

Kahn, P. H., Jr., & Turiel, E. (1988). Children's conceptions of trust in the context of social expectations. *Merrill-Palmer Quarterly, 34*, 403–419.

Kant, I. (1964). *Groundwork of the metaphysic of morals* (trans: Paton, H. J.). New York: Harper Torchbooks. (Original work published 1785.)

Kling, R., & Star, S. L. (1998). Human centered systems in the perspective of organizational and social informatics. *Computers and Society, 28*(1), 22–29.

Kling, R., Rosenbaum, H., & Hert, C. (1998). Social informatics in information science: An introduction. *Journal of the American Society for Information Science, 49*(12), 1047–1052.

Kyng, M., & Mathiassen, L. (Eds.). (1997). *Computers and design in context*. Cambridge, MA: MIT Press.

Leveson, N. G. (1991). Software safety in embedded computer systems. *Communications of the ACM, 34*(2), 34–46.

Lipinski, T. A., & Britz, J. J. (2000). Rethinking the ownership of information in the 21st century: Ethical implications. *Ethics and Information Technology, 2*(1), 49–71.

MacIntyre, A. (1984). *After virtue*. Notre Dame: University of Notre Dame Press.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review, 20*(3), 709–734.

Millett, L., Friedman, B., & Felten, E. (2001). Cookies and web browser design: Toward realizing informed consent online. In *Proceedings of CHI 2001* (pp. 46–52). New York: ACM Press.

Moldan, B., Billharz, S., & Matravers, R. (Eds.). (1997). *Sustainability indicators: A report on the project on indicators of sustainable development*. Chichester: Wiley.

Moore, G. E. (1978). *Principia ethica*. Cambridge: Cambridge University Press. (Original work published 1903).

Nass, C., & Gong, L. (2000). Speech interfaces from an evolutionary perspective. *Communications of the ACM, 43*(9), 36–43.

Neumann, P. G. (1995). *Computer related risks*. New York: Association for Computing Machinery Press.

Nielsen, J. (1993). *Usability engineering*. Boston: AP Professional.

Nissenbaum, H. (1998). Protecting privacy in an information age: The problem with privacy in public. *Law and Philosophy, 17*, 559–596.

Nissenbaum, H. (1999). Can trust be secured online? A theoretical perspective. *Etica e Politca*, *2* (Electronic journal).

Nissenbaum, H. (2001). Securing trust online: Wisdom or oxymoron. *Boston University Law Review, 81*(3), 635–664.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Northwest Environment Watch (2002). *This place on earth 2002: Measuring what matters*. Northwest Environment Watch, 1402 Third Avenue, Seattle, WA 98101.

Noth, M., Borning, A., & Waddell, P. (2003). An extensible, modular architecture for simulating urban development, transportation, and environmental impacts. *Computers, Environment and Urban Systems, 27*(2), 181–203.

Olson, J. S., & Olson, G. M. (2000). i2i trust in e-commerce. *Communications of the ACM, 43*(12), 41–44.

Olson, J. S., & Teasley, S. (1996). Groupware in the wild: Lessons learned from a year of virtual collaboration. In *Proceedings of the conference on computer supported cooperative work (CSCW 96)* (pp. 419–427). New York: Association for Computing Machinery Press.

Orlikowski, W. J., & Iacono, C. S. (2001). Research commentary: Desperately seeking the "IT" in IT research—a call to theorizing the IT artifact. *Information Systems Research, 12*(2), 121–134.

Palen, L., & Dourish, P. (2003). Privacy and trust: Unpacking "privacy" for a networked world. In *Proceedings of CHI 2003* (pp. 129–136). New York: ACM Press.

Palen, L., & Grudin, J. (2003). Discretionary adoption of group support software: Lessons from calendar applications. In B. E. Munkvold (Ed.), *Implementing collaboration technologies in industry*. Heidelberg: Springer.

Palmer, K. (Ed.). (1998). *Indicators of sustainable community*. Seattle: Sustainable Seattle.

Phillips, D. J. (1998). Cryptography, secrets, and structuring of trust. In P. E. Agre & M. Rotenberg (Eds.), *Technology and privacy: The new landscape* (pp. 243–276). Cambridge, MA: MIT Press.

Pruitt, J., & Grudin, J. (2003). Personas: Practice and theory. In *Proceedings of DUX 2003*. Cambridge, MA: ACM Press.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York/Stanford: Cambridge University Press/CSLI Publications.

Riegelsberger, J., & Sasse, M. A. (2002). Face it – Photos don't make a web site trustworthy. In *Extended abstracts of CHI 2002* (pp. 742–743). New York: ACM Press.

Rocco, E. (1998). Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. In *Proceedings of CHI 1998* (pp. 496–502). New York: ACM Press.

Rosenberg, S. (1997). Multiplicity of selves. In R. D. Ashmore & L. Jussim (Eds.), *Self and identity: Fundamental issues* (pp. 23–45). New York: Oxford University Press.

Sawyer, S., & Rosenbaum, H. (2000). Social informatics in the information sciences: Current activities and emerging direction. *Informing Science, 3*(2), 89–95.

Scheffler, S. (1982). *The rejection of consequentialism*. Oxford: Oxford University Press.

Schiano, D. J., & White, S. (1998). The first noble truth of cyberspace: People are people (even when they MOO). In *Proceedings of the conference of human factors in computing systems (CHI 98)* (pp. 352–359). New York: Association for Computing Machinery.

Schneider, F. B. (Ed.). (1999). *Trust in cyberspace*. Washington, DC: National Academy Press.

Schoeman, F. D. (Ed.). (1984). *Philosophical dimensions of privacy: An anthology*. Cambridge: Cambridge University Press.

Schwartzman, Y., & Borning, A. (2007, January). The indicator browser: A web-based interface for visualizing UrbanSim simulation results. In *HICSS 2007. 40th annual Hawaii international conference on system sciences, 2007* (pp. 92–92). IEEE.

Shneiderman, B. (1999). *Universal usability: Pushing human-computer interaction research to empower every citizen* (ISR Technical Report 99–72). College Park: University of Maryland, Institute for Systems Research.

Shneiderman, B. (2000). Universal usability. *Communications of the ACM, 43*(5), 84–91.

Simpson, J. A., & Weiner, E. S. C. (Eds.). (1989). "value, *n*." *Oxford English Dictionary.* Oxford: Clarendon Press, 1989. *OED Online*. Oxford University Press. 30 May 2003. http://dictionary.oed.com/cgi/entry/00274678

Smart, J. J. C., & Williams, B. (1973). *Utilitarianism for and against*. Cambridge: Cambridge University Press.

Stephanidis, C. (Ed.). (2001). *User interfaces for all: Concepts, methods, and tools*. Mahwah: Lawrence Erlbaum Associates.

Suchman, L. (1994). Do categories have politics? The language/action perspective reconsidered. *CSCW Journal, 2*(3), 177–190.

Svensson, M., Hook, K., Laaksolahti, J., & Waern, A. (2001). Social navigation of food recipes. In *Proceedings of the conference of human factors in computing systems (CHI 2001)* (pp. 341–348). New York: Association for Computing Machinery.

Tang, J. C. (1997). Eliminating a hardware switch: Weighing economics and values in a design decision. In B. Friedman (Ed.), *Human values and the design of computer technology* (pp. 259–269). New York: Cambridge University Press.

The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. (1978). The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.

Thomas, J. C. (1997). Steps toward universal access within a communications company. In B. Friedman (Ed.), *Human values and the design of computer technology* (pp. 271–287). New York: Cambridge University Press.

Turiel, E. (1983). *The development of social knowledge*. Cambridge: Cambridge University Press.

Turiel, E. (1998). Moral development. In N. Eisenberg (Ed.), *Social, emotional, and personality development* (Vol. 3 of W. Damon, Ed., Handbook of child psychology. 5th ed. pp. 863–932). New York: Wiley.

Turiel, E. (2002). *The culture of morality: Social development, context, and conflict*. Cambridge: Cambridge University Press.

Turkle, S. (1996). *Life on the screen: Identify in the age of the internet*. New York: Simon and Schuster.

Ulrich, R. S. (1984). View through a window may influence recovery from surgery. *Science, 224*, 420–421.

Ulrich, R. S. (1993). Biophilia, biophobia, and natural landscapes. In S. R. Kellert & E. O. Wilson (Eds.), *The biophilia hypothesis* (pp. 73–137). Washington, DC: Island Press.

United Nations (2002). *Report of the United Nations Conference on Environment and Development, held in Rio de Janeiro*, Brazil, 1992. Available from http://www.un.org/esa/sustdev/documents/agenda21/english/agenda21toc.htm

Waddell, P. (2002). UrbanSim: Modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association, 68*(3), 297–314.

Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M., & Ulfarsson, G. (2003). Microsimulation of urban development and location choices: Design and implementation of UrbanSim. *Networks and Spatial Economics, 3*(1), 43–67.

Weiser, M., & Brown, J. S. (1997). The coming age of calm technology. In P. Denning & B. Metcalfe (Eds.), *Beyond calculation: The next 50 years of computing* (pp. 75–85). New York: Springer.

Weizenbaum, J. (1972). On the impact of the computer on society: How does one insult a machine? *Science, 178*, 609–614.

Wiener, N. (1985). The machine as threat and promise. In P. Masani (Ed.), *Norbert wiener: Collected works and commentaries* (Vol. IV, pp. 673–678). Cambridge, MA: MIT Press. (Reprinted from *St. Louis Post Dispatch*, 1953, December 13.).

Winograd, T. (1994). Categories, disciplines, and social coordination. *CSCW Journal, 2*(3), 191–197.

World Commission on Environment and Development (Gro Harlem Brundtland, Chair). (1987). *Our common future*. Oxford: Oxford University Press.

Wynne, E. A., & Ryan, K. (1993). *Reclaiming our schools: A handbook on teaching character, academics, and discipline*. New York: Macmillan.

Zheng, J., Bos, N., Olson, J., & Olson, G. M. (2001). Trust without touch: Jump-start trust with social chat. In *Extended abstracts of CHI 2001* (pp. 293–294). New York: ACM Press.

4.8 Addendum: Practical Considerations of Value Sensitive Design

4.8.1 Practical Value Sensitive Design Challenges

Value sensitive design (VSD) has evolved over time and proven its ability to guide developers and researchers in considering human values in the designs of their systems. Although applied successfully for almost 20 years in diverse projects focusing on different values, e.g. informed consent in browsers (see Sect. 4.4.1), independence for blind and deaf transit riders (Azenkot et al. 2011) or security of implantable medical devices (Denning et al. 2010), VSD is subject to critical and constructive reviews by its creators and other researchers. Learning from practice, reflecting on the methodological assumptions and creating new tools and methods is crucial for its further development and its widespread acceptance by researchers and practitioners.

In their recent work Borning and Muller (2012) have identified a number of issues in VSD research that could lead to a lack of adoption of VSD in value-focused

---

Alina Huldtgren
Department of Media, Duesseldorf University of Applied Sciences,
Josef-Gockeln-Str. 940474 Düsseldor, Germany
e-mail: alina.huldtgren@fh-duesseldorf.de

human computer interaction (HCI) research. Related critique has been presented earlier, e.g., from Le Dantec and colleagues (2009). In the following we will discuss three points that are relevant within the scope of this book: the nature of values that are considered in a project, the role of different stakeholders in the design process and concrete methods used in VSD.

4.8.1.1 Nature of Values

VSD refers to values as “what a person or group of people consider important in life.” (this chapter) While this definition is kept broad, Friedman and colleagues also provide “a list of human values with ethical import that are often implicated in system design”. When talking about ethical or moral values the question arises whether certain values are universal. In this chapter the authors state that VSD “builds from the psychological proposition that certain values are universally held” [p.4] which may play out differently in a given culture and time.

Borning and Muller (2012) discuss the problematic nature of taking a stance on the universality of values and conclude that the existence of universal values has little impact for the practical application of VSD. As values play out sufficiently different in each design context, universal designs that account for a certain value are not attainable. Instead, it is important that the values at stake are identified and analyzed carefully as well as defined with respect to the particular context and new design solutions for the given context have to be created.

With regard to identifying the values at stake, an important question that has recently been discussed within the VSD community is “should VSD single out certain values as particularly worthy of consideration?”

During the evolution of VSD, lists of values with ethical import have been presented from having “a distinctive claim on resources in the design process” (Friedman and Kahn 2003) to being heuristics for designers. Le Dantec and colleagues’ (2009) stance on this issue is that given lists of values may bias researchers and designers towards these values. While expressing classifications of ethically principled values was an important step, more scaffolding is needed to guide the value discovery, i.e. to uncover values as they are lived in-situ, through empirical exploration relevant to the design context. After these so-called local values have been discovered, lists can be used as an analytical tool.

In my opinion, heuristic lists could also be beneficial from the start, especially for practitioners with limited time at hand, as the lists highlight important values and mitigate the odds that these are overlooked.

Borning and Muller (2012) suggest that lists presented in the literature should be contextualized by emphasizing who wrote them and for what purpose. Additional careful empirical investigations can highlight stakeholder values that have not been considered initially by the design team. In this respect an important distinction is to be made among *explicitly supported values* (i.e., ones that the system is designed to support), *stakeholder* values (i.e., ones that are important to some but not necessarily all of the stakeholders) and *designer values* (i.e., ones that the system designers hold).

4.8.1.2 Role of Stakeholders

When defining whose values are accounted for, the stakeholder concept is the focus of attention. Stakeholders in VSD are not only the clients or end-users, but all people involved directly or indirectly in creating, using or being affected by the new technology. Therefore, developers, designers, researchers, users and other people can be regarded as stakeholders. Borning and Muller (2012) emphasized that giving voice to the participants in VSD studies and clearly expressing the voice of researchers in publications is important to aid others in understanding the setting in which VSD was carried out and how the ethical analysis has taken place. Similar to the motivation of Participatory Design (Schuler 1993), sharing responsibility and power among stakeholders and designers/researchers is beneficial to investigate value questions. To this end, contextualized VSD methods could, e.g., include ethical analysis by stakeholders in-situ, thereby letting them act as "lay ethicists".

4.8.1.3 Concrete Methods

VSD does not prescribe concrete methods for empirical investigations, but states: "the entire range of quantitative and qualitative methods used in social science research is potentially applicable here, including observations, interviews, surveys, experimental manipulations, collection of relevant documents, and measurements of user behavior and human physiology." (see this chapter)

While Borning and Muller (2012) suggest that researchers examined the value *suitabilities* of these methods, Le Dantec and colleagues (2009) propose that more specific methods are needed to capture values as lived experiences and to give stakeholders the power to express and share their comprehension of local values.

In my opinion, the question of concrete methods for VSD is not only closely related to the methods' abilities to facilitate participation, but also to the competences within a design team. An important question is "does a design team need to include a social scientist or someone trained in the methods above to be able to carry out VSD?" In reports of past VSD projects the researchers' background and expertise is not always transparent and VSD does not offer a concrete proposition on the composition of design teams. Given that VSD has until now mainly been carried out by HCI researchers, a combination of knowledge of technology engineering and social science research is often present. Nonetheless, conceptual analyses of values with ethical import may also benefit from professional philosophers or ethicists. Considering the use of VSD in industry practice, expertise in ethics cannot always be easily acquired. Design teams may benefit from value advocates, but reports from the field show that in a business-oriented setting value advocates may meet challenges (Manders-Huits and Zimmer 2009). Their role has to be considered carefully, e.g., with respect to how much leadership they take and how other design team members receive such leadership.

Another way to empower technology developers who are untrained in social science or ethics are specific tools or techniques to deliberately consider and account for values in design. Since the first publication of this chapter, VSD researchers have developed several methods for value discovery and definition (e.g. Le Dantec et al. 2009; Woelfer et al. 2011) and for the consideration of the broader and long-term socio-technical context (e.g. Nathan et al. 2007; Friedman et al. 2012), which can be utilized by researchers, practitioners and other stakeholders. Some of the methods will be elaborated on below.

4.8.1.4 Summary of Practical Questions

Summarizing the discussion above, I compiled a list of practical questions to be considered when using VSD in practice.

- Which values are important in a given design case? Whose values are they and how are they defined with respect to the given context?
- Which methods are suited for conceptual/empirical/technical investigations in VSD? Which methods, in particular, are suited to discover, elicit and define values?
- What kind of social science knowledge or skill set is needed to engage in VSD?
- How can methods give power and responsibility to stakeholders and make them "lay" ethicists?

The remainder of the addendum will provide an ongoing design case that exemplifies the use of methods recently developed in VSD and partially addresses these practical questions.

4.8.2 VSD Case: Safety for Homeless Young People

4.8.2.1 Socio-technical Context

The socio-technical context for the design case consists of homeless young people, mobile technologies, and safety. People generally want to be safe, know that their families are safe and help others to be safe. For homeless young people, life can be very difficult when securing basic needs, such as safety, food, and shelter, while sometimes even managing physical and mental health problems. They encounter unsafe situations in their struggle to meet their needs, often with civility laws being implicated (Woelfer and Hendry 2011).

Across social classes mobile phones are becoming essential for safety, as they are carried closely to people's bodies at all times and can be accessed in emergency situations to connect to others. At the same time "overreliance on its safety functions may undermine a person's resilience" (Woelfer et al. 2011). For homeless young people mobile phones have beneficial safety functions ranging from functionality, e.g. calling or texting in unsafe situations, to form factors, e.g. held in particular ways the phone may resemble a gun. However, mobile phones may also create unsafe situations, e.g. if homeless young people trespass at secluded power outlets in order to recharge their phones. Thus, the use of mobile technology by homeless young people and its relation to safety is multi-faceted and a topic worth investigating.

The central question in the ongoing research is "How can mobile technology be designed to keep homeless young people safe?"

This design case is representative for recent developments of theory and method within VSD. In particular, it engages in design work concerning multiple stakeholder groups with different perspectives and values as well as value tensions within and among individuals and groups. Further it reflects how empirical research of current conditions and co-design activities to envision the future can be integrated.

4.8.2.2 Stakeholder Analysis

Unlike other design and engineering methodologies VSD is a holistic approach to the design and introduction of new technologies considering not only primary (and maybe secondary) users, but also other stakeholders. VSD makes a deliberate distinction between direct stakeholders and indirect stakeholders, and by that dictates an explicit consideration of people affected by the system, who are not users. This allows for an early analysis of benefits and harms of new technology for the whole social environment in which it is situated.

In the given design case the homeless young people were considered as the primary direct stakeholders, but the researchers did not exclude the emergence of another relevant direct stakeholder group throughout the research and later design phases. Three indirect stakeholder groups identified: service providers, police officers, and community members.

4.8.2.3 Value Analysis & Value Tensions

As explained briefly above the use of mobile technology bears benefits and obstacles for homeless young people with regard to their own needs and values and in interactions with other stakeholders (e.g. other urban dwellers). Therefore, value tensions on three levels were anticipated in the project: (1) within the individual, (2) between an individual and another stakeholder group and (3) between stakeholder groups. In order to get a detailed understanding of how these tensions play out in real life situations, several methods (following in the next subsections) were used. Important to note is that VSD does not require a definition of the concept of a value at the onset of the project. In the given design case, the researchers explicitly did not define their conception of safety, but instead used methods, that were open-ended, yet gave enough structure (through precise tasks) to guide the participants' reflection. By using verbal and visual methods a rich set of data has been elicited that reveals the nuances of stakeholders' perceptions of safety and its situational nature.

4.8.2.4 Value Sketches

Sketching is often used in design work to uncover knowledge for "physical and conceptual structure" (Woelfer et al. 2011). Value sketches in particular are meant to emphasize participants' values. In this project value sketches were especially useful to uncover *situated* perceptions of safety, as these are often time and location based. For example, one could feel safe in a specific location during the day but not during the night. Therefore, participants were given two identical maps of their living area, one for daytime and one for night-time activities, and asked to use red and green (graphic and textual) marks to represent their perception of safety of different regions. Participants used different marks to denote safe and unsafe areas, spots and paths.

Through detailed coding and analysis of the sketches, Woelfer and colleagues could retrieve a detailed picture of temporal and location sensitive perceptions of place, mobility and safety for each stakeholder group.

4.8.2.5 Stakeholder Generated Value Scenarios

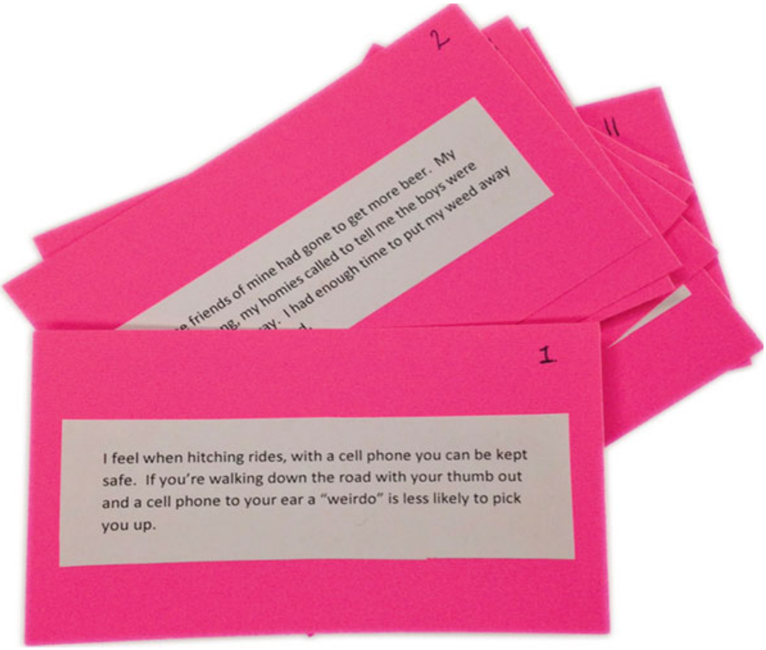
Scenarios have a long-standing tradition in scenario-based design (Rosson and Carrol 2003). These scenarios often tell short-term and functionality-focused stories about how the designers intend the system to be used by direct stakeholders. While being a powerful tool to analyze aspects of functionality and usability, these scenarios lack in portraying long-term systemic effects that new technology has on the social and political environment.

Value scenarios (Nathan et al. 2007) are a VSD tool that combines the narrative power of traditional scenarios in design processes with five new key elements that help to engage in (ethical) issues of long-term and emergent use of new technology: indirect stakeholders (additionally to direct ones), pervasiveness (effects from the widespread adoption of the technology), time (long-term effects), systemic effects and value implications. By describing possible positive and negative effects and value tensions that come along with widespread adoption, value scenarios support technologists and policy makers to consider the creation and introduction of new technologies.

While value scenarios were originally intended for early strategic planning of technology projects or as touchstones for policy-making discussions, the design case at hand provides a new way to use value scenarios. In this design case the stakeholders wrote value scenarios. They were prompted to write a true or fictional story of how a mobile phone could keep a homeless young person safe. In this way value scenarios became tools in the design process to elicit stakeholders' views and experiences. One benefit of writing fictional stories is that participants could mask their identity while still providing perspectives and ideas, which would be too risky to portrait openly. One example of such a fictional story is the following:

> Once upon a time there was three little pigs, one lived in a house, one lived on the street, and the last one lived in a squat. One day a big bad wolf was looking for a squatter, the big bad wolf was out to get all the little pigs. The first little pig called the second pig, and he found the third pig through word of mouth. Thank cellphone.

In a recent iteration of VSD investigations in this project the stakeholder-generated scenarios (Fig. 4.4) were utilized in co-design activities with homeless young people, police and service providers (Yoo et al. 2013).



friends of mine had gone to get more beer. My
ng, my homies called to tell me the boys were
ay. I had enough time to put my weed away

I feel when hitching rides, with a cell phone you can be kept safe. If you're walking down the road with your thumb out and a cell phone to your ear a "weirdo" is less likely to pick you up.

Fig. 4.4 Value Scenarios written by stakeholders and used in design iteration

4.8.2.6 Envisioning Cards

Creating awareness of and considering values and other systemic effects of new technology can be difficult – for designers, technologists and also for other stakeholders engaged in co-design activities. To put technology development in a broader socio-technical, long-term perspective, by highlighting "diversity, complexity and subtlety of human affairs, as well as the interconnections among people and technologies" (Friedman and Hendry 2012), the Envisioning Card toolkit provides a promising means.

Envisioning Cards (see Fig. 4.5) incorporate similar elements to the Value Scenarios: stakeholders, time, values and pervasiveness. The aim of the card toolkit is to raise awareness of long-term and systemic issues in technology design. To this end each card has an evocative image and title on one side and the envisioning criterion, theme description and concrete design activity on the backside.

In the context of the project the cards were used as an iteration step in a co-design activity with homeless young people, police and service providers. After creating 3D prototypes to keep homeless youth safe, participants were asked to select an Envisioning Card, consider the theme and refine their designs if needed (Yoo et al. 2013). The Envisioning Cards stimulated the creative exploration of the design space. They helped participants "to reframe technical problems, to reconsider technical aspects of their designs, and generally to catalyze their technical imaginations." (Friedman and Hendry 2012).

Technologies are not always used in ways that the designers intended. Who might use the system for unplanned or nefarious purposes (e.g., frustrated stakeholder or an identity thief)? In what ways?

Identify three roles that involve non-intended use of the system.

Fig. 4.5 Envisioning Card front (*left side*) and back (*right side*) (source: VSD lab, University of Washington (UW), permission to reprint the image and copyright remains with UW. See also: http://www.envisioningcards.comwww.envisioningcards.com)

Overall, the Envisioning Cards are a versatile tool that can be used in many design processes including ideation, co-design, heuristic evaluation, and critique or as educational tools. The cards are self-explanatory and open for different types of use, which makes them equally accessible to designers, technologists and end-users and supports them in ethical reflection. This is especially of interest for design situations in which ethical or social science knowledge is lacking within the design team.

4.8.2.7 Reflection on the Use of VSD

Reflecting on the project as carried out until now, I would like to highlight a few aspects that can be attributed to the use of VSD. As the project is ongoing, there is no final product to inspect and no definite answer to the initial research question: "how can mobile technology be designed to keep homeless young people safe?" Still the VSD process has already uncovered several interesting results, which common engineering or user-centered design (UCD) approaches would have missed. The latter approaches would have focused solely on the homeless young people as they are defined as the primary users of the technology. Focusing only on this stakeholder group, however, would have prevented the researchers to gain insights from the police and service providers. Such insights and specifically the prototype designs of these groups provided many new opportunities to design not simply a new device that promotes safety, but also the social network and necessary socio-technical solutions (e.g. power supply or back-up phone at service providers' station).

The VSD methods further provided rich insights into how the different stakeholders perceive safety instead of using merely a designer's definition of the concept. It became clear that for the homeless people safety is linked to basic needs and also other values such as being part of a community and affordability. In addition, the complex interplay of safety, time and place as well as mobile phone use (leading to increased or reduced safety) was the result of value sketches and scenarios. Although we cannot be certain, I believe, that such complexity is hard to be obtained by standard interviews. Furthermore, as mentioned earlier, face-to-face interviews do not allow participants to mask their identities and may therefore be less apt to obtain sensitive data for this stakeholder group. Last, it was noticed that the envisioning cards supported stakeholders in re-considering and adapting their ideas (in forms of prototypes) in the light of long-term consequences, pervasive use and other stakeholder groups. A standard usability evaluation would have missed such aspects.

4.8.3 Discussion

In the following, we review the design case to address the questions posed in Sect. 4.8.1.4.

The presented design case set out to investigate the role of mobile technology and safety for homeless young people. While safety was a central value in the project, the researchers deliberately avoided a definition of safety from the onset of the project and allowed other values to emerge. The various methods presented above helped to define the nuanced perceptions of safety of the different stakeholder groups. Especially the value sketches and value scenarios highlighted that safety is fundamentally situational for the homeless young people. Whether a place, another stakeholder group, or the mobile phone use is considered safe or unsafe depends on the situational (e.g. temporal) context. Ongoing co-design work with homeless young people and service providers revealed more dimensions of safety and its relation to other basic needs.

The methods presented in this addendum can be used or combined for different types of VSD investigations. While we focused above on gathering empirical data around the perceptions of safety and mobile technology, more recent work used parts of this data (i.e. the value scenarios) in technical investigation of concrete designs of envisioned mobile technology.

The HCI researchers, who carried out this work, have extensive background in VSD and co-design and experience in working with the homeless community and service providers. HCI and VSD researchers are trained in social science methods and have affinities towards ethics and technology design. Therefore, to consider the taxonomy given in the introduction of this book, VSD can be considered as having a joint project organization, i.e. the ethical assessment is done from within the project. Ideally, VSD projects would include professionals trained in ethics, social sciences, computer science/engineering and design.

However, researchers and developers without such expertise at hand should not refrain from using VSD. Especially in industry practice, where it is equally important to design in a value sensitive manner, it cannot be assumed that a design team is sufficiently trained in ethics or social sciences. One way to address a shortcoming may be to have consultants or value advocates from outside the projects providing these skills. Another way would be to develop more specific toolkits to trigger value sensitive deliberation and discussion within the design team, and tools to work out value definitions and tensions with stakeholders. The Envisioning Cards provide one example of such a toolkit.

Further, Borning and Muller (2012) have introduced the term "lay" ethicists, i.e. stakeholders that act as ethicists in the given design context. While "lay" ethicists are not able to substitute professional ethicists or social scientists, providing stakeholders with specific methods that allow them to consider and discuss ethical design questions in-situ is a step towards accounting for values in design. The methods presented above trigger thinking about ones' (Dunne and Raby 2001) values and I believe that especially value scenarios or related methods, e.g. design noir, can be used to make stakeholders more aware of the (unforeseen, long-term) ethical issues at stake. More work needs to be done in VSD research that allows for a common ground between designers, researchers, practitioners and other stakeholders in understanding each others' perspectives and situated value definitions.

Important for methods that allow stakeholders to voice themselves safely is that they provide means for controlling precision and ambiguity for the data they elicit. In the project presented here safety is a sensitive topic for homeless young people as it may be linked to embarrassing or dangerous experiences. Discussing such experiences may be uncomfortable for the participants or even put them at risk, as one homeless person mentioned, "letting people know where I feel safe, makes me feel unsafe" (Woelfer et al. 2011). Such effects were mitigated by the ambiguity of fictional scenarios and variable precision of the value sketches.

4.8.4 Conclusions and Future Work

In this paper I have pointed out the recent developments within VSD and clarified how VSD fits into the taxonomy presented in the introduction of this book. As an evolving framework, its assumptions and practicability have been under ongoing critical review. This has led to adaptations (e.g. value lists as heuristics) and new methods for VSD investigation. By elaborating on these methods in the context of a recent VSD case and addressing practical questions of VSD, I extended the previous publication of this chapter.

Considering current paradigm shifts in innovation towards co-creation in several public sectors (e.g. healthcare) to achieve solutions, which integrate technology in community-based and social practice, I believe, VSD will play an important role as a framework to envision and design long-term socio-technical change. To facilitate this role HCI researchers will continue to evolve VSD and put more focus on multi-lifespan systems, participation of stakeholders and shared value investigation. An essential part of supporting widespread VSD practice is the early education of researchers and practitioners in various fields, which are being addressed in academic courses and workshops at major research venues (e.g., Detweiler et al. 2012).

References

Azenkot, S., Prasain, S., Borning, A., Fortuna, E., Ladner, R.E., & Wobbrock, J. O. (2011). Enhancing independence and safety for blind and deaf-blind public transit riders. In-*Proceedings of CHI 2011*. New York: ACM Press.

Borning, A., & Muller, M. (2012). Next steps for value sensitive design. In *Proceedings of the 2012 annual conference on human factors in computing systems* (pp. 1125–1134). New York: ACM Press.

Dunne, A., & Raby, F. (2001) *Design Noir: The secret life of electronic objects*. Basel: Birkhäuser.

Friedman, B., & Hendry, D. G. (2012). The envisioning cards: A toolkit for catalyzing humanistic and technical imaginations. In *Proceedings of the 2012 annual conference on human factors in computing systems* (pp. 1145–1148). New York: ACM Press.

Friedman, B., & Kahn, P. H., Jr. (2003). Human values, ethics, and design. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook* (pp. 1177–1201). Mahwah: Erlbaum.

Denning, T., Borning, A., Friedman, B., Gill, B., Kohno, T., & Maisel, W. (2010). Patients, pacemakers, and implantable defibrillators: Human values and security for wireless implantable medical devices. In *Proceedings of CHI 2010 conference on human factors in computing systems* (pp. 917–926). New York: ACM Press.

Detweiler, C., Pommeranz, A., & Stark, L. (2012). Methods to account for values in human-centered computing. In *CHI '12 extended abstracts on human factors in computing systems (CHI EA '12)* (pp. 2735–2738). New York: ACM Press.

Le Dantec, C. A., Poole, E. S. & Wyche, S. P. (2009). Values as lived experience: Evolving value sensitive design in support of value discovery. In *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)* (pp. 1141–1150). New York: ACM Press.

Manders-Huits, N., & Zimmer, M. (2009). *Values and pragmatic action: The challenges of introducing ethical intelligence in technical design communities. International Review of Information Ethics, 10*, 37–44.

Nathan, L. P., Klasnja, P. V., & Friedman, B. (2007). Value scenarios: A technique for envisioning systemic effects of new technologies. In *CHI '07 extended abstracts on human factors in computing systems* (pp. 2585–2590). New York: ACM Press.

Schuler, D., & Namioka, A. (Eds.). (1993). *Participatory design: Principles and practices*. Hillsdale: Lawrence Erlbaum Associates.

Woelfer, J. P., & Hendry, D. G. (2011). Homeless young people and technology: Ordinary interactions, extraordinary circumstances. *Interactions, 18*(6), 70–73.

Woelfer, J. P., Iverson, A., Hendry, D. G., Friedman, B., & Gill B. T. (2011). Improving the safety of homeless young people with mobile phones: Values, form and function. In *Proceedings of the 2011 annual conference on human factors in computing systems* (pp.1707–1716), New York: ACM Press.

Yoo, D., Huldtgren, A., Woelfer, J. P., Hendry, D. G., & Friedman, B. (2013). A value sensitive action-reflection model: Evolving a co-design space with stakeholder and designer prompts. In *Proceedings of the SIGCHI conference on human factors in computing systems* (CHI '13, pp. 419–428). New York: ACM.

Chapter 5
Socio-technical Integration Research: Collaborative Inquiry at the Midstream of Research and Development

Erik Fisher and Daan Schuurbiers

Abstract Midstream modulation is a framework for relating changes in research and innovation to changes in practitioners' contextual awareness. It is used by the socio-technical integration research (STIR) program to help elucidate and enhance the capacities of laboratory practitioners to participate more deliberately in the governance of science, technology and innovation. STIR involves collaborative inquiry between embedded humanists or social scientists and the scientists, engineers and others who host them. The collaborative inquiry takes place during routine research and innovation activities, generating feedback that can modulate these activities in real-time. Reflexive midstream modulations can disrupt and enhance the conditions under which research and innovation practitioners engage the social and ethical contexts of their work. This chapter presents the conceptual backbone and the overall philosophy behind midstream modulation, and surveys the concrete outcomes that typically result from laboratory engagement studies.

This material is based upon work supported by the National Science Foundation under Grant No. #0849101 and #0937591.

E. Fisher (✉)
CSPO, Arizona State University, PO Box 875603, Tempe, AZ 85287–5603, USA
e-mail: efisher1@asu.edu

D. Schuurbiers
De Proeffabriek (The Pilot Plant), Lookwatering 36, 2614 KA Delft, The Netherlands
e-mail: daan@proeffabriek.nl

N. Doorn et al. (eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*, Philosophy of Engineering and Technology 16, DOI 10.1007/978-94-007-7844-3_5,

5.1 Introduction

Midstream modulation is a framework for relating incremental changes in research and innovation processes, whether momentary or gradual, that take place as a result of practitioners' changing cognitive interactions with their social and ethical contexts. In combination with intervention-oriented interdisciplinary collaboration, such as the approach of socio-technical integration research (STIR), it can help elucidate and even enhance the capacities of laboratory researchers and other practitioners to participate more deliberately in the governance of science, technology and innovation. Developed by Erik Fisher during a 3-year laboratory engagement study, midstream modulation has been used to structure STIR studies in a range of academic laboratories, industrial organizations, and other knowledge and service settings around the world.[1] Midstream engagements involve ongoing, intensive, collaborative inquiry between embedded humanists or social scientists and the scientists, engineers and others who host them. The collaborative inquiry takes place during routine research and innovation activities, generating feedback that can modulate these activities in real-time. As demonstrated both in STIR quasi-experimental studies and ethnographic accounts, midstream modulations tend to follow discernible patterns and can take several distinct forms. They can disrupt and also enhance the conditions under which research and innovation practitioners engage in critical reflections on the social and ethical contexts of their work. Specifically, practitioners who participate in the STIR experience have been shown to become more reflexively aware of socio-ethical contexts, to alter their decision processes, and to change the nature and direction of their work in light of the collaborative inquiry. This chapter presents the conceptual backbone and the overall philosophy behind midstream modulation, and surveys the concrete outcomes that typically result from laboratory engagement studies.

5.2 The Midstream Modulation Framework

Midstream Modulation is a three-part dialectical framework for relating the gradual alteration of science and innovation processes and practices in response to practitioners' ongoing cognitive interactions with their broader social contexts.

---

[1] The Socio-Technical Integration Research (STIR) project is a coordinated set of 20 laboratory engagement studies to assess and compare the varying pressures on – and capacities for – laboratories to integrate broader societal considerations into their work. STIR is co-funded through the NSF programs in Science, Technology & Society; Biology and Society; Mathematical and Physical Sciences and Society; Science of Science and Innovation Policy; and Office of International Science and Engineering. The project is administered through Arizona State University's Center for Nanotechnology in Society. Since initial NSF funding, STIR has been taken up more generally as a program of inquiry and interaction within an increasingly diverse application in a variety of research and innovation governance settings.

Cognitive interactions refer to reflection and learning that is stimulated or enhanced by means of observation of oneself and one's context. The framework is scalable insofar as it applies to both routine laboratory decisions and to larger science and innovation policy processes as they unfold over time. It is also meant to explain socio-technical change regardless of whether intervention-oriented activities—such as STIR—are undertaken or not. In keeping with this volume's theme of "opening up the laboratory," we focus on the role of midstream modulation engagements in enhancing the deliberative and socially responsive capacities of scientists and engineers as they participate in the broader governance of research and innovation.

The metaphorical content of "the midstream" is rich and, more importantly, points to practical opportunities for practitioners to make concrete changes "from within." In its most literal sense, the "midstream" denotes the phase of research and development in between 'upstream' research authorization and 'downstream' research outcomes. For instance, it can denote the implementation stage of a broader public policy process that is distributed over place, time and a variety of social actors (e.g., Fisher et al. 2006). The "midstream" can also refer to the intersection of human social, and material influences that make up an established yet continually evolving course of action. It refers to the ongoing "flow" and malleable nature of research and innovation activities, but also to the more or less durable institutional and material structures that contain and channel those activities.

Midstream modulation represents the incremental effects of human social, and material perturbations on this flow—both in terms of its meaning and its direction. Practical adjustments take place as a matter of course as scientists and engineers follow normal routines and modify their behavior to take into account new and constantly changing circumstances. This much is evident in trial and error learning and other micro-processes by which innovation trajectories are able to evolve at macro levels. For a variety of reasons, however, these patterns of learning and behavior do not normally involve scientists and engineers systematically reflecting on the social and ethical context of their research as they conduct it. *What happens if they do?*

Without completely overrunning the organizational and cultural constraints that structure "normal" midstream modulation (i.e. the delicate interactions of cognitive and social perturbations that routinely occur as research decisions in the laboratory unfold), laboratory engagement studies attempt to elucidate the conditions for midstream modulation to become more reflexive and more deliberate. These studies thus add another layer of perturbations, or 'stir', the ongoing modulation processes within the research laboratory (rather than trying to forcefully shape those processes). Hence, midstream modulation is not just about conceptually relating research and development to its policy, use and regulation; it's about being in the flow of work-related and other decisions, and being able to trace and stir ongoing modulations while the research is taking place.

As stated, practitioners are already reflexive in that they constantly take social and other contextual considerations into account. But they tend do so tacitly and implicitly and only when prompted, for instance, by a crisis. Midstream modulation studies inquire into the effects of systematically integrating into research regular reflection on the societal dimensions of that research. They are premised on the assumption that, as scientists and engineers become more consciously aware of the social and ethical contexts within which they already work, they are more likely to critically reflect upon these contexts. Critical reflection can in turn enable innovation actors to “take more into account” (Mitcham 1994), to the point where they may choose to alter their routine decisions so as to respond to a broader array of concerns. Ideally, this type of critical reflection and deliberate modulation will enhance the social value of science.

This process is formalized in a dialectical framework that begins with *de facto* or “normal” midstream modulation. This initial stage is assumed to be what takes place in research laboratories and other midstream sites of research and innovation as a matter of course. When a crisis emerges, or is perceived to emerge—whether at the scale of a Kuhnian revolution, or simply at the level of human agency—what was normal can become “post-normal” (Funtowicz and Ravetz 1993; Kuhn 1962). Midstream modulation highlights the cognitive function of scientific individuals or institutions in the process of science becoming post-normal. When practitioners experience a heightened sense of the normal, *de facto* modulation has become reflexive. In a state of reflexive modulation, practitioners do not necessarily make any deliberate changes in what they do—they simply become more aware of one or more dimensions of the broader socio-ethical landscape within which they are operating. Depending upon the meaning that they and other stakeholders attribute to these broader dimensions, reflexive awareness may or may not precipitate a “crisis.” While heightened awareness may be the goal of some interventionist programs, reflexive modulation is theorized here as a necessary but not sufficient condition for deliberate changes in practice that reflect this heightened awareness. The third stage of midstream modulation is thus called deliberate modulation, and it entails productive changes in practice, materially or otherwise. Midstream modulation is what happens when reflection on existing practices changes those practices. In the language of John Dewey, when a midstream actor experiences reflexive modulation, they find themselves in a “problematic situation” (which could happen when a humanist enters a laboratory), and they may then choose at that point to undertake “reflective inquiry” (Dewey 1929). Midstream modulation may pass from reflexive to deliberate, but this is by no means guaranteed and depends on multiple conditions, including the practitioner’s own perceptions of the appropriate course of action in any particular situation.

To date, the STIR project has enabled laboratory engagement studies in nearly 30 scientific and engineering laboratories throughout North America, Europe and East Asia. Embedded humanists and social scientists employ an ensemble of concepts and techniques in order to document and assess the conditions and effects of self-critical science and engineering. These studies have empirically shown that there are two main forms that deliberate modulation can take: first order reflection on the means of pursuing given ends, and second-order reflection on the ends as such (Schuurbiers 2011). In either case, deliberate midstream modulation can lead to changes in material and reflective practices in surprising ways, as we will show below. First, we briefly discuss the connection between the midstream modulation framework and socio-technical integration research methodology.

5.3 STIRring the Midstream

Historically, tensions between the performance of science and engineering work and various normative expectations for how science and engineering should be performed have been addressed by numerous governmental efforts and attempts at reform (e.g., Guston 2008).[2] These tensions have become more pronounced in recent public policy mandates for "upstream public engagement" and "responsible innovation" (Fisher 2011). Governmental efforts tend to consist of a combination of formal policies and attempts at intervention, usually imposed from without, that seek to alter the largely self-governing routines of scientists and engineers (Fisher et al. 2006). We have argued elsewhere (Fisher 2007; Schuurbiers 2010) that such governmental efforts should ideally be complemented by supporting activities that can be described as 'governance from within' (Fisher et al. 2006).

Midstream modulation was developed to "probe the capacity" of technical experts to respond to calls for more socially responsive scientific practices. Laboratory engagement studies can be used to test the logic behind these mandates, to engage the responsive capacities of research and innovation, and to conceptualize the role of midstream activities within broader processes of innovation governance. As such, they complement recent policy initiatives for responsible innovation (US Congress 2003; Netherlands Organisation for Scientific Research 2008; Von Schomberg 2013). In the STIR program, midstream modulation can be explicated, engaged and even enhanced through situated or embedded collaborative inquiry. While STIR can be used instrumentally in order to enhance the social and public value of midstream modulation, it is also a form of inquiry in and of itself, which can shed light on the contemporary cultural and institutional conditions that enable and constrain scientific and engineering practices from taking broader concerns and considerations more explicitly and effectively into account.

Fisher developed the midstream modulation framework and techniques for elucidating it for practitioners in real-time during a 33 month laboratory engagement study in a nanoscale engineering lab (Fisher 2007; Fisher and Mahajan 2006a). The primary instrument by which midstream modulation capacities are elucidated, or "stirred," takes the form of a decision protocol (Fisher 2007). This protocol structures interviews, conversations and interactions between the embedded scholar and his or her laboratory hosts. The protocol consists of four conceptual components—opportunity, considerations, alternative, outcomes—and attendant questions that are meant to unpack the socio-ethical dimensions of laboratory research (Fisher 2007). Opportunities may take the form of a research problem, an occasion to take advantage of, or any situation eliciting a response. Considerations are enabling or constraining selection criteria that potentially influence or determine the response to the opportunity. Alternatives are perceived options or courses of action available for

---

[2] These attempts are paralleled by an 'empirical turn', a growing interest for more 'relevant and focused' approaches, in disciplines like engineering ethics, science studies and in the philosophy of science and technology: (Kroes and Meijers 2000; Nordmann and Rip 2009; Van de Poel and Verbeek 2006; Webster 2007).

selection in response to the opportunity. Outcomes are understood as a particular response to the opportunity, through selecting one or more alternatives, in light of one or more considerations. Outcomes occasion new opportunities. Ongoing and regular use of the protocol and its questions during the actual conduct of research and innovation activities enables the circulation of both social and material knowledge among the research participants. The protocol thus serves as the primary feedback mechanism in the STIR program by which humanistic and social scientific inquiry is integrated with natural scientific and engineering inquiry.

5.4 Changing Practices

STIR has been applied in academic and industrial laboratories, both at the bench science level and the level of management and administration. Participants in the program generally conduct paired, comparative laboratory engagement studies in two laboratories that carry out similar work but are located in different nations. STIR studies combine ethnographic and ethnomethodological forms of observation with engaged modes of feedback and collaborative inquiry. It has also been used in various educational and residential management settings. The results demonstrate that laboratory engagement studies not only lead to changes in thinking and learning, but can also generate meaningful changes in individual and collective behavior.

5.4.1 Fisher's Pilot Study

Fisher developed and tested the decision protocol as a means of elucidating midstream modulation during the final 12 weeks of his longer study in early 2006 (Fisher and Mahajan 2006a, 2010; Fisher 2007). Whereas Fisher had previously conducted interviews, archival research and participant-observation in the lab, this 12-week study sought to probe the capacities of research decisions to integrate broader societal considerations as mandated by the US *21st Century Nanotechnology Research and Development Act of 2003* (see Fisher 2005; Fisher and Mahajan 2006b). What became the STIR pilot study combined ethnographic and quasi-experimental methods: while Fisher as an "embedded humanist" engaged in unstructured interactions with both individual practitioners and groups within the lab, the focal point of the study involved two research participants, one of whom had semi-structured weekly and bi-weekly ("high") interactions with Fisher, and one of whom had semi-structured monthly ("low") interactions. A third research participant, who had no interactions with Fisher during the study, acted as a control.

Each of these three participants took part in pre- and post- program interviews to identify and assess changes in researchers' "reflexive awareness" of various specified contextual social, institutional, cognitive and axiological factors, including awareness of the researcher's own decision making. Fisher found that, whereas the control demonstrated little if any changes in reflexive awareness at the end of the 12-week study, both the “high” and “low” interaction participants did, and that the quality and intensity of the changes corresponded to their levels of participation in the study.

Although the goal of the study was not to actually alter the research or material practices of the lab, Fisher was told by the “high interaction” participant that, had they not interacted as they did over the course of the 12 weeks, one particular engineering project “could have been a whole different thing.” Fisher and Mahajan (2006a, b) elaborate on this statement:

> The collaborations between [Fisher] and [the “high interaction” participant] seem to have influenced the direction of [the participant’s] research, even though such goal-oriented modulation was beyond the objectives of [Fisher’s] study. Altered decision outcomes included:
>
> - Developing a new disposal method for carbon nanotubes (there had existed no unique method for their disposal)
> - Making a change in catalyst used for carbon nanotube synthesis (ferrofluid versus Ferrocene)
> - Making a change in the experimental setup that was deemed safer than the previously established method (inserting a three way valve)
> - Developing safety strategies for working with carbon nanotubes (there had existed none before)[3]
>
> These altered outcomes were presumably stimulated by the activities of the study. They are attributed to the reflexive modulation that was documented, since otherwise they would not (in the views of [the participant], [Fisher], and the [laboratory] director) have occurred.

Fisher further recounts the second “decision outcome” from the above list (Fisher 2007):

> While the protocol rendered elements of [the “high interaction” participant’s] decision making more visible to the author, it also made them more visible to [the participant]. His reflexive capacity to identify and align ‘social’ considerations with ‘technical’ alternatives in order to solve a complex problem…was conditioned by a reformulation of the social in terms of the technical.…The design and inhabitation of structured spaces for socio-technical integration can thus, by increasing reflexive awareness, affect decision making.

Fisher’s pilot study had sought primarily to elucidate “normal” midstream modulation, and to investigate the possibility of enhancing “reflexive” midstream modulation. That laboratory decisions and material practices were deliberately altered as a result of enhancing reflexivity—this was the view of various observers of and participants in the study—despite the fact that Fisher had made no conscious attempt to alter them, suggests the transformative potential of observation, description and reflexive awareness.

---

[3] Fisher’s study was designed to investigate the possibility and utility of socio-technical integration in light of science and technology policy mandates. It employed what we now refer to as the STIR methodology, as a form of collaborative midstream modulation. Hence, the material outcomes were unexpected but salient indications not only of the possibility and utility of integration, but also of its transformative capacity.

5.4.2 Schuurbiers's STIR Studies

In 2008–2009, Schuurbiers adopted the midstream modulation framework in two laboratory engagement studies to address the question of social responsibility in research practice, focusing on researchers' critical reflections on the broader socio-ethical context of their work (Schuurbiers 2010). These studies sought to gauge whether the midstream modulation framework could increase the visibility of and reflection on the broader context of laboratory research (Schuurbiers 2011). Eight laboratory researchers participated in the studies, four of whom had regular interactions (up to 12 h per week) with Schuurbiers as an 'embedded ethicist' during a period of 12 weeks. In keeping with the pilot study, interactions with research participants were documented by means of pre- and post interviews, participant observation and regular application of the protocol. In addition, visual representations of the research process drafted in collaboration indicated the links between the interrelated series of decision processes mapped over the 12-week period. As documented elsewhere (Schuurbiers 2011), the protocol and visual representations served to render normative issues that were directly related to the research at hand more visible to both the laboratory practitioner and the embedded ethicist, bringing the socio-ethical context to life within the context of research.

The elucidation and enhancement of midstream modulation ('stepping into the helicopter', as one participant put it), enabled researchers to creatively address the socio-ethical context of their work through collaboration and in real time. Initial reticence from research participants transformed over the course of each study's duration into enthusiasm for discussing both the progress and the broader aspects of their research. This enthusiasm may derive in part from the attention being given to research participants: for 12 weeks their work was in the centre of attention, and every move was noticed, recorded and discussed. For PhD students, bathing in attention in that way can be especially motivating, considering that daily routines in the laboratory do not involve such extensive daily discussions. The psychological effect of being in the centre of attention in turn opened up research participants to a more conscious and direct experience of how their work was situated in larger research and policy frameworks, and that experience in turn exemplified the relevance of broader considerations to the work in the lab. Schuurbiers's studies indicate that broader socio-ethical dimensions can be productively engaged during laboratory research. The midstream modulation framework was found to help engender fruitful and meaningful collaborations between the social and natural scientists, encouraging second-order reflective learning while respecting the lived morality of research practitioners (Schuurbiers 2011).

Corresponding to the interventions, Schuurbiers documented specific changes in research practice. For instance, laboratory participants reported hitting upon new research ideas and directions as a result of their interactions with Schuurbiers and/or because of their own critical reflections that these interactions sparked. An interesting instance of deliberate midstream modulation occurred at the level of the lab group itself: after Schuurbiers concluded his second study with a presentation of his initial findings to the group, various discussions ensued among the group regarding the nature of laboratory safety practices, including the question of when to wear laboratory safety coats. Not long afterwards, Schuurbiers ran into one of the members of the lab and was informed of the discussions, which had continued to take place over the course of the next few days and led to a collective decision among lab members to wear their lab coats. This vignette illustrates several relevant insights into the nature and possibility of more socially responsible and responsive science. Note that it is not the questions of laboratory safety or compliance with rules and regulations that commanded our attention. Rather, the vignette demonstrates that *deliberation among the laboratory members* took place, and that it took place *at the group level.* Furthermore, it took place as a result of earlier collaborative inquiry between individual researchers and the embedded ethicist. Therefore, this example of *second-order deliberate modulation* was also a collective form of deliberation that took place on its own, and that led to collective action, in this case changes in material practices. This has several important implications.

First, the vignette underscores the importance of reflection and deliberation as a first step towards changing the material practices of a research organization (we will come back to this point below, in our discussion of the distinction between reflection and action). Second, even if this is but one example of second-order deliberation, it illustrates that the kind of deliberative capacities that democratic theorists believe are important for the broader health of a democracy (Elster 1998; Leibj 2006) can be built through efforts to elicit and enhance midstream modulation. Semi-structured feedback of observational findings typically invites broader reflection on those findings. Finally, it is a proof of concept for the scalability of the STIR experience. The "soft" interventions of the STIR program can lead to critical reflection, deliberation and subsequent alteration of practices at the group level. This suggests that this type of engagement may encourage reflection and deliberation at the scale of an organization, if not a community, even though this may require modifications to the protocol. Attempts to elucidate and enhance midstream modulation, such as in the case of STIR, of necessity depend upon the voluntarism of the scientific and engineering practitioners not only to reflect critically on their routines and behavior, but also to alter them in light of these reflections. As the studies of Fisher and Schuurbiers show, dependence upon the voluntary behavior of practitioners is by no means an impediment to learning or changes in practice; instead, this is an example of co-responsibility (Mitcham 2003) and it underlines the robustness of the approach among expert practitioners.

5.4.3 Subsequent STIR Studies

Since 2009, numerous more STIR studies have sought to engage with midstream modulation. Chapter 8 of this volume discusses some of the outcomes of the STIR laboratory engagement studies of researchers Shannon Conley and Antonio Calleja-López, analyzing the material dimensions of their interactions from the standpoint of their acquisition of interactional expertise. As Conley has elsewhere noted when reflecting on her STIR studies, "collaboration…cannot be a one-way street" (Conley 2011). She goes on to relate an instance in which her embedded interactions pulled her into the midstream modulation of the lab:

> Just as my natural science collaborators express interest (and sometimes confusion) in my work, I have expressed an interest in their work, and have consequently gained interactional competence in the process. At one point, when, in a casual venue outside of the lab, I signaled my intention to audit a molecular genetics course in order to deepen my understanding of some of the laboratory's experiments, I was confronted by a member of the lab who had learned of my expressed interest. "Why would you want to do that?" he demanded. After I explained my motives, another lab member accompanying us asked if I had "done a PCR" (polymerase chain reaction), a technique that is commonly used to amplify regions of DNA. When I responded that I had not, my collaborators immediately suggested that I try my hand at one. Within a matter of days, donning a lab coat and rubber gloves, and armed with a pipette, I was doing my own PCR, no longer "benchside," but at the bench itself (Conley 2011).

As Conley reported at a 2011 public workshop,[4] her material integration into the laboratory culture in this way also continued to have ripple effects. For instance, her experimental results were used as a basis to teach others to perform PCR, both in the laboratory in question as well as in the second of her paired laboratories.

Paul Ellwood also presented the results of his STIR studies at the same workshop. One aspect in particular of his findings was picked up by workshop participants from Genome British Columbia, who subsequently published the following account of Ellwood's research:

> STIR doctoral students have had the opportunity to shape lab-based practices related to safety, the environment, patient outreach and other issues that can influence research decisions through their interdisciplinary projects. For example, one project examined the use of nanotechnology in dishwashers. Paul Ellwood noted that during his lab-based collaboration with scientists "a key decision point [in the research process] concerned the switch from studying the 'wet-cycle' of the wash to the drying phase, when it was found that nanoparticles then accumulated on surfaces (e.g. plates)." Social scientists and humanists can raise important questions during the collaborative process and this may lead to technical modifications.

Other STIR studies have produced material changes in research practices as well as alterations at the level of research agendas and laboratory policies. Steven Flipse found several instances of transformed research ideas and strategic thinking within two industrial laboratories (Flipse et al. 2013).

---

[4] The workshop was held at the Woodrow Wilson International Center for Scholars in Washington, D.C. and has been archived for online viewing. See the STIR project website at: http://cns.asu.edu/stir/workshops.php?ws=4. Also at the workshop were presentations by Ellwood, Schuurbiers and other STIR researchers, who presented their work alongside the laboratory directors and researchers that hosted them. This allowed for further confirmation of the material effects that collaborative inquiry into midstream modulation has had.

5.5 Material Deliberation

The results presented here from a handful of STIR laboratory engagement studies show that midstream modulation can be observed and documented, and that feeding these descriptions back into the research environment in real-time can be the basis for materially productive forms of collaborative inquiry. They further demonstrate that socially-sensitive, materially situated inquiry can have tangible effects on the laboratory itself, its research, as well as on the thinking and behavior of both individuals and the research groups as a whole. These findings hold promising implications not only because they enhance critical reflection and practices of individuals and small groups, but because they point to the possibility and utility of engaging routine research and innovation decision-making. Taking "normal" research and innovation activities as the starting point for critical, collaborative inquiry and reflection turns out to be a potentially powerful way to build deliberative and responsive capacities that can transfer over time and place. In other words, engaging material practices is relevant for enhancing the responsible governance of research and innovation. As Susanna Hornig Priest has suggested:

> That science is largely 'sold' on a promise that it will address fundamental human needs and desires entails a profound obligation to contemplate, at least on occasion, just what these needs and desires might be (Hornig Priest 2005, 298).

The question remains who is to do the contemplation, where it is supposed to occur and to what effect. We have embarked with the STIR program in an empirical investigation of the practical and material dimensions of such contemplation when it is physically, temporally and culturally situated within the worlds of expert practitioners. Precisely what STIR 'does', or how collaborative inquiry in the laboratory affects those worlds, are theoretical and empirical questions that should remain open to investigation. While we are documenting findings on various levels, we would like to warn here against overly simplified interpretations of what is at stake.

Implicit in normative injunctions for *plus respicere*, enhanced reflexivity, more critical reflection and the like is that shifts in cognitive ability will result in corresponding shifts in material practices and the governance of science in society. Indeed, critical discussions and reflections that do not yield clear, concrete and transformative outcomes run the risk of being ineffectual if not self-serving. This is why the results of public engagement and interdisciplinary collaborations are considered "generally disappointing" by some (Rip 2009). On the other hand, privileging the material outcomes of a laboratory engagement runs the risk of overstressing the instrumental role of science studies (Jasanoff 2011). The results of the STIR studies suggest that clear-cut distinctions between action and reflection miss the point. What matters in the end is the interplay between the changes in material practice that we have focused on here, and the crucial role that changes in thought and reflective practice played in enabling them in the first place.

In a 2009 piece we published while Schuurbiers's laboratory engagement studies were still underway, we reported that one of his research subjects initially made the following statement about the STIR experience: "Does it change my thinking? Yes.

Does it change what I do on a daily basis? No." At the time, we were not convinced that "changes in thinking" could be considered relevant, if one were evaluating laboratory engagement studies from the standpoint of socially responsible science, sustainability, or anticipatory governance. Subsequent laboratory engagement studies in the STIR program however pointed out the intricate and unpredictable ways in which changes in "reflexivity" lead to changes in "materiality", without being able to exactly specify each causal component in the system: consider Fisher's instance of carbon nanotube synthesis, Schuurbiers's account of the laboratory coats, Conley's acquisition and contributions of material expertise, and Ellwood's example of nanoparticles in dishwashing detergent.

Ironically, the same researcher who made the statement distinguishing his thoughts from his actions came back on that statement after the paper went to press, indicating that his thinking inevitably changes what he does. The point we are making is that studies in midstream modulation and engagement ought not to be evaluated by one side or the other of a dichotomous pair of metrics. While instances of reflection may not correspond one-on-one to instances of material redirection, the former is a vital condition for the latter. Reflexivity and materiality should be treated symmetrically where possible, keeping in mind that they are both resources for governance. Moreover, while both are necessary evaluative metrics to bear in mind, neither is sufficient for more anticipatory, socially robust or even democratic forms of governing research and innovation. These must be underwritten with the normative and interpretive obligations carried by any social scientific or humanistic endeavor (Jasanoff 2011).

As STIR studies show, reflection and action feed into one another in surprising and often unpredictable ways. Similarly, the clear-cut distinction between individual practices and institutional structures may conceal how the two are linked. Again, we suggest that each of the multiple levels of agency need to be attended to as necessary but not sufficient. Individual and even group level instances of reflexive and deliberate modulation of laboratory behavior will not by themselves bring about larger adjustments in the 'rules of the game'. Conversely, changes in institutional design, university regulations, policy mandates and the like are not in themselves worthwhile if they do not actually advance important public norms and values. Furthermore, the discussion is moot if expert practitioners in research and innovation themselves are not acknowledged as, at least in some respects, self-determining and autonomous participants.

Laboratory engagement studies can produce, or be used as, demonstrations for those who take part in them and their broader communities, allowing specific examples to serve as proofs of concept for the possibility and utility of self-critical science and engineering. For while researchers are likely to initially only make changes in their routines and decision making on the basis of "research value" and self interest, we have found that these first-order modulations can be followed by second-order modulations, which in turn can produce outcomes that are more clearly linked to "public value" and common interest (e.g., Fisher et al. 2010).

Whether STIR and midstream modulation studies can help build longer-term capacities that in turn allow research and innovation to be more responsive to societal concerns and public values—either through democratic or expert forms of deliberation—remains to be seen. It will presumably require further expansion into policy and public engagement domains (Wynne 2011). But the concrete changes in both research practice and agendas that are the result of STIR activities suggest that broader questions like this can and ought to continue to be studied empirically.

References

Conley, S. N. (2011). Engagement agents in the making: On the front lines of socio-technical integration. Commentary on: "Constructing productive engagement: Pre-engagement tools for emerging technologies". *Science and Engineering Ethics, 17*(4), 715–721.

Dewey, J. (1929). *The quest for certainty: A study of the relation of knowledge and action*. New York: Minton, Balch and Company.

Elster, J. (Ed.). (1998). *Deliberative democracy*. Cambridge: Cambridge University Press.

Fisher, E. (2005). Lessons learned from the Ethical, Legal and Social Implications program (ELSI): Planning societal implications research for the National Nanotechnology Program. *Technology in Society, 27*, 321–328.

Fisher, E. (2007). Ethnographic invention: Probing the capacity of laboratory decisions. *Nanoethics, 1*(2), 155–165.

Fisher, E. (2011). Editorial overview: Public science and technology scholars: Engaging whom? *Science and Engineering Ethics, 17*(4), 607–620.

Fisher, E., & Mahajan, R. L. (2006a, November 5–10). Midstream modulation in an academic research laboratory. In *Proceedings of the American Society for Mechanical Engineers International Mechanical Engineering Congress and Exposition*. Chicago.

Fisher, E., & Mahajan, R. L. (2006b). Contradictory intent? US federal legislation on integrating societal concerns into nanotechnology research and development. *Science and Public Policy, 33*(1), 5–16.

Fisher, E., & Mahajan, R. L. (2010). Embedding the humanities in engineering: Art, dialogue, and a laboratory. In M. E. Gorman (Ed.), *Trading zones and interactional expertise: Creating new kinds of collaboration*. Cambridge, MA: MIT Press.

Fisher, E., Mahajan, R. L., & Mitcham, C. (2006). Midstream modulation of technology: Governance from within. *Bulletin of Science, Technology & Society, 26*(6), 485–496.

Fisher, E., Biggs, S., Lindsay, S., & Zhao, J. (2010). Research thrives on integration of natural and social sciences. Correspondence. *Nature, 463*, 1018.

Flipse, S. M., Van der Sanden, M. C. A., & Osseweijer, P. (2013). Midstream modulation in biotechnology industry: Redefining what is 'part of the job' of researchers in industry. *Science and Engineering Ethics, 19*, 1141–1164.

Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the post-normal age. *Futures, 25*(7), 735–755.

Guston, D. H. (2008). Innovation policy: Not just a jumbo shrimp. *Nature, 454*(7207), 940.

Hornig Priest, S. (2005). Commentary – Room at the bottom of Pandora's box: Peril and promise in communicating nanotechnology. *Science Communication, 27*, 292–299.

Jasanoff, S. (2011). Constitutional moments in governing science and technology. *Science and Engineering Ethics, 17*(4), 621–638.

Kroes, P., & Meijers, A. (Eds.). (2000). *The empirical turn in the philosophy of technology*: Vol. 20 of Research in philosophy and technology, ed. Carl Mitcham. Oxford: Elsevier.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Leibj, E. J. (2006). Can direct democracy be made deliberative? *Buffalo Law Review, 54*.

Mitcham, C. (1994). Engineering design research and social responsibility. In K. Shrader-Frechette (Ed.), *Research ethics* (pp. 153–168). Totowa: Rowman & Littlefield.

Mitcham, C. (2003). Co-responsibility for research integrity. *Science and Engineering Ethics, 9*(2), 273–290.

Netherlands Organisation for Scientific Research (2008). *Responsible innovation — Description of thematic programme*. http://www.nwo.nl/files.nsf/pages/NWOA_7E2FMH_Eng/$file/MVI_description_April2008.pdf. Accessed 16 Oct 2012.

Nordmann, A., & Rip, A. (2009). Mind the gap revisited. *Nature Nanotechnology, 4*, 273–274.

Rip, A. (2009). Futures of ELSA. *EMBO Reports, 10*, 666–670.

Schuurbiers, D. (2010). *Social responsibility in research practice: Engaging applied scientists with the socio-ethical context of their work*. Delft: Simon Stevin Series in the Ethics of Technology.

Schuurbiers, D. (2011). What happens in the lab: Applying midstream modulation to enhance critical reflection in the laboratory. *Science and Engineering Ethics, 17*(4), 769–788.

US Congress (2003). 21st century nanotechnology research and development act. P.L., 108–93, 1–24.

Van de Poel, I. R., & Verbeek, P.-P. (2006). Editorial: Ethics and engineering design. *Science, Technology and Human Values, 31*, 223–236.

Von Schomberg, R. (2013). A vision of responsible innovation. In R. Owen, M. Heintz, & J. Bessant (Eds.), *Responsible innovation*. London: Wiley.

Webster, A. (2007). Crossing boundaries – Social science in the policy room. *Science, Technology and Human Values, 32*, 458–478.

Wynne, B. (2011). Lab work goes social, and vice versa: Strategising public engagement processes. Commentary on: "What happens in the lab does not stay in the lab: Applying midstream modulation to enhance critical reflection in the laboratory". *Science and Engineering Ethics, 17*(4), 791–800.

Chapter 6
Ethical Parallel Research: A Network Approach for Moral Evaluation (NAME)

Ibo van de Poel and Neelke Doorn

Abstract Research and Development (R&D) of new technologies increasingly takes place in networks of different organizations and actors. In this contribution, we present an approach that was developed at Delft University of Technology for addressing ethical issues in R&D. The approach takes the engineers and scientists involved in R&D as entry point for discerning and discussing ethical issues and is to be carried out parallel to the R&D trajectory. On the basis of two cases studies, the network approach is described in detail including its strengths and weaknesses. Two procedural norms for assessing an R&D network are discussed, viz. inclusiveness and second-order learning. Some of the main advantages of the approach are that it offers the possibility to identify moral issues in situations of uncertainty and indeterminacy about the final consequences of technological innovations, while being applied already at the early stages of technological development. Because the moral issues are identified in their real-world context, the approach can generate insights that immediately influence R&D and design decisions. As such, the approach may help focusing the technical work in a way that moral issues are better addressed.

I.R. van de Poel (✉) • N. Doorn
Department of Technology, Policy and Management, Delft University of Technology,
P.O. Box 5015, 2600 GA Delft, The Netherlands
e-mail: i.r.vandepoel@tudelft.nl; n.doorn@tudelft.nl

N. Doorn et al. (eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*, 111
Philosophy of Engineering and Technology 16, DOI 10.1007/978-94-007-7844-3_6,

6.1 Introduction

Our current society is inconceivable without technology. Technologies play an important role in our daily life and they enable us to do certain things we otherwise could not do. Medical technologies, for example, enable early diagnostics of some diseases, which was impossible before. Cell phones enable us to communicate with each other when we are not at home. Usually, the introduction of new technologies is aimed at advancing human well-being.[1] However, the introduction of new technologies may come at the price of negative and unforeseen side-effects and risks. These consequences are often hard to predict with accuracy beforehand and in many cases only materialize during use (Swierstra and Jelsma 2006). Nevertheless, it is desirable that potential moral concerns (e.g., those regarding safety, sustainability, privacy, distribution of welfare, and social inclusion), are identified in as early a stage as possible and are included as additional design criteria in Research and Development (R&D) and technological design, especially since seemingly "neutral" choices during the design process may significantly affect the broader social and political issues pertaining to the final design of the technology (Schot and Rip 1997; Schinzinger 1998; Van de Poel and Van Gorp 2006).

In this contribution, we propose an approach for addressing moral issues in R&D networks. The focus of the approach is on how the actors within an R&D network can actively take up their responsibility for addressing such issues. We are particularly interested in the moral issues that arise due to the use of the eventual innovation or artifact being developed and the need to anticipate these issues. Our focus is therefore less on issues that the research itself may give rise to, like scientific integrity issues or moral issues that have to do with experiments with humans and animals ("research ethics," narrowly conceived), although several of such issues might also be addressed by the approach we have developed.

The approach set out in this paper falls under the more general heading of ethical parallel research. The idea behind this kind of ethical research is that ethical investigations are carried out parallel to, and in close cooperation with, a specific technological R&D project. In ethical parallel research, ethicists, or other humanities and social science researchers, investigate moral issues in parallel to an R&D trajectory in close cooperation with the involved scientist and engineers, and they feed back their results to the actors in the R&D network.

---

[1] Although this claim may in general be true, it should be noted that many technologies are being developed in a host of other contexts, such as the military sector, which are not always uncontroversial. The notion of dual-use technology has been introduced to refer to research and technology with the potential both to yield valuable scientific knowledge and to be used for purposes with potentially serious detrimental consequences. Although dual-use is as old as engineering and design, the terrorist attacks of 9/11 and recent developments in the life sciences have renewed the attention for the topic (Van der Bruggen 2012). The moral assessment of dual-use technologies and the prevention of its harmful use is currently one of the most debated topics engineering ethics (cf. the recent special issue on "The Advancement of Science and the Dilemma of Dual Use" in the journal *Science and Engineering Ethics*; (Spier 2010)).

In this contribution, a specific approach to ethical parallel research is presented, the so-called network approach. Our starting point is the idea that R&D takes place in a network of actors, which we analyze and assess in terms of its composition and (learning) capacity. Hence, rather than focusing on particular individual researchers or starting with specific moral issues, we take the characteristics of the network itself as the primary unit of analysis. The main reason for this is that ethical issues often do not arise at the level of individual researchers and engineers, but at more collective levels, for example because there is a lack of clarity regarding the distribution of responsibilities.

The approach consists of three main parts. The first is aimed at identifying moral issues in R&D networks; the second at ethical reflection and judgment on these issues, and the third on the distribution of responsibilities for addressing these issues. These three aims are reflected in the respective subsections of Sect. 6.2. Taken together, these three aims yield two procedural network norms, which are discussed in Sect. 6.3. In Sect. 6.4, two examples of applications of the approach are discussed. In the concluding Sect. 6.5, the strengths, challenges, and limits of the approach are discussed in more detail.

6.2 Network Approach

In this section, we outline our network approach, which consists of three phases.

6.2.1 Identifying Ethical Issues: Network Analysis

The first element of the approach is the identification of relevant ethical issues in an R&D setting. R&D covers a broad range of activities, and involves many different actors and institutions, working together in complex cooperative ventures (network organizations). These network organizations often lack a strict hierarchy and a clear task division (Rogers and Bozeman 2001; Saari and Miettinen 2001). Building on insights from Science & Technology Studies (STS), we propose to study R&D from the perspective of a network of collaborating actors with different stakes and goals, but also different problem definitions (Callon 1992; Elzen et al. 1996; Klijn 1997a; Mitchell et al. 1997; Mehalik and Gorman 2006).

The network approach assumes that the rationality of professionals is bounded rather than comprehensive. This means that professional behavior can best be understood as a heuristic search process, which is guided by cognitive frames shifting back and forth between problem setting and problem solving (Schön 1983; Grin and Van der Graaf 1996). These cognitive frames, which are referred to as “frames of meaning” guide the behavior of the actors (Spender 1989). In this bounded rationality view, actors are driven by strategic considerations, such as how to generate support or how to frame the problem, and not by a single pre-defined set of goals (Johanson and Mattsson 1992; Klijn 1997b; Marsh and Smith 2000).

Central to the network approach are the different actors' "problem definitions" (Dery 1984; Fischer 1995) and their "agendas" (Kingdon 1984). The latter can be defined as a "coherent set of goals or ends, which these actors want to achieve" (Zwart et al. 2006, p. 671). Problem definitions are linked to the actor's interpretation of what the central issue in the network is. Different agendas translate into differences in opinion about which issues should be addressed in the network.

The main elements in this phase of the approach are the mapping of the network and its characteristics, which can be done by observation, document analysis, interviews, and brain-storming sessions with the involved actors and stakeholders on, for example, relevant moral issues and risks of the technology being developed. The first step in the analysis is the identification of the relevant actors and stakeholders. This step requires a demarcation criterion: which agents can be considered to be an "actor" in the network, and which agents a "stakeholder." Zwart et al. (2006) propose to call an agent an actor if the following three requirements are fulfilled:

1. the agent is a human[2] who can act deliberately and purposefully;
2. the agent can influence the decisions made and the actions undertaken in the network;
3. the agent has regular interactions with the other actors in the network.

In addition, one could distinguish stakeholders as agents who are not participating in the network but who may undergo the consequences of decisions in the network without necessarily being able to influence them. Following this criterion, potential users who do not participate in the development of the new technology can also be considered stakeholders.

Once the actors and stakeholders are identified, their mutual relationships can be mapped. Zwart et al. (2006) distinguish three kinds of relations:

1. information and knowledge sharing;
2. a relationship of financial or labor input. These relationships establish asymmetric bonds of dependency;
3. a relationship of potential deals.

After having distinguished the main actors in a network, the different actors' roles and agendas are to be described in more detail. Most technological projects will include the following roles, but other roles may be possible as well (see Fig. 6.1 for a graphical representation of the R&D network discussed in more detail in Sect. 6.4.1):

1. researchers;
2. producers of a technology, which includes activities like design and consultancy;

---

[2] Note that the authors do not use the broader definition of actor as proposed in Actor-Network-Theory.

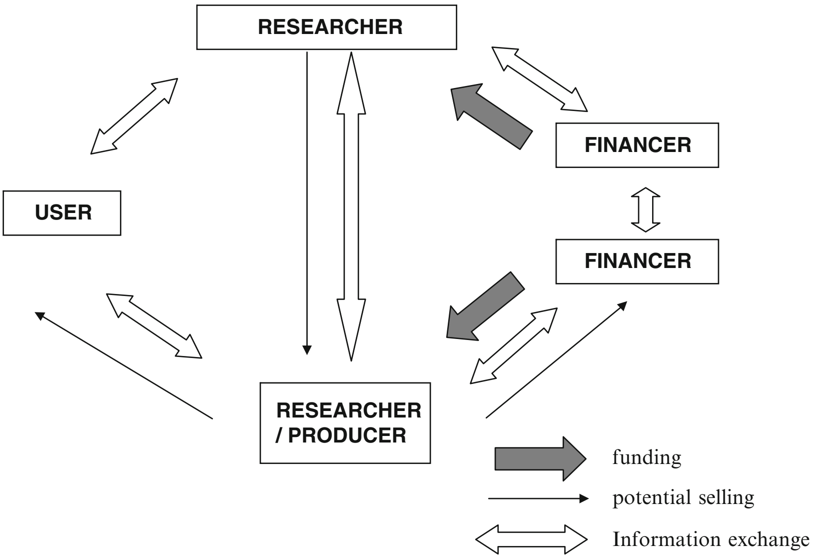


Fig. 6.1 Sample figure of R&D network

3. users of a technology;
4. financers of technical research.

Actors can have more than one role. In the project described in Sect. 6.4.1, for example, one actor was both user and financer. Some roles bring with them particular responsibilities. The producers of the technology, for example, may be responsible for delivering some tangible results at the end of the project. Financers may have the responsibility to spend taxpayers' money in a justifiable way (dependent on whether or not they are related to the government). In addition to these formal responsibilities, the different actors in the network may, tacitly or openly, allocate responsibilities informally to each other in the network. Such informally allocated responsibilities may be relevant, for example, if unforeseen and undesirable consequences of the new technology, which are not accounted for in formal contracts or the law, materialize. These informal responsibilities may evoke strategic behavior and negotiation.

In addition to and partly stemming from their different roles and responsibilities, the actors in the network have different aims for participating in a network and also different interests. These aims and interests will have consequences for the choice of the issues that the actors want to address within the context of the network. Above, we called these the agendas of the actors. The agendas of the different actors may to some extent overlap, but they may as well be in conflict (see Table 6.1 for an example of a hypothetical network).

Together, mapping the actors with their mutual relationships and the identification of (conflicting and/or supporting) agendas provide a detailed description of the research network that can serve as input for the identification of ethical issues. For example, an unclear distribution of responsibilities or conflicting agendas may point

Table 6.1 Example matrix of actors and their agendas

	Fundamental research	Effective technology	Quickly launching technology	Patents	…
Actor 1: Researcher	O		X	O\X	
Actor 2: Producer	X		O		
Actor 3: Financer		O		X\O	
Actor 4: Financer					
Actor 5: User		O			
…					

O means item is on the agenda, *X* means potential friction

at an ethical issue. It should be noted that the actual identification of ethical issues is an additional step to the network analysis itself. This step may be carried out by the actors in the R&D network itself on the basis of a common sense notion, or a simple definition, of what makes an issue an "ethical issue." One might for example call an issue "ethical" if certain moral values are at stake. Moral values that might be at stake in an R&D setting include, but are not limited to, safety, sustainability, human-well-being, fairness, honesty, privacy, trust, et cetera. In many cases, identifying ethical issues requires judgment in the sense that there is not a simple algorithm to decide whether an issue is ethical or not. Because of the importance of judgment it might be desirable to have an ethicist or at least someone with some training in ethics involved in the identification of ethical issues. Conversely, it is clear that the identification of ethical issues requires detailed knowledge of the context of the R&D project to be able to recognize the relevant issues. It is for this reason that in our approach the identification of ethical issues starts with a detailed network analysis to "contextualize" the analysis, which may be necessary to alter innovation trajectories (Wynne 2011).

6.2.2 *Ethical Judgment: Wide Reflective Equilibrium*

The second element in our approach is not just to describe the network and the research decisions taken, but also to evaluate them. Although network approaches have been applied to similar cases both in policy science and in science and technology studies (Hakansson 1989; Schneider 1992; Bressers et al. 1994; Elzen et al. 1996; Smit et al. 1998), our approach differs in that we shift the focus from a primarily descriptive towards a normative point of view.

The challenge here is to develop an approach to ethical judgment that does justice to the plurality of ethical views in R&D networks and that is productive. To this end we have adapted the wide reflective equilibrium (WRE) approach, initially developed by Rawls (1999 [1971]) and further elaborated by Daniels (Daniels 1979, 1996), as a framework for analyzing moral deliberation (see also Doorn 2010). In the concept of WRE, a distinction is made between three layers of moral considerations: (1) descriptive and normative (moral) background theories, (2) moral principles, and (3) considered moral judgments about particular cases or situations. Reflective equilibrium refers to a state of coherence between one's considered judgments and moral principles concerning a certain case. According to Rawls, everyday moral reflection takes place by examining our moral judgments on particular matters (layer 3) with more general or broader beliefs and principles on similar issues (layers 1 and 2). In order to decide how to respond to moral issues, people move back and forth between these various beliefs and considerations, reflect on them, revise them if necessary and try to achieve an acceptable coherence between their moral judgments on particular matters with more general or broader beliefs and principles on similar issues. In this deliberative process, all three layers are open to revision. In the literature, the term WRE can refer both to the state of coherence between these three layers of morality and to the process or method itself of arriving at such equilibrium.

Rawls used the concept of WRE in explicating and defending his theory of justice in the context of political philosophy. He tried to develop a criterion of justice that would be fair to all despite the diversity of moral frameworks people endorse (hence, the idea of justice as fairness). Although Rawls at first tried to develop a substantive conception of justice – viz. the right to equal basic liberties, the right to fair equality of opportunity and the difference principle (which says that inequalities are allowed only if they work to the benefit of the worst-off group) – he later recognized the plurality of incompatible and irreconcilable moral frameworks within a democratic society and he limited the idea of justice as fairness to a purely procedural conception of justice. People with divergent moral doctrines will most probably not agree on a substantive conception of justice but they can overlap in their acceptance of a procedural conception of justice. For these procedural principles to be justified they must cohere with each individual's background theories and considered judgments. Rawls introduced the term overlapping consensus to refer to these justice principles that are agreed upon by all citizens and that are part of each individual's WRE. If there is coherence between all layers of the WRE model, the outcome of the reflective process can be considered justified, Rawls argued.

In real applications, Rawls' idea of WRE can be used in a twofold way, viz. constructively (that is, to encourage discussion) and with a justificatory purpose. Concerning the constructive use of the WRE approach, the WRE approach is built on the idea that moral deliberation takes place by reflecting on different layers of morality and seeking coherence between these different layers. Moral deliberation can then be encouraged by explicitly paying attention to these different layers of morality, rather than focusing a discussion on just one of the three layers. Hence, if a discussion is structured along the lines of the WRE approach (that is, if the three different layers are explicitly made topic of discussion), people are encouraged to include the other layers in their deliberation as well. It is then more likely that people will revise their original judgments that do not fit well with other moral judgments, principles, and background theories and possibly reach a consensus. If the outcome of this discussion fits in the individual actors' WRE, the outcome can be considered justified.

In this phase of the network approach, the relevant moral issues are analyzed in terms of the WRE approach. As stated above, this is an evaluative exercise. This phase is typically done in an interactive setting (for example, a workshop or brainstorm session). In the first phase (network analysis), the moral issues were identified. We explained that, by analyzing the network in terms of agendas, roles and responsibilities, it was possible to identify some potential obstacles, for example because different actors had conflicting agendas or some issues where not on anyone's agenda. In this way, actors may be confronted with issues which they may not have recognized as important until then.

The aim of this phase is to reach a consensus on the issues that are to be addressed in the project. Building on the list of issues identified in the previous phase, a decision need to be made on what is and what is not part of the project, preferably on the basis of a consensus. Important is that this consensus is justified. In the Rawlsian framework this yields the criterion of "overlapping consensus," which means that the agreed upon outcome fits each individual's WRE and that it is more than just a negotiated outcome. That means that for all people, their individual WRE needs to be "reconstructed." In other words, it needs to be assessed what the normative background theories of the different people are, as well as their moral principles and considered judgments, and how the final consensus (if any) fits in these layers. The considered judgments probably form the starting point for the discussion, because this is usually the layer where disagreements become visible. Some people may say that certain issues are certainly not part of the project, whereas others will argue that it is the team's responsibility to address these moral issues as well. By including the moral principles and the normative background theories in the discussion, it can then be explored to what extent a consensus is achievable. People are asked to explain why they think that certain issues should or should not be done by the team. They may, for example, refer to more abstract principles, such as fairness of the workload, efficacy and efficiency, end users, or to normative background theories such as Kantian deontology ("we should not use the end user simply as a means for our technological project") or consequentialism ("we should not include this issue in the project if only few people will benefit and the costs outweigh the benefits"). Analysis of the discussion between the different team members provides information on the question to what extent the outcome coheres with the different layers of their individual WRE.

It is important to make the analysis inclusive, in the sense that all relevant stakeholders have their stakes included, also relevant stakeholders who do not have a formal role in the project but who may potentially be affected by the technology.

Again, correct demarcation is important here. Van de Poel and Zwart (2010) propose the criterion of public reason to demarcate relevant stakes from irrelevant ones. If a consideration can be argued for on the basis of public reason and not only private reason, it can be considered a "reasonable stake" and therefore relevant to address. These reasonable stakes translate into a list of moral issues that are to be addressed in the project.

6.2.3 Distribution of Responsibilities

Building on the previous two phases, the central issue in the third phase is the distribution of responsibilities. The aim for this third phase is to distribute the responsibilities for addressing the issues discerned in the previous phases, in a way that is acceptable to all actors involved and that is "productive" in addressing the issues adequately. Here we are faced with a similar challenge as in the case of moral judgment. Research has shown that, in a pluralist society, people have different views on responsibility, not only in terms of what responsibility *means* but also in terms of the question when a person *is* responsible for addressing something (Doorn 2012). The approach we have developed therefore aims at moving beyond potentially conflicting substantive notions of responsibility.[3] Again the idea of WRE is useful for deliberating and trying to agree upon a distribution of responsibilities that is acceptable to all actors involved, as the approach may encourage people to think in terms of a "fair" workload and the legitimacy of other people's arguments.

Given that people have different conceptions of responsibility, the WRE approach is particularly suitable for deliberating on the fairness of concrete responsibility ascriptions. As explained above, the WRE model is based on a three-tiered view on moral reflection. The most abstract layer consists of people's background theories. These can be normative background theories (such as, deontological ethics, consequentialism, virtue ethics, etc.) but this layer also includes other background theories, for example based on one's professional background. On a less abstract level (the mid-level), we can distinguish people's moral principles. A person's particular conception of responsibility can also be located at this level. In a discussion, such a responsibility conception will appear as the type of argumentation that people use when arguing for their particular position. People who defend a consequentialist conception of responsibility, will probably use arguments related to efficacy and efficiency. People with a merit-based conception of responsibility, will probably refer to arguments related to fairness. The most concrete layer consists of people's considered moral judgments. This level also includes people's particular responsibility ascriptions; that is, the concrete answer to the question "who is responsible for addressing this particular issue?" Since responsibility is related to all three layers, the WRE approach is very suitable for reflecting on responsibility. By reflecting on all three layers, people can try to reach an overlapping consensus on how responsibilities are to be distributed. In order to make it a justified consensus, the consensus

---

[3] The underlying thought is that people do not have to agree on substantive conditions which tell when a person is responsible as long as they agree on the *procedure* for distributing the responsibilities (and given that they have a shared understanding of what responsibility means. The latter is important to prevent people from talking at cross-purposes). If such a procedure, or its outcome, is accepted by all people involved as representing the "fair terms of cooperation," this might help reconciling the different responsibility conceptions and, ultimately, make sure that the important issues are indeed addressed.

should fit each individual's WRE. Similar to the second phase, this phase of the network approach is typically done in an interactive setting, like a workshop or brainstorm session.

This phase starts with the list of moral issues selected in the previous phase. In a number of iterative steps, members of the project team can be asked to distribute the tasks to address the respective issues among the different actors. At the start, there will probably be significant disagreements. This is not surprising, since these "moral tasks" are typically not covered in the working plans of the project. However, by discussing the disagreements, not only at the most concrete level but also in terms of the more abstract moral principles and background theories, people are encouraged to rethink their original responsibility ascriptions. After some time of discussion, the distributing exercise can be repeated. It can then be investigated to what extent people converge to a consensus on how the responsibilities are to be distributed. The same procedure of discussion and re-distribution can then be repeated, if necessary. At the end, people have to evaluate to what extent they think the points on which they agree fit their own WRE (that is, whether it is a justified overlapping consensus). Since it is will probably be difficult for people without any training in political philosophy to make this assessment, it is useful to provide some background information on Rawls' model and political deliberation during of the meeting. This may contribute to the reflective capacity of the participants as well.

6.3 Two Procedural Norms

Above we described our approach for discerning and dealing with moral issues in R&D networks. To recap, we identified three different phases, with three distinct aims:

1. to identify moral issues;
2. to judge how these moral issues are to be addressed;
3. to distribute the responsibilities in a fair way.

We now want to look at two procedural norms that are important for the ability of an R&D network to adequately deal with moral issues: reflective learning and inclusiveness. The perspective here is somewhat different from the section above. The approach we described above can be applied by ethical parallel researchers to help people in the network identify and judge moral issues and to ensure a fair and productive distribution of responsibilities. As such the approach helps in better dealing with moral issues in R&D networks. The norms we will discuss in this section can be used to *evaluate* how well an R&D network is able to adequately deal with moral issues. These norms can be used by outsiders to an R&D network, like the government, potential funders, or ethical parallel researchers, to judge the network in this respect. They may also be used by participants in the networks to judge their own networks and to try to improve it. The approach we set out in Sect. 6.2 might be one way to try to improve the network, although it is probably not the only way.

These norms, which are also used in the literature on policy and innovation networks, contribute to achieving a justified overlapping consensus. Before explaining the relation between these norms and the procedural approach, we first discuss the two norms in somewhat more detail.

6.3.1 Inclusiveness and Openness

As stated in Sect. 6.2, the network approach requires a demarcation criterion. The first norm we distinguish is related to this demarcation criterion; it is the norm of "inclusiveness" or "openness." It can be described as the norm that all relevant actors are included in a network. In the literature on social implications of technology, this norm is often addressed by proposals for involving stakeholders during design, development, and implementation of new technologies to broaden the scope of these processes. Many researchers argue that making the network inclusive will bring social and moral considerations to the table and make them part of the design and implementation process. In the last decades, many participatory methods have been developed to include the different perspectives of stakeholders in the development and decision-making about technology: Integrative Assessment (Van Asselt and Rijkens-Klomp 2002), participatory technology development (Schot 2001) or design (Schuler and Namioka 1993; Kensing 2003), stakeholder learning dialogues (Daboub and Calton 2002), Constructive Technology Assessment (Rip et al. 1995; Schot and Rip 1997), Interactive Technology Assessment (Grin and Hoppe 1995; Reuzel et al. 2001), Participatory Technology Assessment (Schot 2001; Joss and Bellucci 2002), scientists stakeholder workshops (Cohen 1997; Hanson et al. 2006), and consensus conferences (Joss and Durant 1995; Einsiedel et al. 2001). Whereas some people take the position that all relevant *considerations* should be included, others emphasize that inclusiveness is a democratic criterion per se. Adherents of the latter position argue that, based on the very nature of democracy, all relevant *stakeholders* should be able to participate in the network. Participation, in this view, is understood as a way to empower citizens and stakeholders (see, e.g., Sclove 1995). This distinction is also reflected in the different contributions to this volume. The approach described by Rip and Robinson, for example, seems to be an example where inclusiveness serves an "instrument" to broaden technological development. Van der Burg's approach of moral imagination, described in Chap. 10, seems more an example in which inclusiveness is seen as a democratic criterion per se. We take the former position here. We think that the interests of all relevant stakeholders (that is, the relevant considerations) should be included, but that this does not require that all relevant stakeholders become part of the network. In practice, this will probably lead to an unworkable situation, possibly hampering the learning capacities of the network as well (see also Sect. 6.3.2). Including additional people in the network may be a means to making the network inclusive in the sense of addressing all relevant considerations. In our view, the minimal requirement of inclusiveness is that all relevant stakeholders are represented in the network and that their the interests are on the agenda of the network.

The norm of inclusiveness does not only contribute to identifying relevant moral issues, but it also contributes to the ability to judge such issues adequately. As explained above, with respect to dealing with ethical issues we propagate the reflective equilibrium approach. The reflective equilibrium approach is aimed at getting an overlapping consensus and this consensus can only be considered just, if all relevant considerations are included. This means that all relevant stakeholders need to be represented in the network. Otherwise we run the risk that a limited group of people achieve a consensus on the decision how to address, for example, health risks, without taking into account all relevant views on the issue. To decide what considerations and what actors are relevant for inclusiveness one might apply the Rawlsian criterion of public reason (Van de Poel and Zwart 2010). Each actor that can legitimately claim to have a "reasonable stake" or a "reasonable interest" is to be included in the network. Reasonableness is here understood in terms of whether something can be argued upon on the basis of public reason.

Since "relevance" will probably always be debatable, the criterion of openness is added to warrant the possibility that new aspects become relevant. However, openness has an additional, more institutional feature. The criterion of openness calls for an open discourse, which means that it is not only important that all relevant actors are included, but that they have equal opportunities for participating in, and contributing to, the decision-making process as well. If a group of actors from different fields and with different levels of expertise are engaged in a conversation, it is important that the vocabulary used by the experts is understandable to all. The criterion of openness also requires that people feel free to bring in unwelcome arguments. If some actors are discouraged to do so and remain silent, the overlapping consensus that is arrived at cannot be considered justified. Together, inclusiveness and openness contribute to an overlapping consensus that is justified. They prevent unjustified shortcuts to a wide reflective equilibrium or overlapping consensus. The latter could, for example, be the case when people with unwelcome arguments are excluded from the network.

6.3.2 Reflective Learning

The second norm is reflective learning. Most scholarly literature on learning goes back to the work of Fischer (1980, 1995) and Schön (1983). Fischer conceptualized his "levels of argumentation" (he does not refer to learning or reflection explicitly) within the context of policy making. Schön refers to the professions of engineering, architecture, management, psychotherapy, and town planning to show how professionals meet challenges by engaging in a process of "reflection-in-action."

A distinction is usually made between two levels of learning or reflection: lower-order versus higher-order discourse (Fischer 1980) or reflection (Schön 1983), single-loop versus double-loop learning (Argyris and Schön 1978; Sabatier and

Jenkins-Smith 1993), or adaptive versus generative learning (Senge 1990). Although the contexts and the exact definitions differ, the distinction between the two types of learning in all cases is more or less similar. In the lower-order category, the learning process is a kind of technical or instrumental learning. It is reactive, short-term focused, within a context of fixed objectives (as applied to policy), a context of fixing new problems within the same problem definition and procedures (as applied to organization), or a context of technological design optimization (Elzen et al. 2004; Brown et al. 2003). Lower-order learning occurs when people become aware of their position in the network and the possible differences in actor roles, agendas, perceptions, values, and interests among the actors. The awareness of these differences enhances the instrumental rationality of the actors in the sense they realize that the other actors enable or constrain the achievement of certain goals (Van de Poel and Zwart 2010, p. 181).

In the higher-order category of learning, the objectives, problem definitions and procedures are not taken for granted but questioned and explored (Elzen et al. 2004). It therefore involves the redefinition of policy goals and changes in norms and values (Brown et al. 2003). This higher-order learning is usually more long-term focused, involving reflection on goals and values. In the remainder of the text, we will use the term "reflective learning" to refer to these higher-order learning processes.

The effect of learning in organizational settings, like networks, can be conceived as a threefold shift (Brown et al. 2003): (1) a shift in framing of the problem; (2) a shift in the selection of the principle approaches to solve the problem and in comparing and choosing between alternative solutions, and (3) a shift in the relationships among actors in a professional network as well as the broader sphere. It is especially this third shift (a shift in the relationships among actors) together with the object of reflective learning (appreciative systems and overarching theories) which makes reflective learning such an important phenomenon for achieving a justified overlapping consensus.

In case of reflective learning, actors are not only aware of the differences in actor roles, agendas, perceptions, values and interests among the actors, but they also recognize the *legitimacy* of these other views as a consequence. People may begin to question and explore their own views and relate these to those of others. Additionally, reflective learning may help to distinguish between private and public values, that is, between arguments that are and that are not legitimate and important for an actor fulfilling a specific role in the network. This may lead to the situation that people reach a consensus on particular issues or on the values that should dominate decision-making. But even if people do not reach such a consensus, reflective learning might enable what Dryzek and Niemeyer (2006) call a normative meta-consensus, that is, recognition of the legitimacy of a variety of the disputed values, without agreement on the values that should predominate (which would establish a normative consensus). This normative meta-consensus ("agreement to disagree") is different from Rawls overlapping consensus because it might also include non-public values that are excluded by Rawls. Dryzek and Niemeyer suggest that a normative meta-consensus might be instrumental in achieving agreement on specific issues (2006, p. 640).

In the specific situation of R&D or innovation networks, reflective learning may also include reflection on the desirable properties of the network as a whole. By reflecting on the project objective and problem definition, people may reconsider their original views regarding the project and frame the problem differently. They may, for example, realize that they need to include other actors in the network to address particular issues or to represent important stakeholders. Or they may come to realize that certain aspects should be on the agenda, even though they thought at first that it was beyond the scope of their project or not their own responsibility. Reflective learning might thus contribute to achieving an overlapping consensus among actors within a network displaying a large variety of value systems and background theories.

6.4 Two Examples of Application of the Approach

In this section, we illustrate the approach described in Sect. 6.2 with an application to two research projects: the development of a new waste water treatment technology, the so-called Granular Sludge Sequencing Batch Reactor (GSBR), and the development of an in-house monitoring system for elderly people (ALwEN).

6.4.1 The GSBR Project

The GSBR project concerned the development of an innovative wastewater treatment technology. One drawback of traditional biological wastewater treatment plants is their large space demand or footprint, which is caused by the use of separate settling tanks and the slow settling velocity of the sludge. In the aerobic GSBR technology both size increasing factors are addressed. By using high-density granules, the time needed for the sludge to sink to the bottom at the end of each cycle is substantially reduced. Subsequently, the shorter deposit time increases the throughput of the installation and reduces the footprint. Second, it is hoped that different ecological zones inside the granules will be able to take care for the entire treatment process in one reactor instead of several separate tanks.

The GSBR technology has been developed at the Department of Biotechnology, Delft University of Technology, the Netherlands. After successful laboratory experiments, the Dutch Foundation for Applied Water Research (STOWA) was found willing to invest in the project. Additional funds were acquired for a PhD-project, which was funded by the Dutch Technology Foundation STW. Finally, an international engineering and consulting firm, with water management technology as one of its main domains, showed interest in the commercial exploration of the GSBR technology. This firm was in charge of the research at the pilot plant, operated by a local water board.

One of the crucial elements in the development of the technology was the upscaling of the three-liter laboratory reactor to an outdoor pilot plant of 1.5 m^3. This upscaling was partly based on several unproven assumptions about which microbiological mechanisms are at work. The ethical parallel research, therefore, focused on the question of how this incompleteness of knowledge was dealt with in the choice of scaling-up steps. Incomplete knowledge can lead to the introduction of unknown risks, which may become manifest in the research done during the development of the technology, but also later in the eventual use of this technology. The aim of the ethical parallel research was to find out how risks and uncertainties were handled and how this was open to improvement. The ethical parallel research consisted of qualitative research, based on interviews, document analysis, attendance of technical meetings, and the organization of an interactive session in the Group Decision Room (GDR; an electronic brainstorming facility) with the different stakeholders, where questions related to risks and responsibilities were addressed.

During the ethical parallel research, it was observed that the risks due to so-called secondary emissions (i.e., unwanted but not yet regulated substances in the effluent) were not addressed by the engineers and researchers involved. The users of the technology delegated the responsibility for dealing with the risk of secondary emissions to the research phase, and most of the researchers also allocated the risk to a phase for which they in turn bore no responsibility. Nobody therefore assumed responsibility for dealing with this risk. The argument of the researchers was that the impact of the risks due to these secondary emissions was negligible and that potential problems were to be addressed during the operational phase. This was based on the presumed similarity between biological processes in traditional sewage plants and the biological processes in the GSBR technology. As a result, the issue who is responsible for checking or preventing secondary emissions never became an object of discussion. This was reinforced by the fact that, on the basis of the existing knowledge, it could not be concluded that such emissions are a serious cause of concern. The situation was rather one of insufficient knowledge. Thus the question arose which of the actors in the network were responsible for reducing this knowledge deficiency, and which actors were responsible for reducing potential secondary emissions in case they turn out to be a serious concern" (Van de Poel and Zwart 2010).

The ethical parallel researchers confronted the technical researchers with their observation that the secondary emissions were currently not investigated in the project. This prompted discussion among the technical researchers about the need to address these emissions. In a co-authored paper, written jointly by the technical and ethical researchers, it was argued that ethics in innovation "forces the researchers to think about all aspects of their innovation and to discuss where and at which moment certain aspects are investigated and who's responsibility it is. Even if the outcome would be that it is not an important issue at that specific moment in the development (because of lack of legislation or minor impact on environment or humans), all risks need a conscious judgment where and when to be investigated" (De Kreuk et al. 2010, p. 221).

Another observation that the ethical researchers shared with the technical researchers and users was the mismatch between their risk perceptions. The technical researchers estimated the risks stemming from the new technology considerably lower than the users did. This prompted discussion and led to a better understanding between the technical researchers and users.

This increased awareness of additional (secondary) risks and the discussions prompted by the ethical parallel research, ultimately led to new research proposals in which explicit attention was given to the behavior of the technology under extreme conditions and possible toxic compound. Especially the additional risks foreseen by the users, were an “extra stimulus for the researchers to continue investigating the limits of the system” (ibid.). New research questions were formulated and additional funding was applied for, in the hope to decrease both the economic risk of introducing the technology to the market and the environmental risk.

6.4.2 The ALwEN Project

The second project where the network approach was applied is the ALwEN project, which concerned an R&D project aimed at developing Ambient Intelligence-based technology (where ALwEN was an acronym for Ambient Assisted Living with Embedded Networks). The European term Ambient Intelligence (AmI) – or equivalently ubiquitous and pervasive computing (USA) or ubiquitous networking (Japan) – reflects a vision of the future of ICT in which “intelligence” is embedded in virtually everything around us. This intelligence is built into tiny processors and sensors which are integrated into everyday objects (such as clothes and furniture) and which are able to communicate directly with each other without the need of traditional PC input and output media (Mattern 2004). Already with the composition of the research team, the ALwEN consortium tried to differentiate itself from other projects by capturing the whole trajectory of fundamental research to the development of a prototype application and ultimately commercial exploitation. In order to do so, four universities, two independent industrial research institutes, one clinical partner, and a consortium of 12 SMEs (Small and Medium Enterprises) cooperated. At the start, the ambitions of the ALwEN team were high. In the project proposal it had set itself the goal of bringing the engineering science for such a technology to the level of commercial product viability. The aim was to develop a prototype Ambient Assisted Living (AAL) type application to monitor and assist the activities of the elderly in the context of an elderly home. The original project proposal mentioned a pilot application, in which the concepts and techniques, which were required to safeguard security and privacy of the information collected through the wireless sensor networks (WSN), could be tested and further developed. Rather than focusing on isolated aspects of the technology, the ALwEN consortium aimed at a more systematic and integral approach to scientifically understand all interactions, interferences, and cross-relations of WSN technology, such as to find the right balance and trade-offs on the system level.

In the project, a use case[4] was developed to serve as an example of what can be done with WSNs and to focus the work of the demonstration activities of the project. The use case described a situation of in-house monitoring of the daily activities of a patient with COPD, a chronic lung disease. End users, including health care professionals, were consulted to clarify their wishes and demands with respect to the environment to be created. Although this use case was supposed to support the cooperation between the technical and the clinical partners in the project, this cooperation proved difficult and the team members adopted an attitude of waiting. The technical partners seemed to be waiting for instructions "how to establish social acceptance," whereas the clinical partners seemed to be unaware of the possibilities of WSN technology.

At the start of the project "social acceptance" was identified as one of the crucial points for the successful implementation of the technology. In addition to technical and economic goals, the project consortium had therefore set itself the following two goals related to the social acceptance of the application:

- Quality of life: the project will develop a pilot application to monitor and assist the activities of the elderly in the context of an elderly home. The main societal criterion for the success of this application is that it contributes to the quality of life of the elderly, in the sense that it helps them to maintain their independent living.
- Security and privacy: even though personal information may be pervasively collected and distributed over wireless communication channels, the security of the information and the privacy of the patient must be guaranteed.

Since "social acceptance" was identified as a necessary criterion for the success of the project, this was taken as the starting point for the ethical parallel research. This research consisted of the following phases. After an observation period of several months, a series of interviews was carried out with 13 representatives of the different institutional partners involved in the project. On the basis of these interviews, a list of relevant moral issues was established. Some issues were mentioned by several of the interviewees, other issues were mentioned only once or twice. In the interviews, it was also asked who – according to the interviewee – was responsible for addressing these issues.

The next step in the ethical parallel research was the organization of an interactive workshop. On the basis of the interview results, some striking issues were selected and explored in more detail during this workshop. The issues that were selected did not represent a complete set of sufficient conditions to get the technology accepted. Since it would not be possible to discuss all issues extensively, those

---

[4] A use case is a prose description of the system's actions that are required to perform a certain task. It describes the system's behavior when interacting with an agent outside the system (i.e. a user). The use case is detailed into several use-case scenarios, which each describe a different use-case "flow of events" or "'path" through the use case (Jacobson and Ng 2005, p. 54). Since its introduction in 1987 by Jacobsen, use-case modeling has become the standard in software and systems engineering to elicit the needs of stakeholders and to capture requirements. As such it provides early validation of what needs to be built into the system (Jacobson 1987; Jacobson and Ng 2005).

issues were selected that were, on the basis of the interview results, expected to prompt some disagreements (either because the interviewees gave different answers to the question whether or not that issue should in fact be addressed within the context of the project, or because the people disagreed on the question whose responsibility it was to address it). This list included the following issues/tasks: making sure that the application does not interfere with everyday life (invisibility of technology); setting the requirements of the security of this application (how secure is secure enough?); striking the right balance between user friendliness, reliability and functionality; making sure that end users (patients, their family & friends, clinicians) are able and willing to use the application; starting a broad societal discussion about the desirability of these kinds of (monitoring) applications; addressing questions related to data storage and data access (legal aspects); inventorying/monitoring potential risks of the present application; identifying how technological choices affect the social acceptance.

At the start of the workshop, the participants were first asked to list the different project phases or activities. This resulted in a list of project activities (such as, project management; simulation; experimentation; exploitation of application). The participants were then asked to distribute the tasks over the different project activities. For some issues, there was immediate agreement. An example is the task of identifying how technological choices affect the social acceptance. It was agreed that this is something that needs to be addressed in the experimentation phase. There was disagreement on the question which issues to include in the project. Regarding the task to address legal issues related to data access and storage, some participants argued that this was beyond the scope of the project, whereas other thought it should be included.

The next step in the workshop was to discuss the remaining points of disagreement. Meeting support software was used to allow anonymous discussions in the hope that this would give the participants more freedom to speak freely and not to give the “desirable answer” for the sake of the atmosphere of the project. The moderators of the workshop also gave some “theoretical background” on responsibility, stemming from the philosophical literature.

After the discussion, the participants were asked to distribute the selected tasks again. Compared to the previous distribution exercise, there was significantly more agreement. Regarding the scope of the project, the people reached a consensus on the question what should be part of the project. At this stage, everyone agreed that also the legal issue of data storage and access should be addressed by them, even if only in terms of informing the respective regulatory bodies that this issue is currently not adequately covered by existing regulations and law.

The main conclusion of the workshop was that it is hard to single out one partner responsible for all “moral issues.” During the workshop, the people agreed that establishing social acceptance really is a joint effort. Although it is primarily up to the management to coordinate this joint effort, all partners should have a commitment to the “ethics” of the project. Another conclusion was that the project, until then, was too much focused on fundamental research and too little on more applied research, including experimentation. This result prompted some refocus of the work and soon after the workshop, a meeting was scheduled in which plans for more realistic experiments were made. In this meeting, both the technical and non-technical requirements pertaining to the use case were discussed.

Ten months after this workshop, two additional interviews were carried out with two project members to evaluate the intervention by the ethicist and to discuss the lasting effects of the workshop. In these interviews, two team members with a formal role in the management of the project were asked after their experiences with the ethical parallel investigations. They both expressed their appreciation of the involvement of an ethicist in the project and they argued that it had helped them giving "ethics" a more profound role in the project. One interviewee argued that the involvement of an ethicist can help making technical people more aware of things they otherwise overlook. In the current project, for example, the technical researchers at first overlooked the issue of interference with everyday life and the effect a technology can have on other people than the end users. At the start of the project, the technical researchers did not fully realize that intended end users of a medical technology may not be willing to use this technology if the technology requires the presence of technical people on a too frequent basis. This led, for example, to the new requirement to use as much as possible energy saving batteries with a longer lifetime because that would allow for a longer use of the technology without technicians having to change the batteries too often. Regarding future projects, both team members thought it should be common practice to give an ethicist a formal role in technical projects during the whole course of the project. Both interviewees indicated that they see ethics as a relevant, but for themselves unknown, field of expertise. Since they considered themselves lacking the ethical expertise, they thought future projects would gain in quality by composing multi-disciplinary teams. They considered ethics not just as instrumental to successful technology implementation but rather as an end in itself. Ideally, ethics should be seen as a "non-functional requirement that you cannot ignore," one of the interviewees remarked.

6.4.3 Evaluation of the Experiences

In both cases, the interventions by the ethicist led to an increased awareness for the ethical and social aspects of the technology. In the GSBR case, discussion of the differences in risk perception made the technical researchers realize that they should make a more well-informed trade-off between the different risks. In the ALwEN case, the researchers realized that they needed to shift their focus to experimentation in order to assess how the technology affects the end users. By discussing responsibility on different levels of abstraction, they became aware of what is needed to be responsible, and came to agree on who is responsible for doing what. In the ALwEN case, this led to the realization that addressing the ethical and social aspects requires a joint effort of the various partners and a more prominent role of the project management for coordinating this.

In both cases, the researchers were able to achieve an overlapping consensus on the issues that should be addressed within the framework of the project. Moreover, there was also convergence on the actual distribution of the corresponding responsibilities, leading to a consensus on some specific points. This consensus was lacking before the ethicists "intervened" in the project. If analyzed in terms of the two procedural norms, it seems that in both cases (increased) reflective learning and (increased) openness and inclusiveness were important ingredients for achieving a justified overlapping consensus. In the GSBR case, some reflective learning was witnessed during the last period of the ethical research, possibly because of the intervention in the network by the ethical parallel researchers. After the interviews, and partly as result of the feedback provided by the ethical parallel researchers, the technical researchers applied for new research funding, including investigations into ecological impacts. This might point at some form of second-order reflective learning because the technical researchers had adjusted their norms about which risks deserve serious attention.

In the ALwEN project, both levels of learning seemed to have occurred. Various participants in the workshop remarked that they had become more aware of moral issues, which is a clear sign of first-order learning. However, the workshop prompted second-order learning processes as well. Some senior participants worried about the fairness of the load for the PhD and postdoctoral researchers, which indicates an openness to other people's interests and values. One of the workshop participants with a consequentialist approach to the project, came to realize that the fairness of the work load and the rights of end users should be taken into account, whereas one of the participants with a more "deontological stance," realized that efficacy and efficiency matter as well.

Moreover, the emphasis that the work requires a joint effort, spanning all the project activities, also points to (second-order) reflective learning processes. Lastly, the problem definition itself became object of discussion, which is an indication of reflective learning. The ALwEN project also showed increased inclusiveness. Soon after the ethical workshop, some tangible attempts were made to include end users.

To conclude, these two cases suggest that the network approach can be a fruitful way of intervening in an R&D project. Mapping the network is instructive for it lays bare some relevant issues without assuming them a *priori*. In the other two phases, the idea of pluralism was paramount. Recognition of this pluralism contributes to gaining the trust of the technical researchers. Both cases suggest that confrontation with the observations may already prompt sufficient discussion to raise awareness for these issues. It is then, within the bounds of reasonable pluralism, up to the relevant actors and stakeholders to assess which issues fall beyond the scope of the project and consequently, how to address these issues.

The two cases also suggest that the two procedural norms, reflective learning and openness and inclusiveness, are instrumental in achieving a justified overlapping consensus in the network. This means that these norms could be used as indicators for the moral assessment of R&D networks. Moreover, the norms suggest clues for how the achievement of a justified overlapping consensus in an R&D network can be improved.

6.5 Strengths, Limits, and Challenges

Compared to early TA approaches as described in Chap. 2, which are sometimes claimed to have little effect on the technical work being done, and supported by our experiences with the GSBR and ALwEN project, we think that the network approach has four major strengths. While the first and second point may be true for the other approaches discussed in this volume as well, the third and fourth point reflect two additional strengths, which are more specific for the network approach.

First, the network approach fits in the family of recently developed approaches that offer the possibility to identify moral issues in situations of uncertainty and indeterminacy about the final consequences of technological innovations. The moral issues related to the future use of the technology are often still unknown or in flux. To an important extent, this is often due to distance between the research that is being done in the scientific laboratories and the eventual applications. The network approach has been specifically developed to deal with these difficulties. It uses the researchers and the networks in which they participate, including the other relevant parties in these networks, as entry for identifying moral issues. What is important is that ethical issues often do not arise at the level of individual researchers and engineers, but at more collective levels. The two examples discussed illustrate this. For example, the problem that nobody assumes responsibility for investigating secondary emissions is only visible at the collective level at not at the level of individuals in isolation. Also the awaiting attitude between the technical and clinical partners in the ALwEN project was only visible at the collective level. The network approach is thus especially appropriate to identify moral issues that arise at the collective level.

Second, the network approach can be applied already at the early stages of technological development and it addresses the researchers involved. It can therefore generate insights that immediately influence R&D and design decisions. As such, the approach may help focusing the technical work in a way that moral issues are better addressed. In both the GSBR and ALwEN project, the feedback provided by the ethicists to the engineers and scientists involved actually led to adjustments in the original R&D trajectory. In the ALwEN project, the workshop that was part of the ethical parallel research raised awareness among the researchers to pay more attention to the application of the technology. This led to a shift in focus towards clinical experimentation, which was initiated soon after the workshop.

Third, the network approach identifies moral issues in their real-world context. The network of people involved in R&D is used as entry for identifying moral issues. This is not to say that issues are only seen as ethical if the actors themselves recognize them as such; the ethicist has the room to articulate certain ethical issues even if the actors themselves do not recognize them as such. However, this articulation takes into account the empirical facts of the situation. Hence, the approach is normative but with due attention to the particularities of the project. This requires that the ethicist is well-informed about the project, including some of its technical details. Of course, an ethicist does not need to be as specialized as the technical team member themselves, but some familiarity with the topic and the application domain is required. Without this familiarity, the ethicist's involvement runs the risk of becoming either irrelevant (for example when the ethicist does not sufficiently understand the application of the technology) or moralizing (when the ethicist does not understand what could and what could not reasonably be asked from the technical researchers). Analyzing the network in terms of the two procedural norms may also be necessary for the effectiveness of the approach. Other early engagement approaches are sometimes criticized for focusing too much on what happens in the laboratory and not taking into account actors and conditions in the external context of laboratory science (such as institutional constraints and democratic attunements of science in society), as a result of which "the enhanced reflexivity of the lab science component itself may not be sufficient to alter innovation trajectories so as to better achieve the public good, as an open-ended dynamic process" (Wynne 2011, p. 791; see also Swierstra and Jelsma 2006). By analyzing the network itself in terms of the two procedural norms, the impact of this approach can be taken beyond the individual level.

Fourth, by using insights from political philosophy, the approach can do justice to moral pluralism. People have different views on what to include in the project and by whom it should be addressed. This does not mean that it is up to people's own private opinions whether or not to address particular issues, but neither is it something to be decided by someone outside the project without any attention to the existing pluralism. By including procedural elements in the approach (for example, the demand of openness and the requirement to base one's arguments on public reason) a balance can be struck between a top-down approach in which the engineering and technological researchers have little to say about which moral issues to address in the project and an approach in which the individual people only address those issues which happen to attract their attention.

However, notwithstanding these advantages, the approach has its limits as well. Limits to the approach are mainly due to the need to attain the cooperation and trust of the technological researchers. This is not just a practical challenge but raises also more fundamental questions about what ethical issues and perspectives can be addressed and about the independence of the ethical researchers. In the ALwEN project, the issue of trust was explicitly evaluated in a second series of interviews. The interviewees indicated that an open attitude towards the project is a prerequisite for trust.

For the ethicist involved, the major challenge is to maintain her independence. On the one hand, the ethicist should become an insider. This is important because the ethicist should get sufficiently acquainted with the project and the technology to be able to describe the network and identify the relevant issues. At the same time, the ethicist should maintain her critical stance as well. Independence is necessary to assure that there is room left to give and accept criticism (Van de Poel 2008: 36; see also Schuurbiers 2011; Doorn and Nihlén Fahlquist 2010). A combination of personal skills and institutional safeguards is probably required to deal with this challenge. Personal skills are required to maintain the relations of trust and criticism, to temper the eagerness to report moral concerns, and discuss those with the technical researchers first. Clear arrangements about reporting the ethical parallel research, including the right to respond to claims made by the ethicist, or complaints about unwarranted claims, should be part of the institutional safeguards to make the cooperation a successful one.

Second, the applicability of this approach may be limited by the topic being studied. When the R&D project concerns a relatively controversial research topic (for example, the development of military equipment), the project team may be more inclined to exclude particular organizations, especially those that are critical towards the application domain. The same holds when the cooperation is vulnerable, for example because the internal competition is high, which may in turn lead to a tension due to conflicting loyalties. In those situations, the project team will probably be less inclined to invite an ethicist in their midst. This also holds for the type of partners involved. Larger industrial companies are usually less open towards outsiders, which is to some extent understandable given the scope of their work and stakes of these types of companies.

On a more fundamental level, the challenge is to deal with broader ethical issues that transcend the boundaries of the R&D networks. Such issues include the question whether R&D on a particular technology is desirable in the first place, whether or not certain applications should be banned, but also questions related to global justice, and the availability of and access to a particular technology. Answering these kinds of questions requires, again, attentiveness to moral pluralism. As such the approach seems promising. However, although insights from R&D may be quite relevant here, it can be argued that these issues need to be addressed at other locations than the laboratory or the R&D network. Compared to the other approaches discussed in this book, the network approach already takes into account extra-laboratory factors. However, the main focus still lies with the R&D activities. The network approach is not principally limited to R&D, though, and it is possible to shift the focus to upstream issues like, for example, R&D funding. More research is needed to see whether the network approach can be adjusted to this specific context and effectively be applied to discuss these upstream ethical issues.

Acknowledgments Neelke Doorn acknowledges the Brocher Foundation, which provided her with the opportunity to work on this paper during her stay at the Centre as visiting researcher in fall 2011.

References

Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective*. Reading: Addison-Wesley. 344.

Bressers, H., Huitema, D., & Kuks, S. M. M. (1994). Policy networks in Dutch water policy. *Environment Politics, 3*, 24–51.

Brown, H. S., Vergragt, P., Green, K., & Berchicci, L. (2003). Learning for sustainability transition through bounded socio-technical experiments in personal mobility. *Technology Analysis & Strategic Management, 15*, 291–315.

Callon, M. (1992). The dynamics of techno-economic networks. In R. Coombs, P. Saviotti, & V. Walsh (Eds.), *Technological change and company strategies* (pp. 84–106). London: Academic.

Cohen, S. J. (1997). Scientist-stakeholder collaboration in integrated assessment of climate change: Lessons from a case-study of northwest Canada. *Environmental Modelling and Assessment, 2*, 281–293.

Daboub, A. J., & Calton, J. M. (2002). Stakeholder learning dialogues: How to preserve ethical responsibility in networks. *Journal of Business Ethics, 41*, 85–98.

Daniels, N. (1979). Wide reflective equilibrium and theory acceptance in ethics. *Journal of Philosophy, 76*, 256–282.

Daniels, N. (1996). *Justice and justification: Reflective equilibrium in theory and practice* (Cambridge studies in philosophy and public policy). Cambridge: Cambridge University Press.

De Kreuk, M. K., Van de Poel, I. R., Zwart, S. D., & Van Loosdrecht, M. C. M. (2010). Ethics in innovation: Cooperation and tension. In I. R. Van de Poel & D. E. Goldberg (Eds.), *Philosophy and engineering: An emerging agenda* (pp. 215–226). Dordrecht: Springer.

Dery, D. (1984). *Problem definition in public policy*. Lawrence: University Press of Kansas.

Doorn, N. (2010). Applying Rawlsian approaches to resolve ethical issues: inventory and setting of a research agenda. *Journal of Business Ethics, 91*, 127–143.

Doorn, N., & Nihlén Fahlquist, J. A. (2010). Responsibility in engineering. Towards a new role for engineering ethicists. *Bulletin of Science, Technology & Society, 30*, 222–230.

Doorn, N. (2012). Exploring responsibility rationales in Research and Development (R&D). *Science, Technology & Human Values, 37*, 180–209.

Dryzek, J. S., & Niemeyer, S. (2006). Reconciling pluralism and consensus as political ideals. *American Journal of Political Science, 50*, 634–649.

Einsiedel, E. F., Jelsoe, E., & Breck, T. (2001). Publics at the technology table: The consensus conference in Denmark, Canada, and Australia. *Public Understanding of Science, 10*, 83–98.

Elzen, B., Enserink, B., & Smit, W. A. (1996). Socio-technical networks. How a technology studies approach may help to solve problems related to technical change. *Social Studies of Science, 26*, 95–141.

Elzen, B., Geels, F. W., & Green, K. (2004). *System innovation and the transition to sustainability: Theory, evidence and policy*. Cheltenham: Edward Elgar.

Fischer, F. (1980). *Politics, values, and public policy*. Boulder: Westview.

Fischer, F. (1995). *Evaluating public policy*. Chicago: Nelson-Hall.

Grin, J., & Hoppe, R. (1995). Toward a comparative framework for learning from experiences with interactive technology assessment. *Industrial & Environmental Crisis Quarterly, 9*, 99–120.

Grin, J., & Van der Graaf, H. (1996). Technology assessment as learning. *Science, Technology & Human Values, 21*, 72–99.

Hakansson, H. (Ed.). (1989). *Industrial technological development. A network approach*. London: Routledge.

Hanson, C. E., Palutikof, J. P., Dlugolecki, A., & Giannakopoulos, C. (2006). Bridging the gap between science and the stakeholder: The case of climate change research. *Climate Research, 31*, 121–133.

Jacobson, I. (1987). Object-oriented development in an industrial environment. *Proceedings of OOPSLA', 87*, 183–191.

Jacobson, I., & Ng, P.-W. (2005). *Aspect-oriented software development with use cases*. Upper Saddle River: Addison-Wesley.

Johanson, J., & Mattsson, L. G. (1992). Network positions and strategic action: An analytical framework. In B. Axelsson & G. Easton (Eds.), *Industrial networks: A new view of reality*. London: Routledge.

Joss, S., & Bellucci, S. (Eds.). (2002). *Participatory technology assessment. European perspectives*. Gateshead/Tyne/Wear: Athenaeum Press.

Joss, S., & Durant, J. (Eds.). (1995). *Public participation in science: The role of consensus conferences in Europe*. London: Trustees of the Science Museum.

Kensing, F. (2003). *Methods and practices in participatory design*. Copenhagen: ITU Press.

Kingdon, J. W. (1984). *Agendas, alternatives and public policies*. Toronto: Little, Brown and Company.

Klijn, E. H. (1997a). Policy networks. An overview. In W. J. M. Kickert, E. H. Klijn, & J. F. M. Koppenjan (Eds.), *Managing complex networks. Strategies for the public sector* (pp. 14–34). London: Sage.

Klijn, E. H. (1997b). Policy networks: An overview. In W. J. M. Kickert, E. H. Klijn, & J. F. M. Koppenjan (Eds.), *Managing complex networks. Strategies for the public sector* (pp. 14–34). Thousand Oaks: SAGE Publications.

Marsh, D., & Smith, M. (2000). Understanding policy networks: Towards a dialectical approach. *Political Studies, 48*, 4–21.

Mattern, F. (2004). Ubiquitous computing: Scenarios for an informatized world. In A. Zerdick et al. (Eds.), *E-merging media: Communication and the media economy of the future* (pp. 155–174). Berlin: Springer.

Mehalik, M. M., & Gorman, M. E. (2006). A framework for strategic network design assessment, decision making, and moral imagination. *Science, Technology & Human Values, 31*, 289–308.

Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Towards a theory of stakeholder identification and salience. Defining the principle of who and what really counts. *Academy of Management Review, 22*, 853–896.

Rawls, J. (1999 [1971]). *A theory of justice* (Rev. ed.). Cambridge, MA: The Belknap Press of Harvard University Press.

Reuzel, R. P. B., Van der Wilt, G. J., Ten Have, H. A. M. J., & Robbe, P. E. D. (2001). Interactive technology assessment and wide reflective equilibrium. *Journal of Medicine and Philosophy, 26*, 245–261.

Rip, A., Misa, T. J., & Schot, J. (Eds.). (1995). *Managing technology in society. The approach of constructive technology assessment*. London: Cassell Publishers Limited.

Rogers, J. D., & Bozeman, B. (2001). "Knowledge value alliances": An alternative to the R&D project focus in evaluation. *Science, Technology & Human Values, 26*, 23–55.

Saari, E., & Miettinen, R. (2001). Dynamics of change in research work: Constructing a new research area in a research group. *Science, Technology & Human Values, 26*, 300–321.

Sabatier, P. A., & Jenkins-Smith, H. C. (1993). *Policy change and learning: An advocacy coalition approach* (Theoretical Lenses on Public Policy). Boulder: Westview Press, Inc.

Schinzinger, R. (1998). Ethics on the feedback loop. *Control Engineering Practice, 6*, 239–245.

Schneider, V. (1992). The structure of policy networks. A comparison of the 'chemical control' and 'telecommunications' policy domain in Germany. *European Journal of Political Research, 21*, 109–129.

Schön, D. A. (1983). *The reflective practitioner. How professionals think in action*. New York: Basic Books.

Schot, J. W. (2001). Towards new forms of participatory technology development. *Technology Analysis & Strategic Management, 13*, 39–52.

Schot, J. W., & Rip, A. (1997). The past and future of constructive technology assessment. *Technological Forecasting and Social Change, 54*, 251–268.

Schuler, D., & Namioka, A. (Eds.). (1993). *Participatory design: Principles and practices*. Hillsdale: Erlbaum.

Schuurbiers, D. (2011). What happens in the lab: Applying midstream modulation to enhance critical reflection. *Science and Engineering Ethics, 17*, 769–788.

Sclove, R. E. (1995). *Democracy and technology*. New York: The Guilford Press.

Senge, P. M. (1990). The Leader's New Work: Building learning organizations. *Sloan Management Review, 32*, 7–23.

Smit, W. A., Elzen, B., & Enserink, B. (1998). Coordination in military socio-technical networks: Military needs, requirements and guiding principles. In C. Disco & B. Van der Meulen (Eds.), *Getting new technologies together. Studies in making sociotechnical order* (pp. 71–106). Berlin: Walter de Gruyter.

Spender, J. C. (1989). *Industry recipes: The nature and sources of managerial judgement*. Oxford: Basil Blackwell.

Spier, R. E. (2010). “Dual use” and “Intentionality”: Seeking to prevent the manifestation of deliberately harmful objectives A summary and some reflections on ‘The advancement of science and the dilemma of dual use: Why we can’t afford to fail’. *Science and Engineering Ethics, 16*, 1–6.

Swierstra, T. E., & Jelsma, J. (2006). Responsibility without moralism in techno-scientific design practice. *Science, Technology & Human Values, 31*, 309–332.

Van Asselt, M. B. A., & Rijkens-Klomp, N. (2002). A look in the mirror: Reflection on participation in integrated assessment from a methodological perspective. *Global Environmental Change, 12*, 167–184.

Van de Poel, I. R. (2008). How should we do nanoethics? A network approach for discerning ethical issues in nanotechnology. *NanoEthics, 2*, 25–38.

Van de Poel, I. R., & Van Gorp, A. C. (2006). The need for ethical reflection in engineering design: The relevance of type of design and design hierarchy. *Science, Technology & Human Values, 31*, 333–360.

Van de Poel, I. R., & Zwart, S. D. (2010). Reflective equilibrium in R&D networks. *Science, Technology & Human Values, 35*, 174–199.

Van der Bruggen, K. (2012). Possibilities, intentions and threats: Dual use in the life sciences reconsidered. *Science and Engineering Ethics, 18*, 741–756.

Wynne, B. (2011). Lab work goes social, and vice versa: Strategising public engagement processes. *Science and Engineering Ethics, 17*, 791–800.

Zwart, S. D., Van de Poel, I. R., Van Mil, H., & Brumsen, M. (2006). A network approach for distinguishing ethical issues in research and development. *Science and Engineering Ethics, 12*, 663–684.

Chapter 7

Political TA: Opening Up the Political Debate. Stimulating Early Engagement of Parliamentarians and Policy Makers on Emerging Technologies – Attempts by the Rathenau Instituut

Rinie van Est

Abstract Political technology assessment (TA) is aimed at informing and contributing to opinion formation of both members of parliament and policy makers. The modern practice of political TA presents a mix of investigation, interaction and communication activities. At the beginning of the century the TA community realised that good communication skills and (personal) links with politicians and policy makers are crucial for improving the impact of TA on the political debate. This chapter zooms in on this communicative turn and the way it influences the practice and methodology of political TA. It does so by describing various attempts made by the Rathenau Instituut, the Dutch national technology assessment (TA) organization, to involve parliamentarians and policy makers in the field of nanotechnology, or broader converging technologies. The Rathenau Instituut is institutionally positioned within the scientific domain, and outside the parliament and the government. Still its institutional task is to stimulate the political debate on science and technology. Getting out of the scientific domain and going into the political sphere requires so-called *boundary work*. This chapter illustrates and reflects on the way the Rathenau Instituut performs such boundary work. The basis for this is a trustworthy identity based on the scientific quality of its products and its quality as an organizer of participatory events. In addition, the TA organisation has to build up connections of trust to the parliament and government. Seizing opportunities to cooperate with MPs or policy makers is an important way to actively involve them in the debate on emerging technologies and building up a long-term relationship of trust between the TA institute and the political system.

R. van Est (✉)
Technology Assessment Department, Rathenau Instituut, The Hague, The Netherlands

School of Innovation Sciences, Eindhoven University of Technology, Eindhoven,
The Netherlands
e-mail: q.vanest@rathenau.nl

N. Doorn et al. (eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*,
Philosophy of Engineering and Technology 16, DOI 10.1007/978-94-007-7844-3_7,

7.1 TA and the Political Sphere

The arrival of technology assessment (TA) coincides with the political embracement of permanent innovation as a means to stimulate economic growth. Techno-economic change evokes hope for a better world – in terms of human welfare, better health care, more sustainable products –, but also creates public concerns about its impact on safety, environmental degradation, human dignity, and social equity and equality. As a consequence, science and technology have become crucial factors in many of the dominant issues and controversies of our time. The perspective of TA fits well in such a technological culture, since it combines political awareness about potential positive and negative effects of technological change with the belief or hope that one can anticipate on these effects (cf. Rip 1986).

TA can be targeted towards various relevant groups within the innovation system. TA can be directed towards members of the parliament. This is called *parliamentary TA*, that is, "technology assessment specially aimed at informing and contributing to opinion formation of the members of the parliament as main clients of the TA activity" (Deuten et al. 2011, 1). Policy makers may also be clients of TA. The function of TA then is to contribute to the policy making process by informing policy makers about the societal aspects of technological change. In this case, one might speak of *strategic TA* (Smits and Leyten 1991), or even more straightforward, *policy TA*. Activities of TA organisations can also be aimed at the general public in order to stimulate the public debate on science and technology. This can be referred to as *public TA* (Van Eijndhoven 1997). Finally, scientist and engineers can be the clients of TA. In that case TA can be used as a means to guide research and technology development from a societal perspective. Regularly the term *constructive TA* is used to pinpoint TA which is aimed at influencing technological choice and design processes (Schot and Rip 1997).

This book focuses mainly on recent experiences in various countries with implementing the latter form of TA, which is institutionally positioned very close to the R&D community. Constructive TA is guided by the political idea that social impacts of technology should not be anticipated and accommodated from *outside* the innovation process, but *within* science and technology development itself (Schot and Rip 1997). Over the last decade this approach has been promoted and implemented in many countries in various forms and under various headings, like *anticipatory TA* (Barben et al. 2008) and *real-time TA* (Guston and Sarewitz 2002) in the United States, constructive TA in the Netherlands (Schot 2003), *upstream public engagement* in the UK (Wilsdon and Willis 2004; Royal Society 2004), and *reflective action research* in Flanders (Goorden et al. 2008). Also the European Commission came to officially support involving informed citizens in science, 'particularly in defining the priorities of publicly funded research' (EC 2002, 8).

This article seems to be the odd man out. It does not describe a TA program or project which aims at 'opening up the laboratory.' In contrast, this chapter is about 'opening up the political sphere', including the parliamentary debate and the policy decision making process. This form of TA might be named *political TA*, that is, technology assessment specially aimed at informing and contributing to opinion formation of both members of parliament and policy makers as main clients of the TA activity. Political TA, thus, includes or combines both parliamentary TA and policy TA.

Reflective TA practitioners describe TA as a “scientific, interactive and communicative process” (Bütschi et al. 2004, 14). Accordingly, the modern practice of TA presents a mix of investigation, interaction and communication activities (cf. Van Est 2011b). Since TA developed in the United States in the 1960s as a special branch of policy analysis that dealt with the social impact of science and technology, doing science and writing reports has always been a core aspect of TA. Since the mid-1980s, TA also became regarded as a process for involving experts, stakeholders, and citizens. Setting up public participatory processes required new managerial and methodological skills from TA organisations. Over the years, this participatory turn has led to the development of a large toolbox of participatory methods, including citizens’ panels, and scenario workshops (cf. Joss and Bellucci 2002; Slocum 2003). At the beginning of this century the (parliamentary) TA community woke up to the fact that scientific investigation and organising dialogues within TA projects are not sufficient to achieve an impact in the political domain (Decker and Ladikas 2004). It was realised that good communication skills are crucial for improving the impact of TA on the political debate.

Again this communicative turn also requires novel managerial and methodological skills from TA organisations and new ways of thinking and talking about TA methodology, and defining and managing TA projects (Van Est 2011b). Through this communicative turn seducing MPs and policy makers to become engaged and tackle various relevant public issues has become a central goal of political TA (Van Est & Brom 2012). This implies, for example, that TA practitioners build up personal contacts in the political domain, and keep up to date with the way in which governmental agencies and the parliament deal with certain issues. Moreover it implies delivering custom-made information to the political community, that is, timely scientific information that is framed in a way that connects to the agenda and vocabulary of the political community. In addition, instead of one TA study or participatory event, building up a steady relationship with the political domain may require a whole series of TA activities over a considerable period of time.

This article zooms in on the communicative turn within the modern TA practice and the way it influences the TA methodology. For this, several experiences of the Rathenau Instituut, the Dutch national TA organisation, are described with engaging parliamentarians and policy makers in an early stage with emerging technologies. It is important to acknowledge that the way in which the Rathenau Instituut tries to inform and involve MPs and policy makers is enabled and constrained by its institutional and organisational contexts. The following section, therefore, discusses some characteristics of those contexts and the way they impact political TA. Section 7.3 describes various concrete project activities that aimed to engage MPs and policy makers in the debate on nanotechnology and converging technologies. Finally, Sect. 7.4 will sum up some insights with regards to the question of how to involve political actors in an early stage in the debate on emerging technologies.

7.2 Political TA at the Rathenau Instituut[1]

This section aims to characterise in general terms the way the Rathenau Instituut performs political TA. As said, this is enabled and constrained by its institutional and organisational level. On both levels we want to describe how the Rathenau Instituut relates to the parliament, government, science and society. Moreover, we will discuss the implications of its institutional and organisational settings for the way in which the Rathenau Instituut performs political TA.

On the institutional level the central question is where the TA organisation is positioned within the institutional space that is constituted by the four above named societal spheres. Various elements define the institutional position of a TA organisation. A first question that needs to be addressed is to what perceived institutional problem is the TA organisation regarded as an institutional solution. Second, what is the institutional task of the TA organisation, and (part of that) who are its clients?

On the organisational level, the question becomes which actors from which social spheres have a say in the TA organisation's doing? Various relevant activities can be discerned: who funds the organisation, who are the clients, who evaluates the organisation and who decides on the members of the evaluation committee, who are the members of the Board, who appoints them, who is involved in setting up the work program, who needs to be consulted, and who performs the TA project activities, and who is involved in these various activities?

7.2.1 Dual Nature of the Institutional Position of the Rathenau Instituut

In 1975, physicist and MP Jan Terlouw proposed to install a TA organization closely linked to the Dutch Parliament. Terlouw had been inspired by the American Office of Technology Assessment (OTA). The establishment of that congressional TA bureau in 1972 was meant to provide Congress with "*unbiased* information concerning the physical, biological, economic, social and political effects of (technological) application" and "*early indications* of the probable beneficial and adverse impacts of the applications of technology …" (U.S. Congress 1972) (italics added by authors). Terlouw's motion was rejected, mainly because technological development did not play central stage in the political debate at that time.

This changed, however, rapidly during the second part of the 1970s. For example, the influence of automation in industry on employment became a major topic, and an advisory group Rathenau (named after its chairman) was set up to study the

---

[1] This subsection is inspired by and based on research done within work package 2.1. of the PACITA-project, which is a EU financed project under FP7. PACITA stands for Parliaments and Civil Society in Technology Assessment (Ganzevles and Van Est 2012; see also www.pacitaproject.eu/)

societal effects of micro-electronics (Adviesgroep Rathenau 1980). The group recommended to set up a permanent TA institute. In particular, the rise of social activism related to the technology-dominated issue of nuclear power put science and technology on the public and political agenda. At the end of the 1970s public resistance against nuclear power led to a political stalemate in the field of energy policy, and created a legitimacy crisis of the State. To get out of this crisis, the Dutch government, guided by a political tradition of accommodation, initiated a grand public debate on energy between 1981 and 1984.

In that same period the Ministry of Education and Science developed the *White Paper on the Integration of Science and Technology in Society* (O&W 1984). This so-called IWTS-paper proposed to set up a bureau within the ministry with the TA task to signal and study both potential positive and negative societal effects of science and technology, but also to stimulate societal opinion forming and to bring these insights and opinions into the political decision making process (O&W 1984, 12). The Parliament agreed with the idea to set up a TA organisation, but opposed the idea to position it within the government. The parliament demanded a more independent TA institute at arm’s length of the government. The parliament also did not plea for a TA bureau in or near the parliament. It was finally decided to set up the Rathenau Instituut (then NOTA) in 1986 as a TA organisation which falls under the administrative responsibility of the Royal Netherlands Academy of Arts and Sciences (KNAW).

The Rathenau Instituut got the task to stimulate the public and political debate on the societal consequences of scientific and technological change (see Box 7.1). The installation decree states that for performing its tasks it will “search for connections with relevant societal actors” (OC&W 2009, 1). Actively involving various kinds of social actors in TA is thus seen as a legitimate way to integrate societal interests and values in the TA and ultimately in the policy making process. In other words, Dutch MPs regard public engagement and deliberation as a legitimate add-on to representative democracy (Van Eijndhoven 1997). In practice, therefore, most Rathenau projects contain a mixture of research activities and participatory events to involve experts, stakeholders and citizens in TA. In 2004, at the request of the Minister of Education, Culture and Science, a new task was added to the Rathenau Instituut’s remit: Science System Assessment (SciSA), with the aim to study the dynamics of the Dutch innovation system. The current two formal tasks of the Rathenau Instituut are described in Box 7.1.

It can be concluded that the the Rathenau Instituut’s position towards the four discerned societal spheres has a dual nature. On the one hand the Rathenau Instituut is positioned within the ‘heart’ of the scientific community, namely the KNAW. But on the other hand, the Rathenau Instituut has a special position within the KNAW, because it is the only institute within the KNAW that has its own board, and the KNAW has to place the state aid at the Institute’s disposal unaltered and without delay. Moreover, the Rathenau Instituut is not doing research for scientific reasons, but with the aim to contribute to societal debate and political opinion forming. One could argue that the fact that the Rathenau Instituut is not primarily focussed on publishing in peer reviewed journals put the organisation at the fringes of the

Box 7.1: Formal Task Description of the Rathenau Instituut (OC&W 2009, 1)

Technology Assessment Task

The role of the institute is to contribute to societal debate and the formation of political opinion on issues that relate to or are consequence of scientific and technological developments. This specifically includes the ethical, social, cultural and legal aspects of such developments. In particular, the institute facilitates the formation of political opinion in both chambers of the Parliament of the Netherlands and in the European Parliament.

Science System Assessment Task

The institute continues to work on increasing our understanding of how science works as a system, and, in doing so, integrating all available data and making it more easily accessible as well as acquiring any data that remains unavailable. As part of this process, it is the role of the institute to make information available to the Dutch Cabinet, to both chambers of the Parliament of the Netherlands.

scientific community. Also the Rathenau Instituut's position towards the political process shows a dual nature. On the one side, the institute is positioned at arm's length from the political process. On the other side the Rathenau Instituut is physically located in the political centre of the Netherlands (The Hague) and has the Dutch Parliament and Cabinet as its main clients. Finally, with regards to society the Rathenau Instituut has no formal bonds with any societal organisation. At the same time, its task is to stimulate public debate and actively search for connections with relevant societal actors.

7.2.2 *Shared Control Over the Rathenau Instituut*

Basically two types of TA organisations that perform parliamentary TA can be distinguished: (1) organisations within or very close to the parliament with the sole task of informing the parliament, and (2) organisations at arm's length of the parliament and often funded by the government with the task of informing the parliament and stimulating social debate (Deuten et al. 2011, 20, Figure 2). Petermann (2000) named these models *instrumental* and *discursive*, respectively.

The parliamentary TA office in France, OPECTS (Office Parlementaire d'Evaluation des Choix Scientifiques et Technologiques) exemplifies the first model. Van Eijndhoven (1997, 271) argues that of all the parliamentary TA organizations in Europe OPECST is "the one most intimately linked with parliament, because it is the parliamentarians themselves who conduct the assessments". In France, the TA process is fully integrated in the committee structure of the parliamentary system. The TA project is led by so-called rapporteurs, who are selected from the members of OPECST, and are responsible for writing the TA report (Deuten et al. 2011: 21). Rapporteurs can organize hearings and missions in France or abroad. They are assisted by parliamentary civil servants, and if needed, supported by a working group or steering committee consisting of experts outside the parliament. One might say that the French parliament has full control over the TA process.

The Rathenau Instituut fits the second model. The influence of the parliament on the activities of the Rathenau Instituut is rather weak. First of all, the Rathenau Instituut is not funded by the Parliament but by the Ministry of Education Culture and Science. Moreover, active MPs are not represented in its Board. Half of the Board members are appointed by the Ministry, the other half by the Royal Academy of Arts and Sciences. The majority of the Board consists of scientists, while a number of seats in the board are reserved for representatives from civil society. The government also installs the evaluation committee, mainly consisting of representatives from science and society, and often led by a former well-known politician. Finally, the government and wider society are included as formal addressees, next to the parliament. In its role as client the parliament has, an indirect, but crucial, influence on the way the TA organization functions. In contrast to OPECST, however, MPs do not determine the activities of the Rathenau Instituut, since also actors from government, science and society have a say in the work of the Rathenau Instituut.

7.2.3 Political TA as Boundary Work

The institutional position and tasks and organizational control mechanisms have a profound influence on the way in which political TA is performed. In France parliamentary TA is performed within the confines of the Parliament. The involvement of MPs in the TA is self-evident, since the French MPs perform TA themselves, supported by the staff of OPECST. The situation in the Netherlands is very different. The Rathenau Instituut is institutionally positioned inside the scientific domain, but outside the Parliament and outside the government. Moreover, no formalized organizational ways of informing and involving MPs and/or policy makers exists. Still the Rathenau Instituut has the political task and mandate to inform MPs and policy makers about societal issues related to science and technology.

Getting out of the scientific domain and going into the political sphere requires, what Gieryn (1983, 1995) named, *boundary work*. Gieryn studied how scientists acted outside the academic sphere, for example in advising governments and communicating to the public. Scientists tried to establish and maintain an authoritative position by distinguishing themselves to 'non-science' (Van Rijswoud 2012). Boundary work also refers to the efforts going into obtaining and maintaining a clientele. With regards to political TA this implies that TA organisations need to build up connections of trust to the parliament and government. At best, MPs and policy makers should feel a kind of ownership over the information communicated to them by the TA organisation (Cruz-Castro and Sanz-Menédez 2004: 108).

The involvement of political actors in the TA activities of the Rathenau Instituut, therefore, can never be taken for granted. First of all, because the Rathenau Instituut has to compete with many other organisations for the "limited attention" of political actors. This challenges the Rathenau Instituut to acquire its own special identity in association to the political system. To connect to the political sphere the TA organization, therefore, has to create a distinct profile, which is seen as important for political actors. The next section provides an impression of the boundary work performed by the Rathenau Instituut in order to involve MPs and policy makers in social issues related to nanotechnology and converging technologies. These activities are part of the wider public and political debate on nanotechnology in the Netherlands (Van Est et al. 2012).

7.3 Engaging Parliament and Government in an Early Phase

Since 2003 the Rathenau Instituut has performed many TA studies related to nanotechnology and the broader development of NBIC convergence; the convergence of nanotechnology, biotechnology, information technology and cognitive sciences (Roco and Bainbridge 2002). Also numerous meetings and events were organised to discuss various issues in public. Some were specifically targeted at MPs or policy makers, like a public meeting in the parliament, a parliamentary hearing on the opportunities and risks of nanoparticles, and a study trip for MPs about medical nanotechnology to the High Tech Campus in Eindhoven. We have written letters to the Parliament and opinion articles in national newspapers that were published on the day a parliamentary debate was held on nanotechnology. During this period employees of the Rathenau Instituut also had many personal contacts with MPs and policy makers via private meetings and e-mail.

The remainder of this section zooms in on two specific activities. The first concerns the cooperation between the Rathenau Instituut and the Theme Commission on Technology Policy in organising the public meeting "Small technology – Big consequences" in October 2004. Besides it is described how the Rathenau Instituut interacted with the interdepartmental working group on nanotechnology (ION), which developed the Dutch Action Plan on Nanotechnology (Rijksoverheid 2008). To situate these activities first some milestones in the public debate on nanotechnology will be described.

7.3.1 Milestones in the Dutch Political Debate on Nanotechnology[2]

The Rathenau Instituut started its activities in the field of nanotechnology in Spring 2003. A conference in the European Parliament in July to examine social and ethical issues raised by nanotechnology signalled that nanotechnology had reached the political agenda. The Dutch Ministry of Science & Education anticipated on this by commissioning the Royal Netherlands Academy of Arts and Sciences (KNAW) to set up a committee to look at the opportunities and risks of nanotechnology. In February 2004 the Rathenau Instituut organised the first public workshop in the Netherlands on the chances and risks of nanoparticles. Two months later it published a first overview of applications of nanotechnology and related social issues (Van Est et al. 2004). The study presented an initial concept agenda for the public debate on nanotechnology. Together with the Parliament the institute organised a public meeting "Small technology – Big consequences" in October 2004 (Van Est and Van Keulen 2004). Two months before the KNAW (2004) had published its advisory report, which asked for a rigorous risk policy and wider public engagement.

These activities raised enough awareness among policy makers to commission the National Institute for Health and Environment (RIVM) and the Health Council to prepare inventory studies on risk issues related to nanotechnology. Moreover, some MPs got involved in the debate on nanotechnology, and stimulated the government to look at the governance of nanotechnology in an integrated manner. In November 2005 the government agreed to develop such an integral vision on nanotechnology. The government set up an interdepartmental working group on nanotechnology (ION) to prepare such a vision. The Cabinet's vision on nanotechnology was released at the end of 2006 (Rijksoverheid 2006). In July 2008 the final Action Plan Nanotechnology was published (Rijksoverheid 2008). This plan paid attention to the research agenda, the risk issue, the social and ethical issues, and announced plans for a national dialogue on nanotechnology. This dialogue ran from September 2009 to the beginning of 2011 (cf. EZ 2009).

7.3.2 Interactions Between the Rathenau Instituut and Parliament

In the autumn of 2003 the Rathenau Instituut developed a 1-year project plan for its activities in the field of nanotechnology. The plan was to organise a public event, in which MPs would play a central role. Its goal would be to find out whether it would

---

[2] A more detailed description of the activities of the Rathenau Instituut within the field of nanotechnology can be found in Van Est and Van Keulen (2004) and Van Est and Walhout (2010). A description of a wider range of activities that were organised in the Netherlands to bring a public perspective into the development of nanotechnology can be found in Van Est et al. (2012).

be necessary to organise a large public debate on nanotechnology in the Netherlands. And if not, what (if any) kind of actions should be taken? Below it is described how the Rathenau Instituut managed to get the Parliament involved in organising the public meeting *Small technology – Big consequences* on October 13, 2004 in the Dutch parliament building.

A first step to achieve this was publishing *To value the small…* (Van Est et al. 2004), which delivered a concept agenda for a public debate on nanotechnology. This study was positioned as a background paper in preparation of such a public meeting. Secondly, to engage actors – ranging from social scientists and nanoscientists to NGOs and policy makers – in the discussion on nanotechnology a series of workshops were organized in the first half of 2004 on the health effects of nanoparticles, nano-electronics, biomedical nanotechnology and nanotechnology in the food sector. In that same period, the Rathenau Instituut aroused the interest of the parliamentary Theme Commission on Technology Policy in organizing a public event on nanotechnology. The Theme Commission presented a new type of parliamentary committee. On May 6 2002 the flamboyant politician Pim Fortuyn was killed. During his life Fortuyn had severely criticised the functioning of the Parliament. After his death the Parliament itself picked up this topic. One of the ideas was to create the possibility of setting up a Theme Commission. Such a Commission would deal with specific important long term societal issues that go across various ministries. It was thought to give MPs a platform to reflect deeply on those issues relatively separate from the direct political decision making process. Moreover, the Theme Commission would provide room to experiment with new working methods in order to stimulate the interaction between the Parliament and society. Three Theme Commissions were set up. One of them was the Theme Commission on Technology Policy, which ran from September 2003 to September 2006.

This Theme Commission was supported by the Research and Verification Bureau (OVB) of the Second Chamber of the Parliament. This Bureau had been set up in 2001 to strengthen the control function of the Second Chamber. As part of its strategy to get in closer contact with the Parliament, the Rathenau Instituut cooperated with the OVB by sending employees of the Rathenau Instituut to work for OVB. Since the rathenau Instituut has a track record in dealing with science, technology and society issues it was only logical that the Rathenau employee who at the time worked at the OVB became the secretariat of the Theme Commission on Technology Policy. This gave the Rathenau Instituut a rather direct connection to the parliamentary Theme Commission. Members of the Theme Commission were particularly interested in experimenting with new working methods. Since the Rathenau Instituut has a lot of experience with participatory TA methods it offered the Theme Commission its services. In addition, it suggested to organize a meeting around nanotechnology, since nanotechnology was seen as a key enabling technology for future innovations. The Theme Commission agreed to cooperate with the Rathenau Instituut. This cooperation provided a good opportunity for the Rathenau Instituut to inform MPs about nanotechnology and to involve them in the debate. Rathenau Instituut employees where invited several times to present their findings at various meetings of the Theme Commission, where the content and set-up of the public meeting was discussed. Instead of opting for the classical hearing, it was chosen to organise four interactive debates between stakeholders from different societal domains, like social scientists, nanoscientists, businesspeople, societal organisations, government, politics, and the public. These different groups of stakeholders got a distinct place within the debating arena. In this way the public meeting was thought to present an early reflection of the rising public debate.

The interactions with the MPs during the preparatory Commission meetings gave the Rathenau Instituut a good insight into the questions around nanotechnology that were on the minds of the MPs. Moreover, the fact that the Theme Commission hosted the public meeting “Small technology – Big consequences” on October 2004 created a sense of ownership among its members. Moreover, the public meeting turned out to be an inspiring way to inform MPs and policy makers about the state of affairs of the public debate on nanotechnology in the Netherlands. At the end of the meeting the chairperson of the Commission stated “We want to start a social debate on nanotechnology. Now only the in-crowd is talking about it.” After the public meeting the discussion of nanotechnology remained on the parliamentary agenda. Nanotechnology became part of the annual debates on innovation within the Standing Committee on Economic Affairs. These debates stimulated the government in November 2005 to start developing an integral policy vision on nanotechnology (Tweede Kamer 2005).

7.3.3 Interactions Between the Rathenau Instituut and Government

To develop an integral Cabinet vision the government set up the Interdepartmental working group on nanotechnology (ION), with representatives of almost all ministries. The ION became *the* policy nexus where different public and policy perspectives – stimulating research and innovation, dealing with risks, identifying ethical concerns and public engagement – had to be integrated. The ION contacted various players within the field of nanotechnology, and also visited the Rathenau Instituut. Over the years the Rathenau employees were regularly invited to present their findings during ION meetings, and developed strong personal contacts with some members of ION.

The government had commissioned the Health Council to advice the Cabinet on nanotechnology. This advice was published in April 2006 (Gezondheidsraad 2006). The Rathenau Instituut had built up close contacts with the Health Council and was therefore prepared to organise a timely follow-up workshop shortly after the advice was published specially directed to ION. At the workshop the implications of the report of the Health Council for the Cabinet’s vision were discussed with all kinds of societal actors. The ION used the results for preparing the Cabinet’s vision (Rijksoverheid 2006), which was published at the end of the year. This green paper set the scene for establishing a strategic research agenda, developing a risk governance policy, identification and monitoring of social and ethical issues, involving stakeholders and organizing a societal dialogue. The next task for ION was to elaborate on these perspectives in order to come up with an Action Plan on Nanotechnology. Responsibilities for each perspective were distributed across the participating ministries. For example, the Ministry of Environmental Affairs got the lead in risk governance, the Ministry of Justice got the task to monitor social and ethical issues, and the Ministry of Economic Affairs became responsible for organising a societal dialogue.

In order to link its activities to the policy making process, the Rathenau Instituut decided to focus on these three governance challenges (Van Est and Walhout 2010). Its activities in the field of risk governance are described in Van Est et al. (2012). The following concentrates on the other two policy lines. At the time the Netherlands had three decades of experience with state-initiated forms of public engagement in science and technology related issues, with varying success (Van Est et al. 2002; Van Est 2011a). In particular, the government had received a lot of criticism for the way it had organized the GM-food debate '*Eten en Genen*' in 2001. In that debate the government chose to focus its efforts towards the 'general public.' The existing engaged civil society organizations (CSOs) were merely positioned as sources of information for the 'general public.' The environmental CSOs felt sidetracked and decided to boycott the state-initiated debate. Policy makers wanted to avoid nanotechnology becoming 'the next GM', but also wanted to prevent the nanotechnology dialogue becoming 'the next GM-food dialogue .' The working group within the ION responsible for organizing the National Dialogue on Nanotechnology was aware of the many complexities involved and was open to make use of the methodological expertise and hands-on experience at hand at the Rathenau Instituut with organising public debates.

There is a strong awareness within the Rathenau Instituut that the appropriateness of activities to stimulate a certain debate depends heavily on the condition of the debate itself (Van Est and Walhout 2010). To make sensible decisions about how to organise a National Dialogue on Nanotechnology one first has to map the state of the art of the public dialogue on nanotechnology. The Rathenau Instituut took the initiative to do that. This led to the publication *Ten lessons for a nanodialogue: The Dutch debate about nanotechnology thus far* (Hanssen et al. 2008). An important input for this study was a survey and a workshop which asked civil society organisations (CSOs) about their expectations with regards to nanotechnology and a national dialogue. The government implemented the report's key recommendation in its Action Plan on Nanotechnology (Rijksoverheid 2008): make a clear distinction between the debate about risk policy and the exploration of emerging social and ethical issues. For the risk issue a stakeholder platform was set up, while the national dialogue would aim at discussing broader social and ethical issues. The dialogue would be coordinated by a committee, consisting of a broad representation of society. In March 2009 the Cabinet installed the Commission Societal Dialogue Nanotechnology, with the task 'to implement a broad discussion in which viewpoints and opinions could be expressed by all kind of stakeholders and publics' (EZ 2009). The Commission invited the Rathenau Instituut several times to present its ideas on how to set up the national dialogue. The Commission embraced the advice to enable and empower societal actors to organize activities themselves. By mean of an open call for proposals, the Commission facilitated societal actors to provide information, raise awareness and organize dialogue activities.

According to the Rathenau Instituut the social and ethical issues raised by nanotechnology as an enabling technology can only be appropriately assessed by taking the broader perspective of NBIC convergence. The Rathenau Instituut has been involved in the European discussion on NBIC convergence since 2003 (cf. Nordmann 2004; Berloznik et al. 2006; Van Est and Stemerding 2012). At the national level the institute further explored related topics under heading like ambient intelligence (Schuurman et al. 2007), synthetic biology (De Vriend 2006) and human enhancement (Van Est et al. 2008; Coenen et al. 2008). The Rathenau Instituut actively sought to link the issues in these areas to policy makers involved in foresight and strategy building with regard to emerging technologies.

In particular, close contacts developed with the working group on ethical and legal aspects within ION. Personal contacts with members of this working group for example led to further and continuous cooperation in this area between the Rathenau Instituut and the Strategy Department of the Ministries of Justice and Interior Affairs. In 2007 this led to a symposium on Neurosciences and legal issues. In 2009 and 2011 so-called *Knowledge Rooms* on respectively Human Enhancement (Van Keulen et al. 2009) and robotics (De Jong 2011) were jointly organised. Such Knowledge Rooms are intended to inform the highest civil servants within the ministries about relevant emerging issues. As an indirect result of these interactions, the impact of converging technologies has been put on the strategic knowledge agenda of the national government (Strategieberaad Rijksbreed 2010).

7.4 Crossing Borders Requires Building Up a Trustworthy Identity and Connections of Trust

Its position within the scientific domain, and outside the parliament and government, forces the Rathenau Instituut to perform boundary work to get into the political sphere. The former section illustrated how such boundary work is performed in practice. This section reflects on the way this is done. The main argument is that crossing borders requires a trustworthy identity and a helping hand from actors at the other side of the border. The trustworthy identity presents the passport. And just like a real passport, an organisation has to renew its passport – that is, its license to operate – over and over again. A valid passport is necessary, but not sufficient for the border guard to let you in. You need actors within the parliament and government to invite you in. You may inform them, but it is better that political actors invite you to inform them, or even better, to cooperate with them.

7.4.1 Building Up a Trustworthy Identity: Scientific and Organizational Expertise

In promoting itself towards the parliament and government, the Rathenau Instituut has both stressed the scientific quality of its research and its quality as an organiser of participatory events (cf. Van Eijndhoven 1997). The scientific nature of TA is an important part of TA's public identity and political legitimacy. Political actors namely can use scientific information to legitimise decisions (Feldman and March 1981). In order to maintain this image the institute has to maintain a credible position in the academic world. Over the last decade the Rathenau Instituut has built up a proven track record in the field of emerging technologies. It has been involved in an early phase in TA research and discussions on nanotechnology, converging technologies, synthetic biology and ambient intelligence. Such a track record helps to be seen as an expert on the societal aspects of emerging technologies. Moreover being among the first TA organisations to study emerging technologies helps since as the saying goes 'in the country of the blind, the one-eyed man is king.'

The institute's expertise and experience with participatory methods, organising and involving all kinds of societal actors within a debate is regarded by the parliament and actors within the government as a major asset. For example, the Theme Commission's main goal was to experiment with new methods to involve society. Also the Strategy Department of the Ministry of Security and Justice likes to cooperate with the Rathenau Instituut, for its TA expertise, but also for its contacts with the scientific community it brings in and its experience and creativity in organising events. In this respect, it is important that the TA organisation is considered by political and societal actors as a reliable trusted third party.

7.4.2 Building Up Connections of Trust: Information and Cooperation

The results of TA projects need to be communicated across the border, from the scientific towards the political sphere. The agenda of the political decision making process is a crucial element here, both with respect to content and timing. For example, in 2003 the Rathenau Instituut timely anticipated that nanotechnology would soon reach the political agenda. And after the Cabinet had published its vision on nanotechnology, the Rathenau Instituut organised its activities along three policy lines within that policy vision. It is also important to adopt your communication to (the 'language' of) the client. For example, nowadays a study by the Rathenau Instituut is always accompanied by a so-called *Message*, a two page text which presents the core political message and recommendations of the study. The Message is a custom-made communication tool to reach MPs and policy makers. A final element is visibility. To connect to the political sphere one has to be into the field of vision of MPs and policy makers. Getting into the media is an important way to achieve that. The short Message has proven to be a good tool to get media attention. Writing opinion articles is another way of getting into the public media and getting on the radar of relevant political actors.

The former section shows that this information-based communication strategy can be a useful leg up for a participatory model of communication. Sending out timely, custom-made information and being visible in the media all help to achieve an expert position in the debate, which may lead to being invited as an expert to present your opinion. The Rathenau Instituut was regularly invited by individual MPs, the Theme Commission on Technology policy, and the ION to present its work. This one-way communication process (informing political actors) sometimes led to a two-way communication process (interacting with political actors). For example, the Rathenau Instituut cooperated with the Theme Commission on Technology Policy in organising a public meeting on nanotechnology, and with the Strategy Department of the Ministry of Security and Justice in setting up a series of Knowledge Rooms. In this participatory model of communication issue framing and identification of information needs results from interactions between TA practitioners and political actors. In this way MPs and policy makers really get actively engaged in the debate on emerging technologies. Moreover, seizing opportunities to cooperate with political actors is an important way to build up a long-term relationship of trust between the Rathenau Instituut and the political system.

References

Adviesgroep Rathenau. (1980). *Maatschappelijke gevolgen van micro-elektronica*. Den Haag: Staatsuitgeverij.

Barben, D., Fisher, E., Selin, C., & Guston, D. H. (2008). Anticipatory governance of nanotechnology: Foresight, engagement, and integration. In E. Hackett, O. Amsterdamska, M. Lynch, & J. Wajcman (Eds.), *The handbook of science and technology studies* (pp. 979–1000). Cambridge: MIT Press.

Berloznik, R., Casert, R., Deboelpaep, R., van Est, R., Enzing, C., van Lieshout, M., & Versleijen, A. (Eds.). (2006). *Technology assessment on converging technologies*. Brussels: European Parliament, STOA.

Bütschi, D., Carius, R., Decker, M., Gram, S., Grunwald, A., Machleidt, P., Steyaert, S., & van Est, R. (2004). The practice of TA; Science, interaction and communication. In M. Decker & M. Ladikas (Eds.), *Bridges between science, society and policy: Technology assessment – Methods and impacts*. Berlin: Springer.

Coenen, C., Schuijff, M., Smits, M., & Hennen, L. (2008). *Shifting boundaries, changing concepts, and the governance of human enhancement*. Brussels: European Parliament, STOA.

Cruz-Castro, L., & Sanz-Menédez, L. (2004). Shaping the impact: The institutional context of technology assessment. In M. Decker & M. Ladikas (Eds.), *Bridges between science, society and policy: Technology assessment – Methods and impacts* (pp. 101–127). Berlin: Springer.

Decker, M., & Ladikas, M. (Eds.). (2004). *Bridges between science, society and policy: Technology assessment – Methods and impacts*. Berlin: Springer.

De Jong, J. B., red. (2011). *Kenniskamer intelligente robots: Feiten, fabels en fictie*. Directie Strategie, Ministerie van Veiligheid en Justitie. De Jong: Den Haag.

Deuten, J., Enzing, C., Nagle, M., van Til, J., & Arnold, E. (2011). *Parliamentary technology assessment in Europe*. Brussels: European Parliament, STOA. Draft.

De Vriend, H. (2006). *Constructing life: Early social reflections on the emerging field of synthetic biology*. The Hague: Rathenau Instituut.

European Commission (EC). (2002). *Science and society action plan*. Brussels: European Commission.

EZ. (2009). Instellingsbesluit Commissie maatschappelijke dialoog nanotechnologie. *Staatscourant*, nr. 61, 30 maart.

Feldman, M. S., & March, J. G. (1981). Information in organisations and signal and symbol. *Administrative Science Quarterly, 26*, 171–186.

Ganzevles, J., & van Est, R. (Eds.). (2012). *TA practices in Europe*. Deliverable 2.2. PACITA Collaborative project on mobilisation and mutual learning actions in European Parliamentary Technology Assessment.

Gezondheidsraad. (2006). *Health significance of nanotechnologies*. The Hague: Health Council of the Netherlands.

Gieryn, T. F. (1983). Boundary work at the demarcation of science from non-science: Strains and interests in professional ideologies and scientists. *American Sociological Review, 48*, 781–795.

Gieryn, T. F. (1995). Boundaries of science. In S. Jasanoff, G. E. Marke, J. C. Petersen, & T. Pinch (Eds.), *Handbook of science and technology studies* (pp. 393–443). Thousand Oaks: Sage.

Goorden, L., Van Oudheusden, M., Evers, J., & Deblonde, M. (2008). Nanotechnologies for tomorrow's society: A case for reflective action research in Flanders, Belgium. In E. Fisher, C. Selin, & J. Wetmore (Eds.), *Presenting futures: Yearbook of nanotechnology in society* (Vol. 1, pp. 163–182). Dordrecht: Springer.

Guston, D., & Sarewitz, D. (2002). Real-time technology assessment. *Technology in Society, 24*, 93–109.

Hanssen, L., Walhout, B., & Van Est, R. (2008). *Ten lessons for a nanodialogue: The Dutch debate about nanotechnology thus far*. The Hague: Rathenau Instituut.

Joss, S., & Bellucci, S. (Eds.). (2002). *Participatory technology assessment: European perspectives*. London: Centre for the Study of Democracy.

KNAW Werkgroep gevolgen nanotechnologie. (2004). *Hoe groot kan klein zijn? Enkele kanttekeningen bij onderzoek op nanometerschaal en mogelijke gevolgen van nanotechnologie*. Amsterdam: Koninklijke Nederlandse Academie van Wetenschappen.

Nordmann, A. (2004). *Converging technologies: Shaping the future of European societies.* Brussels: European Commission, High Level Expert Group "Foresighting the new technology wave".

O&W. (1984). *Beleidsnota integratie van wetenschap en technologie in de samenleving*. Den Haag: Ministerie van Onderwijs en Wetenschap.

OC&W. (2009). Instellingsbesluit Rathenau Instituut. *Staatscourant*, nr. 11024, 22 juli.

Tweede Kamer. (2005). Vergaderjaar 2004–2005, 29 338, nr. 29. Den Haag: Sdu Uitgevers.

Petermann, T. (2000). Technology assessment units in European parliamentary systems. In N. J. Vig & H. Paschen (Eds.), *Parliaments and technology: The development of technology assessment in Europe* (pp. 37–61). New York: State University of New York Press.

Rijksoverheid. (2006). *Cabinet view 'From small to great'*. The Hague: Dutch Government.

Rijksoverheid. (2008). *Action plan nanotechnology*. The Hague: Dutch Government.

Rip, A. (1986). Controversies as informal technology assessment. *Knowledge: Creation, Diffusion, Utilization, 8*(2), 349–371.

Roco, M., & Bainbridge, W. S. (Eds.). (2002). *Converging technologies for improving human performance: Nanotechnology, biotechnology, information technology and cognitive sciences*. Arlington: NSF/DOC.

Royal Society and Royal Academy of Engineering. (2004). *Nanoscience and nanotechnologies: Opportunities and uncertainties*. London: The Royal Society.

Schuurman, J. G., Moelart, F., Krom, A., Walhout, B., red. (2007). *Ambient Intelligence: Toekomst van de zorg of zorg van de toekomst?* Den Haag: Rathenau Instituut.

Schot, J. (2003). The contested rise of a modernist technology politics. In T. J. Misa, P. Brey, & A. Feenberg (Eds.), *Modernity and technology*. Cambridge: MIT Press.

Schot, J., & Rip, A. (1997). The past and future of constructive technology assessment. *Technological Forecasting and Social Change, 54*, 251–268.

Slocum, N. (2003). *Participatory methods toolkit: A practitioner's manual*. Brussels: King Baudouin Foundation and Flemish Institute for Science and Technology Assessment.

Smits, R., & Leyten, J. (1991). *Technology assessment: Waakhond of speurhond – Naar een integraal technologiebeleid*. Zeist: Kerckebosch.

Strategieberaad Rijksbreed. (2010). *Rijksbrede Kennisagenda – Fase 1: Trends & Ontwikkelingen*. Den Haag: Strategieberaad Rijksbreed.

U.S. Congress. (1972, October 13). *The technology act of 1972. Public Law 92–484, H.R. 10243.*

Van Eijndhoven, J. (1997). Technology assessment: Product or process? *Technological Forecasting and Social Change, 54*, 269–286.

Van Est, R. (2011a). The broad challenge of public engagement in science – Commentary on 'Constitutional moments in governing science and technology'. *Journal of Science and Engineering Ethics, 4*, 639–648.

Van Est, R. (2011b). Keeping the dream alive: What ELSI-research might learn from parliamentary technology assessment. In S. Cozzens & J. Wetmore (Eds.), *Nanotechnology and the challenges of equity, equality and development, part 5* (The yearbook of nanotechnology in society, 1, Vol. 2, pp. 409–421). Dordrecht/Heidelberg/London/New York: Springer Science + Business Media.

Van Est, R., & Brom, F. (2012). Technology assessment: Analytic and democratic practice. In R. Chadwick (Ed.), *Encyclopedia of applied ethics* (2nd ed., Vol. 4, pp. 306–320). San Diego: Academic.

Van Est, R., & Stemerding, D. (Eds.). (2012). *European governance challenges in bio-engineering – Making perfect life: Bio-engineering (in) the 21st century. Final report*. Brussels: European Parliament, STOA.

Van Est, R., & van Keulen, I. (2004). "Small technology – Big consequences": Building up the Dutch debate on nanotechnology from the bottom. *Technikfolgenabschätzung – Theorie und Praxis, 13*(3), 72–79.

Van Est, R., & Walhout, B. (2010). Waiting for nano – Very actively: A long-term view on the role of the Rathenau Instituut in stimulating the Dutch debate on nanotechnology. *Technikfolgenabschätzung – Theorie und Praxis, 2*, 67–74.

Van Est, R., Van Eijndhoven, J., Aarts, W., & Loeber, A. (2002). The Netherlands: Seeking to involve wider public in technology assessment. In S. Joss & S. Bellucci (Eds.), *Participatory technology assessment: European perspectives* (pp. 108–125). London: Centre for the Study of Democracy.

Van Est, R., Malsch, I., & Rip, A. (2004). *Om het kleine te waarderen… Een schets van nanotechnologie: publiek debat, toepassingsgebieden en maatschappelijke aandachtspunten*. Den Haag: Rathenau Instituut.

Van Est, R., Klaassen, P., Schuijff, M., & Smits, M. (2008). *Future man – No future man: Connecting the cultural, political and technological dots of human enhancement*. The Hague: NWO.

Van Est, R., Walhout, B., Rerimassie, V., Stemerding, D., & Hanssen, L. (2012). Governance of nanotechnology in the Netherlands: Informing and engaging in different social spheres. *International Journal of Emerging Technologies and Society (iJETS)., 10*, 6–26.

Van Keulen, I., Brom, F., Quast, J., & de Jong, J. (2009). *Kenniskamer human enhancement: Feiten, fabels en ficties*. Den Haag: Ministerie van Binnenlandse Zaken, Ministerie van Justitie, Rathenau Instituut.

Van Rijswoud, E. (2012). *Public faces of science: Experts and identity work in the boundary zone of science, policy and public debate*. Nijmegen: Radbout University Nijmegen.

Wilsdon, J., & Willis, R. (2004). *See-through science: Why public engagement needs to move upstream*. London: Demos.

Part III
Reflections

Chapter 8
Integrating Ethicists and Social Scientists into Cutting Edge Research and Technological Development

Michael E. Gorman, Antonio Calleja-López, Shannon N. Conley, and Farzad Mahootian

Abstract Ethics is an integral part of scientific and technological thinking, whether the practitioners recognize it or not. The kind of expertise the scientist gains about ethics and the ethicist about science can be labeled interactional. An interactional expert learns the language of a community with sufficient depth to communicate on matters like research strategy. The concept of trading zone is employed to understand how people from different perspectives and agencies can work together to define a common goal in a way that would be acceptable to their core communities. These concepts will be honed by applying them to case-studies of social scientists

---

This material is based upon work supported by the National Science Foundation under Grant No. #0849101.

M.E. Gorman (✉)
Department of Science, Technology & Society, University of Virginia,
351 McCormick Road, Charlottesville, VA 22904-4744, USA
e-mail: meg3c@virginia.edu

A. Calleja-López
Department of Sociology, Philosophy and Anthropology, University of Exeter, Exeter, UK

Department of Metaphysics and Current Trends in Philosophy,
Ethics and Political Philosophy, University of Seville, Seville, Spain
e-mail: alotria0@gmail.com

S.N. Conley
School of Politics and Global Studies, Arizona State University, Tempe, AZ, USA
e-mail: snconley@asu.edu

F. Mahootian
Liberal Studies Program, New York University,
726 Broadway – Room 664, New York, NY 10003, USA
e-mail: Farzad.Mahootian@asu.edu

N. Doorn et al. (eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*,
Philosophy of Engineering and Technology 16, DOI 10.1007/978-94-007-7844-3_8,

and humanists who integrated themselves into science and engineering laboratories. This paper will particularly focus on the value of complementing interactional expertise with the acquisition of somatic tacit knowledge.

Concern with the ethical and social implications of research should be an integral part of scientific and technological thinking (Gorman et al. 2009). There are codes and guidelines for the ethical conduct of research: for example, plagiarism and fraud are both unethical and illegal, there are norms for assigning authorship, for mentoring graduate students,[1] for protecting research subjects and strict rules regarding laboratory safety and in vivo research. Scientists and engineers must not only be in compliance with these guidelines, they must play a role in shaping them, and in thinking about situations not covered in any current guidelines—the kinds of challenges and opportunities that are created on the research frontier, where discoveries and innovations change what is possible.

Scientists are also urged by the National Science Foundation to address the broader impacts of their work.[2] Broader impact criteria make "critical reflection on the relation of scientific discovery to societal priorities" part of "the scientific research process itself." (Frodeman and Parker 2009, p. 304). Scientists and engineers are not trained to do this, so here they could use assistance from those better trained to think about societal and ethical implications.

One method to achieve this objective, and to integrate broader impacts and intellectual merit, is to follow Davis Baird's advice, provided in testimony before the Senate Committee on Commerce, Science and Transportation, May 1, 2003:

> Ethicists need to go into the lab to understand what's possible. Scientists and engineers need to engage with humanists to start thinking about this aspect of their work. Only thus, working together in dialog, will we make genuine progress on the societal and ethical issues that nanotechnology poses.

Indeed, the twenty-first century Nanotechnology Research and Development Act of 2003 calls for the investigation of the societal impacts of publicly funded nanotechnology R&D. The Act explicitly identifies the objective of "integrating research on societal, ethical, and environmental concerns with nanotechnology research and development" (US Congress 2003, Public Law no. 108–153). Moreover, the 2011 National Nanotechnology Initiative (NNI) strategic plan includes engagement with multiple stakeholders:

> Build collaborations among the relevant communities (e.g., consumers, engineers, ethicists, manufacturers, nongovernmental organizations, regulators, and scientists—including social and behavioral scientists) to enable prompt consideration of the potential risks and benefits of research breakthroughs and to provide perspectives on new research directions (NNI 2011, objective 4.3.2, p. 31).

---

[1] The National Academy of Sciences book on the Responsible Conduct of Science is an extremely useful resource on these topics (http://www.nap.edu/openbook.php?record_id=4917)

[2] See http://www.nsf.gov/nsb/publications/2011/nsb1141.pdf for the latest recommendations by the National Science Board.

This chapter and the one by Fisher and Schuurbiers investigate one way of bringing this expertise into the laboratory: by integrating humanists and social scientists into laboratories and research teams. This chapter will focus on the capabilities required to do this sort of integration and compare them with three engagement experiences, one upstream and two midstream. The end result will be suggestions for further research that would provide valuable lessons for others who want to try engagement as a strategy for integrating the ethical and the social into the lab—and for integrating the latest and best expertise on science and engineering into STS.

8.1 Modulation Projects as Trading Zones

How can communication and collaboration occur across C.P. Snow's two cultures, when even participants in different approaches to a single discipline may have perspectives and practices so different from each other that they appear incommensurable? Consider, for example, the ongoing argument in anthropology over whether the field ought to be a science or not (Wade 2010). The scientific and non-scientific approaches to anthropology are apparently incommensurable, in part because some of those uncomfortable with calling anthropology a science regard science as a hegemonic, colonial enterprise. Similar debates between Science, Technology and Society (STS) scholars and scientists were labeled the 'science wars'; some scientists argued that STS scholars did not have the expertise to study science , and STS scholars counter-argued that they wanted to study science like any other culture, without privileging the scientific world-view from the outset. Again, these positions appear incommensurable, but more recent work has showed a way to solve this problem, involving two key concepts: trading zones and interactional expertise.

8.1.1 Trading Zones

Peter Galison's (1997) solution to the problem of incommensurability is the development of trading zones (see also Gorman 2010). Different cultures who do not understand each others' perspectives and practices can still trade; all they need is to establish what each wants and negotiate an exchange that satisfies both sides. From simple trades like this, more complex exchanges may develop that require a shared language; pidgins, for example, develop to facilitate exchanges in places like ports where multiple cultures mingle. Out of a pidgin may emerge a creole that becomes a language of its own. Most of our modern languages began as creoles.

Trading zones can permit scientists and engineers to work across apparently incommensurable barriers of language, culture and practice. Here the exchanges are often in pursuit of a common goal. For example, Galison shows how the development of radar required scientists, engineers, military experts and others to work together to develop a system that was critical to the survival of Britain in World War II.

Patrick McCray has described the kinds of trading zones that go into the development of giant telescopes (2004).[3]

The academic analogy is the kind of exchanges of knowledge, time and resources that led to the establishment of new fields like science and technology studies, born out of a shared interest in studying science among a few sociologists, anthropologists, historians and psychologists who found they needed additional expertise and so started slowly to develop a creole in order to found a new society, journals and discipline.

Therefore, the apparent incommensurability between scientists and STS scholars requires development of trading zones where these disciplines can exchange knowledge and work together on projects. Good examples are provided by the two Centers for Nanotechnology in Society at ASU (www.cns.asu.edu) and University of California Santa Barbara (www.cns.ucsb.edu) where STS work on nanotechnology includes engagement—and sometimes collaboration—with scientists.

But what about the expertise problem posed in the science wars: how can someone without deep background in a scientific domain understand enough to study and/or work with the members of that expertise community?[4]

8.1.2 Interactional Expertise

An interactional expert learns to communicate with members of another discipline or culture by immersing her/himself in the community and learning both the explicit and tacit aspects of the language. Collins (2004) did this with the gravitational wave physics community; after a long immersion period, he could converse fluently with experts and even pass as one on a kind of Turing test in which a gravitational wave physicist had to determine which of two individuals was a member of his community and which not. Collins passed the test by being chosen as the gravitational wave physicist (Collins and Evans 2007). Interactional expertise is therefore a solution to the central concern of scientists' in the science wars: Collins demonstrated it was possible to learn enough to interact intelligently and deeply with members of a specialized community without having to learn all the laboratory and mathematical skills necessary to do the research.[5]

---

[3] A system like radar involved work on the boundaries of several disciplines; therefore, different expertises represented the emerging system and sub-systems in unique ways characteristic of what Leigh Starr called a boundary object (Bowker and Star 2000).

Boundary objects and systems can facilitate coordination in trading zones, especially ones that involve creating systems, where working prototypes and detailed plans can serve a role similar to an emerging creole.

[4] Galison has a PhD in physics as well as in history of science, which facilitates his study of trading zones—but even in his case, he does not have expertise across all the elements of the trading zones.

[5] Collins and Evans have developed a program of research using the imitation game, a kind of Turing test for interactional expertise, but that important work is beyond the scope of this paper (Collins and Evans 2007).

It might therefore be possible for humanist or social scientist to gain sufficient interactional expertise to follow Baird's advice and 'go into the laboratory' to engage with the work. But if this engagement includes participation, there will also be trades involved; the laboratory members must see a benefit for making their time available to this newcomer who has little or no expertise in their area of science and engineering.

What follows are several case studies that illustrate the potential for, and limitations of, this kind of engagement, using trading zones and interactional expertises as a framework.

8.2 A Micro Nanotechnology Trading Zone

Gorman (psychologist) shared a graduate student with Groves (material scientist).[6] Gorman had advised Masters and PhD students in Systems Engineering students on topics like innovative environmental design where ethical and social issues are explicit components of the degree (Gorman et al. 2004). This project with Groves was different because the student wanted to get a Materials Science degree and Gorman and Groves were equal co-advisors. Since Gorman was not an expert in the research domain, he played the role of an embedded social scientist, asking lots of questions and discussing options as the work progressed.

Groves suggested the student make a chart of global problems and see which ones could be mapped onto nanotechnologies that corresponded to the capabilities of this small team and the constraints of a Masters thesis—an apparently impossible set of constraints, until Groves suggested development of a metaphoric language, based on the fact that all three of us liked to hike. He said the global problems were like distant mountains. The student's project corresponded to a bridge across a stream on a route that could lead to the mountain. This meant the student did not have to think about solving a major health or environmental issue; she could simply focus on a bridge that might be built in the course of a Masters thesis.

The small team expanded by adding a bio-medical engineer who was studying blood flow in the hopes of making an eventual contribution to an understanding of artheriosclerosis. Groves made artheriosclerosis a foothill that lay between the bridge and the distant mountain representing a significant reduction in heart disease. The student looked at creating a nano platform that would hold a blood cell in place long enough for its deformation during flow to be modeled. This kind of a nano platform was far too complicated for a Masters thesis, so the student ended up working on what nano materials might be used for such a platform and conducting experiments on one of the best options.[7]

---

[6] Societal dimensions of nanotechnology (SES 0210452).

[7] For more details on this project and its outcomes, see Gorman et al. (2004).

At one point late in the project, Groves rushed into Gorman's office and spent a half-hour explaining why the consideration of these societal dimensions made the science better. Groves said he would not have stuck with such a difficult problem for so long; the societal goal motivated persistence down a line of research that eventually led to a patent application. The science was better because the team tackled a more ambitious problem and stayed with it until they made a small discovery about the way two metal oxides could be deposited on a surface.

The Gorman/Groves project involved starting a new line of research, and therefore constitutes an example of upstream modulation. Mahajan (engineer) encouraged Fisher (humanist) to embed in his laboratory, interacting with researchers in ongoing projects. Therefore, Fisher was engaged in midstream modulation. Downstream modulation would occur in the late stages of a research project, when the insertion of a humanist or social scientist would be least likely to affect the course of that particular line of research, though modulation could certainly affect dissemination and application of the research, as well as recommendations for regulation of the research products, where necessary.

Fisher also found evidence that modulation could improve the science. His questions about decision-making processes in the lab stimulated one researcher to reflect on alternative possibilities. This resulted in a decision to replace a key compound in the synthesis with a related, but as yet un-thought of and untried, compound. The replacement turned out to be successful in increasing yield with a cleaner process that reduced the fouling of the instruments involved in synthesizing the compound (Fisher and Mahajan 2010). This example illustrates the role of the embedded humanist in encouraging the researcher to reflect—which can open up new possibilities that improve the science.

8.3 Interactional Expertise

To ask thoughtful questions about research strategy, Gorman had to learn enough about the relevant concepts to interact intelligently with Groves and the Materials Science student. Gorman was not able to do any of the research proposed by the student, but he had to understand it well enough to have input on what experiments the student ought to do, and why.

This kind of expertise is called interactional. The canonical example is how sociologist of science Harry Collins gained fluency in the language of gravitational wave physicists sufficient to participate in deep conversations about the domain with members of the community and even understand their jokes (Collins 2004). But Collins could not actually conduct a gravitational wave experiment or do the mathematics.

Similarly, Gorman gained his interactional expertise by asking questions when meeting with his collaborators and then trying to make suggestions based on what he had learned. He also visited the laboratory and learned a bit about the processes involved in conducting the research, but could not conduct any experiments himself. Like Collins, Gorman had spent his career studying how scientists thought and worked, which facilitated his ability to grasp the methodological issues.

Ribeiro (2007) studied technology transfer from the steel industry in Japan to Brazil. The transfer was facilitated by interpreters who had to become interactional experts in order to facilitated knowledge transfer. The Japanese interpreters arrived in Japan with no knowledge of the steel industry or of Portugese, and they gained knowledge by embedding themselves in the work process as much as possible, observing operations, learning the Portugese for technical terms until they became fluent in 'blast furnace Portugese'. This initial mastery of jargon follows the trading zone progression, but the end result for the translators was mastery of an existing language, not development of a creole; here the interactional expertise gained by the translators was sufficient to facilitate transfer. Whereas Collins acquired interactional expertise solely through immersion in the language community, the translators acquired interactional expertise not only through conversation with the operators but also through close observation of what they were doing.

Collins and Evans developed a kind of Turing test for interactional expertise and Collins passed it. In the test, a gravitational wave physicist asked a series of brief questions designed to discover which of two respondents was a gravitational wave physicist (Collins et al. 2006). The exchange was done on-line, and Collins was identified slightly more often as the physicist than the actual scientist.

To determine whether Gorman had actually achieved interactional expertise, he could have been put through a similar test in the domain he and Groves and the student worked in. Such a test would have been difficult to set up because Gorman's interactional expertise was research focused, not discipline focused. But Gorman's level of interactional expertise was almost certainly well below Collins'—the latter spent years with the gravitational wave community, and the former only spent parts of 2 years working with Groves and the student.

An embedded humanist or social scientist is going to need to gain a certain amount of interactional expertise similar to what Gorman acquired—otherwise he or she will not be able to ask intelligent, provocative questions about research strategy. In the rest of this paper, we will consider the value added by the acquisition of what Collins calls somatic tacit knowledge (Collins 2010), or the kind of embodied knowledge that seems to reside in the eyes and hands and is virtually impossible to describe linguistically. Erik Fisher's Socio-Technical Integration Research (STIR) project (NSF #0849101) involved midstream modulation studies across 13 countries (see Chap. 5 by Fisher and Schuurbiers). Calleja-López and Conley were part of this project, and their experiences included both the acquisition of interactional expertise and somatic tacit knowledge.

8.4 Doing a PCR

Conley did her first modulation project in a genetics laboratory in Vancouver, that was exploring novel prenatal diagnosis techniques and genetic causes of premature infertility focusing on chromosomal abnormalities, epigenetic changes, and disorders that are linked to the placenta. The laboratory included a female PI, two PhD students, one post doc, one lab manager, one lab tech, one masters student, an undergrad, and a research coordinator. Conley had no prior knowledge of genetics and met with initial hesitance and distrust from the lab. Laboratory members expressed fears that Conley was on a mission to "dig up" dirt on the laboratory. Laboratory members initially viewed Conley (a political scientist) as an "ethics expert" and feared that she would tell them how to be more ethical scientists. In addition to having to navigate the doubts and distrust of the individual laboratory members, she experienced an immediate language gap and her queries about jargon could not be answered by searching on Google. She needed either to develop a creole across the laboratory, or gain interactional expertise herself. Either could facilitate development of a micro trading zone, in which laboratory members exchanged their time and knowledge with her because both understood the value added by the sharing.

Conley decided her best strategy was to "go native" and learn how to act and function like a member of this community. She felt she needed to probe for diverse ways of engagement/understanding, beyond observation and interviews. So when laboratory members urged her to learn how to do a poymerase chain reaction (PCR) in order to copy a DNA sequence (see Rabinow 1996 for the story of the discovery of this technique) she jumped at the opportunity. She worked closely with a postdoc and also benefited from a one page form known in the laboratory as a "PCR sheet" that scaffolded the activities involved in doing a PCR. The rows of the table corresponded to different samples and the columns reminded the researcher to keep track of concentrations of chemicals. Doing a PCR was like cooking: practicing with a recipe and taking notes on improvements.

Conley's postdoctoral mentor felt that her PCR was so good it he photographed it as an exemplar for others. She was able to transfer the PCR skill to a doctoral student in another laboratory who was unable to do a good PCR. She also engaged in other material practices, such as making gels to be used in experiments.

Collins' interactional expert learns to understand the tacit knowledge associated with the language. Conley learned aspects of the discourse of the laboratory and also gained somatic tacit knowledge. This ability to do the procedural work made her more of a member of the laboratory: she was willing to master one of the core activities and could even transfer it to another laboratory.

Engagement is two-way. Conley's credibility in the laboratory helped her train the post doc in basic sociology, ethnography and "Science and Technology Studies 101." The post doc was not only sensitized to the social context of his laboratory, he gained awareness of the ways in which the lab could be perceived "from the outside" and became actively involved in shaping those same perceptions among the rest of the research staff. This re-description resulted in observable changes in his own professional behavior, including, notably, reaching out to Conley for her assistance in finding speakers for a workshop on genetics and society and to draft a paper about the lab and its situatedness in its community.

Following a casual bench-side conversation on the meaning of "responsible innovation" in different contexts, the post doc encouraged Conley (with the approval of the laboratory director) to lead a laboratory meeting on the topic. Conley was scheduled to lead the lab meeting towards the end of her 3-month project. Rather than lecture the laboratory members on responsible innovation, Conley instead utilized the meeting time as a venue for open dialogue and brainstorming with the scientists. In order to stimulate the researchers' own thoughts on responsible innovation, Conley fed insights from individual interviews over the past 3 months back into the larger group. Conley highlighted concerns of one of the PhD students, who in an initial interview, had expressed concerns regarding patient engagement and outreach efforts, and that to keep patients abreast of laboratory research through workshops and other outreach efforts would be one way of engaging in "responsible innovation." The laboratory director responded that workshops might not be the best way to engage patients, as the samples that the laboratory collected were sensitive in nature (placenta samples from failed, aborted, and successful pregnancies), but a newsletter with updates on laboratory projects might be more appropriate. At a subsequent laboratory meeting, the laboratory director instructed the researchers to each write a "lay" account of their research, so the laboratory could send out a newsletter to patients about ongoing research.

8.5 Engagement Agents and Mobile Trading Zones

While traditional ethnographic sensibilities warn against "going native," Conley's time as an embedded researcher in the Canadian laboratory entailed a significant amount of boundary blurring, as she moved between the roles of observing social scientist to novice natural scientist in training. Conley's engagement strategy, which evolved organically from her interactions in the laboratory, entailed stepping into the shoes of those she was observing. During the 3 month period, Conley shifted from impartial observer, to benchside interlocutor, to donning a white laboratory coat and engaging in material practices alongside the other laboratory members. While progressively shifting into (and out of) these roles enabled increasingly rich and dynamic dialogues and interactions, Conley's assumption of the different roles was not static. When she put on the lab coat, she knew that she would be taking it off at the end of the day. Conley's presence became so normalized in the laboratory that her name and picture were added to the list of laboratory members on the door of the laboratory. Laboratory members would share articles they were reading with Conley, and would leave articles at her desk space in the laboratory, with notes pointing out the science-society connections of the research. While Conley's normalization into the laboratory's culture enabled for an evolving and expanding trading zone, it also was accompanied by the risks of going native, of being unable to differentiate between the roles of observer and participant, of being unable to take off the metaphorical lab coat.

Conley's experience in the laboratory equipped her with tools that enabled her to engage "between multiple dimensions of research, innovation, and policy processes" (Conley 2011). Conley was able to apply her experiences in the laboratory to her interactions with policymakers, clinicians, and stakeholders. As an individual with no prior background or training in genetics, she gained enough interactional expertise to dialogue with a multiplicity of actors involved in the socio-technical arena, in multiple international and institutional contexts. For example, Conley was able to dialogue with top British scientists about pre-implantation genetic diagnosis of embryos in relation to genetic conditions that exist on a gradient (such as autism). Conley was able to dialogue with the scientists because in her British laboratory engagement she had worked directly with a scientist researching autism. Such individuals can be thought of as "engagement agents" (see Te Kulve and Rip 2011; Conley 2011), actors who operate almost as "mobile" trading zones, with the necessary interactional expertise and "on the ground experience" to interface across multiple domains of a particular socio-technical system.

8.5.1 Making an STM Tip

Antonio Calleja-López is a philosophy student who had only high school chemistry and physics courses as background when he embedded in a laboratory at ASU working on nanotechnology applications to solar cells. He attended laboratory meetings involving three/five senior scholars and about a half-dozen PhD students, of whom two or three would speak at a given meeting. He became particularly interested in a sub-group working on electrical measurements of conductance at the nanoscale, a critical component in using nanoarrays to develop solar panels. Two female PhD students and the lab director, Stuart Lindsay, made up this sub-group.

Calleja-López noted that making Scanning Tunneling Microscope (STM) tips was a basic laboratory activity, because without these tips, laboratory members could not visualize the nanoscale surfaces they were working on.

The tips have to be made with great precision and care. Calleja-López thought if he worked on making these tips, it would be a form of payback, fostering a trading zone. Laboratory participants thought it would help him understand the work, to get in their shoes, feel what benchwork is like—its details, time consuming and sometime reiterative tasks, etc. Calleja-López had no prior laboratory experience.

8.5.2 Acquisition of Somatic Tacit Knowledge

Conley had learned how to create a PCR partly from a written guide that provided scaffolding for her learning process. Calleja-López, in contrast, had to learn purely through apprenticeship—he was not provided with a written guide. One of the female PhD students taught him via apprenticeship. The first day, she showed him how she made an STM tip. Basically, he had to carve a cone out of a gold string using electrical current. The string was dipped into a 1″ plate that contained an acidic solution. Electrons moving through a copper ring supplied the current, which Calleja-López had to modulate carefully so the cone would be properly shaped.

The graduate student taught Calleja-López how to use the sound of the current to guide his modulation of the current itself. She also looked at the color of the solution, and smelled it to judge the quality of the product. She could only judge the product a posteriori, during the manufacturing process: the colors, sounds and smells generated by the process told her things the instruments did not. Calleja-López was gaining somatic tacit knowledge (Collins 2010).

After the cone was formed, Antionio had to take the gold string out of the holder, bring the tip to microscope, adjust the light and refocus the lenses to fit his vision. so he could look carefully at the shape of the tip. Was it round? Or deformed? Or a cone?

The PhD student initially made the decisions, using experience-based judgment. In some cases a flawed tip could not be fixed, but in other cases, one could put it back into the solution and shape it again using the current. Calleja-López gradually gained somatic tacit knowledge of how to make a tip and the ability to judge a good one. He felt the main by-product of his tips was his integration with the team, so it was not just a trade in a strict sense: the team did not need him to make tips in exchange for their time, although once he became proficient, they did take advantage of his skill . Tip-making, like doing a PCR, is a sign that the embedded humanist or social scientists is willing not only to talk like a member of the culture, but learn to do an important activity.

In his second lab study, Calleja-López had the opportunity to use his background knowledge to suggest new ways of visualizing and communicating aspects of the laboratory work. During his study in Madrid, a PhD student analyzing a STM image used the metaphor of fried eggs, sunny side up or down depending on the voltage, to describe the appearance of the graphene surface displayed on the computer screen. Calleja-López suggested that it looked more like a honeycomb. This second metaphor may be more helpful in thinking the geometry of graphene, as it happened in conversation a few days later with another of the PhD students. The student was trying to describe to Calleja-López the structure of this new material, stressing the fact that graphene molecules have the appearance of arrays of hexagons, sharing sides and corners, what diminishes the total number of atoms and bonds—if we compare with a hypothetical series of discrete molecules. In order to both show understanding and help with the difficulties in the exposition, Calleja-López suggested the image of row houses common in urban areas. Row houses share walls, which means that 20 row houses need only 21 walls—rather than 40, as would be necessary in the case of free-standing houses. The researcher agreed, and mentioned the value of the example for the purposes of exposition.

These examples show how outsiders can encourage new ways of thinking—and perhaps improve the science once they gain even more interactional expertise.

8.5.3 Expertise Is in the Interaction, Not Just in the Individual

In an interaction of the kind stirred by the integration experience, there are outcomes that none of the individuals by themselves can claim hers or his. When a STIR student enters a lab and begins to gain interactional expertise and even, in the cases presented here, some somatic tacit knowledge, the student is not only changing her or his expertise, he/she is changing the collective expertise of the lab. Cognitive scientists use the term distributed to refer to the kind of cognition that relies on technologies like computers and smart phones that perform cognitive functions for the user, e.g., provide greatly extended memory and organize social contacts. These technologies impose their own additional constraints and loads (Norman 1993).

Cognitive scientists use shared cognition to refer to the way in which laboratory members rely on each others expertises and memories to perform tasks. Obviously, the line between distributed and shared cognition is blurry: a smart phone may include information on who is the expert contact on a particular problem, and which member of the team has the memory or knowledge of the state of the project and its requirements. High-functioning teams have good transactive memory, which is the knowledge of who in the team knows what and who can do certain procedures (Gorman 2002). Transactive memory is therefore an act of attributing expertise to laboratory members, based on experience—it is fluid, can evolve over time. In the highest functioning teams, individuals flow to the work, each fluidly adjusting to the others in a way that needs little discussion—the expertise here is truly shared.

The STIR student becomes part of the collective transactive memory of the laboratory, part of the shared and distributed cognition. Calleja-López's lab could distribute some of the STM tip work to him. More importantly, however, the laboratory gained the ability to look at itself the way an outsider would—and this kind of reflexivity has to emerge in the interaction. One possible end result is an expansion of research and outreach possibilities.

8.5.4 Lessons for Achieving Integration, and Suggestions for Future Research

The case studies presented here suggest that it is valuable to think of a laboratory engagement experience as a kind of micro trading zone, because this concept emphasizes that exchanges of knowledge, time and resources can occur across apparently incommensurable boundaries.

It is important for the embedded humanist to acquire interactional expertise—and for one or more members of the laboratory to reciprocate This kind of mutual interactional expertise in a micro trading zone facilitates development of a shared language, beginning with mutual understanding of a few terms and expanding if the micro trading zone lasts long enough.

Calleja-López's and Conley's experiences suggest that interactional expertise can be complemented by the ability to do hands-on laboratory work. Ribeiro's work on how translators became interactional experts provides partial support: the translators did not make steel, but in order to translated, they had to closely observe the processes. Learning laboratory procedures makes it easier for a social scientist or humanist to go from being a participant in a micro trading zone to becoming a full-fledged member of a laboratory team because it shows the social scientist or humanist is willing to leave her/his comfort zone and acquire the same skills as any other laboratory member. The “offer of effort” has more the character of a gift zone than a trading zone (Baird and Cohen 1999). These laboratory interactions are not just trades, they involve giving without exact calculation of what one receives in return. The exchange has symbolic value; it points to something deeper, a shared understanding and identity.

In Gorman’s upstream engagement experience, he learned no laboratory procedures, though he visited the laboratory and observed the student in action. But Gorman was an embedded social scientist coming in as a Masters thesis advisor, in a position of authority. Gorman never became a member of a laboratory; he functioned as a member of a committee. Upstream roles will involve more of the sort of strategic planning he engaged in.

8.6 Does the Acquisition of Interactional and Procedural Expertise Enhance Ethical Reflection?

Conley and Calleja-López think so, because the acquisition of expertise helps the embedded humanist’s put her himself in the scientist or engineer’s shoes. The central tenet of moral imagination is this ability to see another’s point of view, not just in terms of knowledge but also in terms of the underlying values that frame the laboratory. Moral imagination requires deep conversations about why, not just how and what.

While working together at the bench, Conley and the members of her genetics lab discussed a variety of ethical issues like the realization that informed consent is a process, not a document—it involves continuous conversation and communication with stakeholders. The context of this conversation is of particular interest to the midstream modulation approach: Conley reflected back to lab member the practices of sampling and analyzing each other’s blood as part of their research. Until Conley’s probing of the practice agreements about the practice and the status of the samples was very largely tacit and implicit. The reflexive modulation that resulted in this instance was profound and rapid: new explicit protocols were developed, answering many unresolved, poorly formulated and unformulated bioethical concerns.

Calleja-López went into a laboratory that had a ‘science is in society’ perspective already. It is part of ASU’s Bio-Design Institute, which focuses on bio-inspired and use-inspired design for solving problems in healthcare, sustainability and society (http://biodesign.asu.edu/about). In this setting, Calleja-López did not have to introduce thinking about societal implications into this lab ex novo. However, he did influence the PhD student who taught him how to create an STM. After the study they met several times to share updates of her work and discuss topics at the interface of STS, policy and science; she also asked him for advice about courses, projects or potential mentors to join in this area, as well as further readings. As part of this process, she enrolled in a course run by the director of the Consortium for Science Policy Outcomes (CSPO), and approached people within the institution in order to know and potentially join a project. Nevertheless, several factors, such as difficulties the process of entering this new environment, her graduation, and the weakening of their interactions after Antonio's departure, lowered her involvement. Her later movement to a position in industry further reinforced this situation. This lower involvement does not necessarily involve a lost of interest in societal issues or a neutralization of their potential implications for bench work. To ensure a continued effect, integration experiences may require longer-term contact and commitment from the social scientist or humanist, an appropriate institutional environment and processes for successful integration. The longer-term impacts of the engagement experience should be investigated.

8.7 Future Research

The engagement experiences reported and analyzed in this paper suggest three hypotheses. Embedding a humanist or social scientist into a science or engineering laboratory can:

1. Improve the ability of the laboratory to reflect on the ethical implications of their work and the way in which it fits into social systems.
2. Improve the science by opening up the possibility of new research directions based on the laboratory's increased ability to reflect.
3. Increase our understanding of the laboratory via this mutual reflection. The engagement experience is a good complement to the years of important observational work on laboratories.

To determine whether these hypotheses have any validity, we need additional methods for rigorously comparing these engagement experiences, including:

1. Having the embedded humanist or social scientists keep diaries like the one created by the cognitive scientist Jeff Shrager, who documented the way in which he acquired the expertise necessary to become a molecular biologist (Shrager 2005). Unless there is some standardization in the format and content of such diaries, it will be hard to compare them. This standardization could be achieved by developing categories worth recording and by having each student's mentor read the diaries and prompt for more information on key topics. Collection of purely quantitative temporal data along with the diary entries will facilitate comparison and generalization across laboratory engagement experiences, while preserving the unique character of each.
2. Daan Schuurbiers's chronological flowchart of the activities he observed in his engagement, including not only what was done in each of the laboratory activities, as he observed it, but also some of the ethical and societal questions raised by the activity (see Chap. 5 in this volume for a more detailed explanation).
3. Problem-behavior graphs that represent the progression of each student's attempts to learn laboratory procedures. This method involves creating flow-chart

like graphs that show the behaviors and solutions used to make progress towards solving a problem—and also all the steps that turn out not to lead to the solution (Ericsson and Simon 1984). Gorman used this method to graph Alexander Graham Bell's progress towards both a telephone patent and a working device (http://www2.iath.virginia.edu/albell/albell.html), using each change in a device as a node on the graph and Bell's response to the result from testing the device to determine whether the graph led towards or away from his stated goal (Gorman 1997). The method has also been used on Michael Faraday's discoveries (Gooding 1990). These problem behavior graphs could be produced by an observer listening to a laboratory researcher think aloud as he or she did a procedure or an experiment, or could be done by a social scientist or humanist in an effort to keep track of her or his own attempts to gain procedural knowledge. Such graphs could be associated with higher-level stages in Schuurbiers's flowcharts. For example, a box in a higher level diagram that says "make STM tip" could be linked to a detailed problem behavior graph of the procedure.

Problem-behavior graphs could also be used to chart the path of a scientific or engineering project, documenting the places where the humanist or social scientist has input and noting what happens as a result. Such objects can create a kind of visual creole that facilitates detailed discussions and comparisons of integration experiences.

4. Critical incident interviews (Klein 1999; Hemlin 2009) would involve asking students to describe in detail key episodes in their integration into their laboratories—and compare their stories with those of a key laboratory participant with whom they worked. These kinds of comparisons are hard to do post hoc, because episodic memory is reconstructive (Ericsson and Simon 1984); when asked to recall, human beings give plausible reconstructions that often put them in the best light (Neisser 1982). Problem behavior graphs and chronological flowcharts can act as a check on these reconstructions.
5. There is no current method for tracking development of a trading zone, micro or macro. One method would be to look at conferences co-attended by specific social scientists, humanists, scientists and engineers and see if any collaborative papers resulted (see Liberman and Wolf 2013 for more details on how to use this kind of method). But this kind of high-level indicator of possible trading zones would need to be complemented by detailed qualitative work on exchanges involving knowledge, affect, time and resources and how linguistic and other cultural barriers between social scientists, scientists and engineers were surmounted by adapting all three of the methods described above. In this connection, the collection and analysis of metaphors and analogies used in the process of gaining interactional expertise could be relevant in association with critical incident interviews, especially when set in the context of a chronological and/or problem-behavior graph. Mahootian has suggested that metaphoric redescriptions of laboratory engagement processes could be helpful in better understanding and facilitating the process. Specifically he suggests (Mahootian, forthcoming 2012) that (a) specific metaphoric images (like that of the mountain in Groves's example, and the honeycomb/row houses in Calleja-López's) have the potential

to broaden the scope of alternative perspectives during research and outreach; (b) the lab itself is a non-equilibrium system with various cycles and dynamic regimes that can be tracked in several ways including their temporality, and (c) the embedded observer-participant becomes a boundary object the moment they enter a research lab. Shifting interpretations of the status, purpose and activities of the boundary object that *is* the embedded observer-participant can be tracked as s/he negotiate his/her way through the system. Something as simple as tracking shifts in the duration and frequency of contact between the observer-participant and the lab researcher would be very revealing when compared to shifts in interactional expertise, and even shifts in laboratory procedures and routines (as seen in Conley's case). The way to begin this tracking is to adapt Schuurber's chrononological flowchart, putting documents, interviews and diaries in sequence as they occurred and experimenting with ways of organizing the flowchart to suggest patterns. Here the flowchart is a heuristic, not an effort to reduce the laboratory engagement to a linear experience. The end result may be multiple diagrams of a laboratory experience from different perspectives.

6. Longer-term follow up to see whether these integration experience have an impact on the laboratory even after the embedded humanist/social scientist has left. Ideally, these integration experiences will lead to long-term partnerships.

References

Baird, D., & Cohen, M. (1999). Why trade? *Perspectives on Science, 7*(2), 231–254.

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Collins, H. (2004). *Gravity's shadow: The search for gravitational waves*. Chicago: University of Chicago Press.

Collins, H. M. (2010). *Tacit and explicit knowledge*. Chicago: University of Chicago Press.

Collins, H., & Evans, R. (2007). *Rethinking expertise*. Chicago: University of Chicago Press.

Collins, H., Evans, R., Ribeiro, R., & Hall, M. (2006). Experiments with interactional expertise. *Studies in History and Philosophy of Science, 37*, 656–674.

Conley, S. (2011). Engagement agents in the making: On the front lines of socio-technical integration. Commentary on: "Constructing productive engagement: Pre-engagement tools for emerging technologies." *Science and Engineering Ethics, 17*(4), 715–721.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Fisher, E., & Mahajan, R. (2010). Embedding the humanities in engineering: Art, dialogue and a laboratory. In M. E. Gorman (Ed.), *Trading zones and interactional expertise: Creating new kinds of collaboration* (pp. 209–230). Cambridge, MA: MIT Press.

Frodeman, R. L., & Parker, J. (2009). Intellectual merit and broader impact: The National Science Foundation's broader impacts criterion and the question of peer review. *Social Epistemology, 23*(3–4), 337–345.

Galison, P. (1997). *Image & logic: A material culture of microphysics*. Chicago: The University of Chicago Press.

Gooding, D. (1990). Mapping experiment as a learning process: How the first electromagnetic motor was invented. *Science, Technology and Human Values, 15*(2), 165–201.

Gorman, M. E. (1997). Mind in the world: Cognition and practice in the invention of the telephone. *Social Studies of Science, 27*(4), 583–624.

Gorman, M. E. (2002). Types of knowledge and their roles in technology transfer. *The Journal of Technology Transfer, 27*(3), 219–231.

Gorman, M. E. (Ed.). (2010). *Trading zones and interactional expertise: Creating new kinds of collaboration*. Cambridge, MA: MIT Press.

Gorman, M. E., Groves, J. F., & Catalano, R. K. (2004a). Societal dimensions of nanotechnology. *IEEE Technology and Society Magazine, 29*(4), 55–64.

Gorman, M. E., Groves, J. F., & Shrager, J. (2004b). Societal dimensions of nanotechnology as a trading zone: Results from a pilot project. In D. Baird, A. Nordmann, & J. Schummer (Eds.), *Discovering the nanoscale* (pp. 63–73). Amsterdam: Ios Press.

Gorman, M., Werhane, P., & Swami, N. (2009). Moral imagination, trading zones, and the role of the ethicist in nanotechnology. *NanoEthics, 3*(3), 185–195. doi:10.1007/s11569-009-0069-8.

Hemlin, S. (2009). Creative knowledge environments. An interview study with group members and group leaders of university and industry R&D groups in biotechnology. *Creativity and Innovation Management, 18*(4), 278–285.

Klein, G. (1999). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Liberman, S., & Wolf, K. B. (2013). Scientific communication and the process to co-authorship. In G. Feist & M. Gorman (Eds.), *Handbook of the psychology of science* (pp. 123–150). New York: Springer.

Mahootian, F. (2012). *Innovation by disequilibrium*. Paper presented at the fourth annual conference of the Society for the Study of Nanoscience nd Emerging Technologies [S.NET], Oct 22–25. University of Twente, Enschede, The Netherlands.

McCray, P. (2004). *Giant telescopes*. Cambridge, MA: Harvard University Press.

Neisser, U. (1982). *Memory observed*. San Francisco: W.H. Freeman.

NNI Strategic Plan. (2011, February). http://www.whitehouse.gov/sites/default/files/microsites/ostp/nni_strategic_plan_2011.pdf

Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. New York: Addison Wesley.

PUBLIC LAW 108–153—DEC. 3, 2003 http://www.whitehouse.gov/files/documents/ostp/Issues/Nano%20Act%202003.pdf

Rabinow, P. (1996). *Making PCR: A story of biotechnology*. Chicago/London: The University of Chicago Press.

Ribeiro, R. (2007). The role of interactional expertise in interpreting: The case of technology transfer in the steel industry. *Studies in History and Philosophy of Science, 38*, 713–721.

Shrager, J. (2005). Diary of an insane cell mechanic. In M. E. Gorman, R. D. Tweney, D. C. Gooding, & A. Kincannon (Eds.), *Scientific and technological thinking* (pp. 119–136). Mahwah: Lawrence Erlbaum Associates.

Te Kulve, H., & Rip, A. (2011). Constructing productive engagement: Pre-engagement tools for emerging technologies. *Science and Engineering Ethics*, *17*(4), 699–714.

Wade, N. (2010, December 9). Anthropology a science? Statement deepens a rift. *New York Times*, A–16.

Chapter 9

Collaboration as a Research Method? Navigating Social Scientific Involvement in Synthetic Biology

Jane Calvert

Abstract It is now common for social scientists to become involved in emerging technologies in their early stages. This is an exciting development that opens up new opportunities for interdisciplinary collaboration, but it also puts social scientists in novel situations that give rise to methodological quandaries. In this chapter, I explore these tensions and dynamics, giving three examples of the various ways in which I have become involved in synthetic biology. I discuss my methodological struggles, and the fact that the new collaborative arrangements and peer-type relationships that I am engaged in do not easily lend themselves to straightforward distinctions between social scientific researcher and scientific informant. I explore resonances with notions of complicity and paraethnography that are currently being discussed in the anthropological literature, particularly the idea that we should think of collaboration itself as a research method. I end by arguing that methodological reflection can be valuable in helping us navigate these new kinds of interdisciplinary interactions between scientists, engineers and social scientists.

9.1 Introduction

Although I have been studying the social dimensions of the life sciences since 2002, my recent involvements in synthetic biology have given me new opportunities to collaborate directly with scientists and engineers. I have found this experience interesting and challenging. In my previous work on fields such as genomics and systems biology I have not been entangled with the science to the same extent, so I have

---

J. Calvert (✉)
Science, Technology and Innovation Studies, University of Edinburgh,
Old Surgeons' Hall, High School Yards, Edinburgh EH1 1LZ, UK
e-mail: Jane.Calvert@ed.ac.uk

N. Doorn et al. (eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*,
Philosophy of Engineering and Technology 16, DOI 10.1007/978-94-007-7844-3_9,

not had to make sense of the multifaceted nature of these collaborations. There is something about my involvement in synthetic biology that is distinctive, that it is my aim to capture, express and analyse here. To do this I draw on literature from Science and Technology Studies (STS), sociology and anthropology.

There are several different ways of interpreting these new interdisciplinary collaborative arrangements. There is an analysis at the socio-political level, which examines the increasing amount of social scientific involvement in emerging scientific fields and asks why it has occurred, what motivates it, and what it is expected to achieve. This analysis has been carried out extremely well by other commentators, and I will briefly discuss it here. But another way of attempting to make sense of these experiences is through the lens of methodology, and this is the approach that I prioritise. Of course these two approaches cannot be clearly separated. The institutional conditions for science studies have changed. Social scientists are becoming involved in scientific programmes and networks for socio-political reasons, which means that the pressures of doing fieldwork are different, that it is no longer easy to merely be a 'fly-on-the-wall', and that we have to start thinking afresh about our methodologies.

Since I am a social scientist, this is the viewpoint that I adopt here, although it is important to recognise that other groups such as lawyers, bioethicists and policy makers have also become closely involved in synthetic biology from its early stages. In this paper my interest is in what might be understood as particularly social scientific concerns, such as methodology, particularly the role of social scientists in relation to their research subjects, and the reflexivity that this engenders. These concerns are not necessarily shared by other more normatively-oriented disciplines. Fields such as bioethics and law may also differ in terms of the objectives guiding their interactions with synthetic biology. I started out assuming that I could study the field as a detached observer, but quickly realised that this was not possible because I had become involved to a greater extent than I had anticipated. Being 'swept up' by the field in this way meant that my aims had to be reformulated. One of the results of this was that my methods became a topic of study. The collaborations that I was involved in were continually challenging the way I was thinking about my ongoing research. This led to my interest in collaboration as a form of knowledge-production.

I start this paper at the socio-political level by briefly analysing what Guston (2006) has called the 'Elsification' of the social sciences that followed in the wake of the ELSI/ELSA programmes, and which has led to social scientists becoming engaged in areas such as nanotechnology, stem cell research and neuroscience. I then turn specifically to the new opportunities for interdisciplinary research that are emerging in synthetic biology, where social scientists are becoming a required component of synthetic biology research programmes in Europe, the US and beyond. This mandatory involvement has provided new opportunities for collaboration in research and in teaching. I argue that there are features of synthetic biology that make it particularly amenable to these collaborative arrangements. I give examples of three of my own collaborative experiences, stretching over the last 4 years. The first involved being part of a synthetic biology 'sandpit', the second draws on my contribution to educational and teaching initiatives, and the third reflects on my unexpected role a synthetic biology conference. I draw out the recurring themes that emerge from these collaborative experiences, many of which can be understood in methodological terms. Since my own role in the field of synthetic biology is complex, it is perhaps not surprising that the distinctions between me and my ‘informants’ has become blurred. Furthermore, my involvement is not only in the laboratory, but it takes place in various different settings such as scientific conferences and teaching committees. I draw on work on multi-sited ethnography, complicity, paraethnography and collaboration as a research method from the recent anthropological literature. I end by arguing for the value of exploring the methodological dimensions of these novel interdisciplinary collaborations.

9.2 Elsification

Collaborations between natural scientists and social scientists are not new. Work under the heading of ‘Ethical Legal and Social Issues’ (ELSI) was initially funded as a strand of the Human Genome Project, taking 3–5 % of the total funding (Kevles and Hood 1992). Since the controversy over genetically modified foods, it has become common for policy makers in Europe to involve social scientists in the development of scientific research as a way of avoiding ‘another GM’ (Rip 2006).

In recent years ‘ELSI’ programmes have become associated with many new scientific fields, such as nanotechnology (Macnaghten et al. 2005), stem cell research (Robertson 2001), and neuroscience (Illes and Bird 2006). This means there is an increasingly distributed network influencing the governance of science, which includes social scientists, policy makers, lawyers, bioethicists and publics. Several commentators have termed the involvement of social scientists in these new areas ‘Elsification’. This can be defined quite uncontroversially as the idea that “every major research project should be accompanied by research on its ethical, legal and social implications” (Paul and Van den Belt 2006, p.12). But Elsification can be interpreted in a more negative manner. Guston (2006), for example, says there is a danger that “examining the implications of the natural sciences may be taken as the only, or primary, role for the social sciences” (p.306). In this way ELSI can placate or divert other kinds of research, instead of informing the research process itself. And Rip (2005) has warned that STS can become the victim of Elsification, losing critical distance in the process. Furthermore, Fisher (2005) has critically analysed the US ELSI programme, showing that it had a limited capacity to influence policy.

So are we seeing the ‘Elsification’ of synthetic biology? There is certainly a conviction that synthetic biology raises important ethical, legal and social issues, as is demonstrated by the large number of reports written on the field – 39 since 2004, according to a recent count (Zhang et al. 2011). Webster (2007) argues that in emerging technologies, such as synthetic biology, new relations between science, technology and society are being created, and that these provide new spaces for intervention. In other words, there is something about emerging technologies which makes them more open to social and ethical reflection. This may be because of the promissory nature of these technologies, which are expected to have impacts on and benefits for broader society (Selin 2007). This potential gives force to arguments that attention needs to be paid to the social, political and ethical dimensions of these fields (McGregor and Wetmore 2009). It is perhaps for these reasons that social scientists are becoming a required component of synthetic biology research programmes in Europe, the US and beyond (Calvert and Martin 2009).

9.3 Collaborative Experiences

My own involvement in synthetic biology was the consequence of being approached by synthetic biologists who needed social scientists to be named on their research proposal to fulfil the requirements of the funding council. This was the first time that I had been approached by scientists in this way. The research proposal was funded and it resulted in a synthetic biology network called the Synbiostandards Network,[1] which is one of seven interdisciplinary networks funded by four research councils in the UK, all of which require an 'ELSI' component. The obligatory involvement of non-scientists in these networks has produced a new community of social scientists with an interest in synthetic biology in the UK.

There are several features of synthetic biology that may make it particularly open to these kinds of collaboration. First, synthetic biology is already an interdisciplinary field, bringing together engineers, biologists, chemists, and computer scientists, meaning that bringing in social scientists is perhaps not such a stretch as it would be in a more conventional discipline. Second, particularly in engineering-oriented approaches to synthetic biology, there is an explicit attempt to democratise the technology, and to open it up to those who would not normally engage in cutting-edge life sciences, so in this sense it is a field that is looking outward toward new groups (Dyson 2007; Smolke 2009). And finally, synthetic biologists explicitly see themselves as building a community as well as a technology. In this sense, they are 'heterogeneous engineers' (Law and Callon 1988); they are involved in the manipulation of social, political and economic factors, as well as technical ones. This confluence of factors has provided new opportunities for collaboration with synthetic biologists, in both teaching and research. In what follows, I give three examples of my involvements in synthetic biology as a social scientist. At the end of this section, I draw out the recurring themes that emerge from these various encounters.

9.3.1 Early Engagements: The Synthetic Biology Sandpit

The synthetic biology 'Sandpit' (full title: 'IDEAS Factory Sandpit on New Directions in Synthetic Biology') was a week-long residential event which took place just outside Washington DC in March 2009. It was jointly organised by the

---

[1] See http://www.synbiostandards.co.uk/

US's National Science Foundation (NSF) and the UK's Engineering and Physical Sciences Research Council (EPSRC). Sandpits are a common funding mechanism for the EPSRC, and this research council has organised over 30 Sandpits to date in the UK. But this was the first time that a Sandpit had been organised with the NSF, and the first time it took place in the US. This Sandpit had 30 participants, approximately half from the US and half from the UK. Participants were selected from a pool of around 200 applicants, with a selection committee that included an occupational psychologist whose job was to assess the suitability of applicants for this intense residential event.[2] Two social scientists were selected to participate in the sandpit (myself and a political scientist); the other participants were all scientists and engineers.

The aim of Sandpits is for the participants to develop innovative and interdisciplinary research proposals together over the course of the week. These proposals are subjected to 'real time peer review' by all the other members of the Sandpit. Successful projects are announced on the final day, and funding is committed to them at this point. The synthetic biology Sandpit had approximately 10 million dollars to distribute to successful projects, meaning that the 'carrot' was rather large, and injecting the event with tension and expectation.

As participants, we spent the first half of the week doing various kinds of activities to get to know each other and to stimulate our creativity, and the second half developing research proposals. I kept a diary during the sandpit, which I draw on here.

Sunday: The UK participants arrived the day before the official sandpit started. We shared a certain amount of trepidation about the kind of activities we were going to have to do, because EPSRC Sandpits have a reputation of making use of unusual techniques, such as cookery and ukulele playing, to help participants develop new ideas.

Monday: On the first official day of the sandpit we were given group exercises to do and games to play. We were each presented with a playing card as we walked through the door, and these cards were used to group and regroup us in various different ways. For example, at one point we had to get into groups that would produce a winning poker hand. We also had to write 'Wanted' and 'For Sale' advertisements on post-it notes. As was the case on many occasions during the sandpit, these exercises proved harder for me than for some of the scientists and engineers, who could quite easily put down their scientific skills under 'Wanted' or 'For Sale' headings (such as microfluidics, functional genomics etc.). After some deliberation, my 'Wanted' was: "Some people to study. To be part of it. Enlightenment". My 'For Sale' was "Connections to the ELSI community. Putting synthetic biology in broader social, political and economic context". These were not taken up, but nor were the 'For Sale' offers of most of the scientists and engineers.

After the games and exercises we were given a series of lectures from the EPSRC and NSF organisers about what we were expected to do over the course of the week. We were told that the aim of the Sandpit was to "build a world class synthetic

---

[2] Personal communication, selection committee member.

biology community”. We were then asked to think about developing ideas during the week that “would make your mum proud and your colleagues jealous”. We were told that to be funded, a Sandpit project had to be “interactional, multidisciplinary, transformative, novel, innovative, high-risk potential, high-return”. It should also have a “wow factor”, which was described as “something of real excitement”. The aim of the sandpit mechanism, we were told, was to break away from the conservativism of peer review. This meant that we should aim to develop projects that were outside our normal area of expertise.

Tuesday: On the second day there were more games, including a collaborative painting exercise with oil paints, palette knives and canvases, and more lectures, including ELSI lectures which addressed ethics and regulation and warned the scientists that “only the paranoid survive” and that synthetic biology was likely to be a target for critical NGOs. At one point during this day I wrote in my fieldnotes “I am on my own here”.

Wednesday: On the third day we had to start developing our research projects. The questions that were meant to get us thinking about projects were: “What types of problems are you hoping that synthetic biology might be able to solve?” and “What do you think is the biggest barrier to synthetic biology?”. I found these questions problematic because they focused everyone’s attention on problems ‘to be solved’ and barriers ‘to be overcome’, which encourages the idea of the ‘technical fix’ and frames the work in a narrow way. This is a recurring feature of synthetic biology events, as will be discussed below.

Not surprisingly, one of the ‘barriers’ that many people identified was ‘public acceptance’. This led to a team, composed only of scientists and engineers, developing a project about public acceptance, with the aim to “embed the right kind of positive attitude in society” and to induce “subtle changes of perception” about synthetic biology. This was a project that from the start clearly violated many tenets of social scientific work on public engagement over the last 20 years (see Marris and Rose 2010 for an overview).

The other projects that began developing were scientific projects, and it was very hard for me, as a social scientist, to work out how to become part of these projects. I wandered around from group to group, looking unsuccessfully for a ‘home’. But on the afternoon of the third day an idea emerged from one of the brainstorming exercises which resulted in a post-it note with the words ‘Synthetic Aesthetics’ written on it. This title appealed to me and two of the engineers, and we left the venue to walk around the extensive gardens and brainstorm about a project on the aesthetic dimensions of synthetic biology. We decided to take seriously the organisers’ encouragement to “think outside the box”, by doing a dance to present our ideas, instead of giving a normal powerpoint presentation. The dance was rewarded with applause and laughter.

Thursday: The next day we developed our project ideas further. We decided that we would use the project to bring scientists and engineers together with artists and designers in collaborative exchanges. At this stage, the people developing public acceptance project suggested that we were doing the basically the same thing as them, and that we should join forces. The three of us developing the Synthetic

Aesthetics project argued that we were not doing the same thing, that our project was not about public engagement, but it was about exploring the intersection between art and design and synthetic biology.

The project developed further, and we had discussions about the sublime. In response to real time peer review comments, we had to try to define beauty and defend frivolity. These activities were all totally unexpected and led me into discussions I had never anticipated when starting the Sandpit. We had to refine our project ideas and keep presenting and re-presenting them under considerable time-pressure, taking the peer review comments into account. On this day one of my diary entries reads: "Very busy this day, little time to eat".

Friday: On the final day the atmosphere was tense, because the participants realised that there was only enough funding for half the projects. It became clear that we were all competing with each other over the same pot of money. The Synthetic Aesthetics project, however, was an oddball, and it had the smallest budget. Our willingness to do a dance in the early stages of the week led to goodwill towards our project, and we found out at the end of the final day that we were successful in getting funded.

What resulted from this strange and intense experience was a project that for me was completely unexpected. It has also left me in the rather unusual situation of working in a peer-type collaborative relationship with two leading synthetic biologists. None of us has prior expertise in the topic area, so none of us has epistemic authority over the direction of the project. There are two Principle Investigators on the project, one for each country, and I am the UK PI, meaning that I am not working 'for' the engineers on this project, which is the more normal situation for social scientific engagements with scientists and engineers. I will return to these issues below.

9.3.2 Ongoing Experiences: The iGEM Competition

Teaching is an area of collaboration that is often overlooked. In this second example of my involvement in synthetic biology I will focus on a rather unusual set of teaching opportunities offered by the international genetically engineered machines competition (iGEM). This is an undergraduate competition where teams from across the world compete to build the best 'genetically engineered machine'. Edinburgh University has had a team since 2006 and with an STS colleague I have been involved in advising the team since 2008. In the first year we were invited to hold only one discussion session with the students, but in following years we have been involved to an increasing extent. By the summer of 2011 we were meeting the team once a week to discuss their ongoing work.

The competition started as internal event to MIT in 2003. It has grown exponentially since. In 2010 there were over 100 teams with approximately 1,000 students taking part. The competition plays a very important role in synthetic biology, and reflects the desire to democratise the field, and make biology easier to engineer (Smolke 2009). What is particularly interesting about this competition is that as well as training students in technical skills, iGEM also aims to instil certain ideas about safety, security, and open access to the technology, so a community is being built as well as a technology. However, as was the case in the Sandpit, the orientation is heavily around: "what problems can synthetic biology solve?" rather than: "what is the best way of solving this problem?".

iGEM teams are encouraged to do some of their work in the area of 'Human Practices', a 'post-ELSI' term coined by Paul Rabinow in a conscious attempt to avoid some of the negative connotations of ELSI discussed above. In the context of the iGEM competition, Human Practices has been taken to apply broadly to any of the non-technical aspects of the students' work. The teams undertake a range of different types of Human Practices projects, often designing methodologically problematic internet surveys, or running deficit-model style outreach events. But some students develop imaginative and interesting projects, such as in-depth studies of 'Do-It-Yourself' biology or national regulatory frameworks, or 'futures' workshops with designers. Of the teams I have worked with in Edinburgh over the last 3 years, their initial Human Practices interest has always been in public attitudes towards genetic modification (which may reflect a particular European concern).

There are some unusual features of this form of pedagogical involvement. It involves undergraduates, so they are exposed to the 'social' at a very early stage of their scientific training, when they have not fully embraced the identity of being a 'biologist' or 'engineer' themselves, so are perhaps more open to new disciplinary perspectives. Additionally, iGEM is a lab-based project, so it does not provide a formal teaching opportunity like a structured course. But this is not necessarily a disadvantage, it means that discussions with students are focused on the social, political and ethical dimensions of their scientific project, rather than abstract analysis of the 'issues' arising from synthetic biology. I do not have space here to explore the many interesting features of this competition (for an extended discussion see Frow and Calvert 2013), but it is nevertheless important to bear in mind that iGEM is very influential in the field, because the students taking part will be the synthetic biologists of the future.

I attended the iGEM competition at MIT in 2009 as an advisor and in 2010 as a judge. In my 2009 fieldnotes I noted how there are few scientific conferences where the audience of 1,000 is on the edge of their seats, and where an announcement from the podium results in cheering, rapturous applause and embraces. I was struck by the Mexican waves in the huge auditorium, the techno music that was played between the sessions, and the rather wild Sunday night party. As a judge in 2010 I had an official role, because I was there to assess the Human Practices component of the student projects. There were around 10 Human Practices judges and approximately 70 judges in total, some specialised in technical areas (such as food and energy), and some given a particular output to assess, such as the posters. After all the presentations had been given, specialised judging meetings were held and then all the judges convened for a 4 h long meeting, which ended close to midnight on the penultimate day. During the final, which took place the next day, I was told I had to vote on the technical aspects of the finalists' projects even though I felt unqualified to do so. I also felt uncomfortable standing on the stage for the extended prize-giving ceremony, with my official role in this technical competition made clearly public.

As a follow-on to this teaching activity, my STS colleague and I were invited by the synthetic biologists at Edinburgh to propose some postgraduate teaching on the (undefined) 'ethics' of synthetic biology. We developed a proposal for a 10 week course around a broad range of topics on issues that in our experience concern synthetic biologists. Our course proposal went down surprisingly well with the scientists and engineers, who decided to make it a mandatory component of their Masters programme in systems and synthetic biology. As a result of these pedagogical developments, I have come to share many of the same concerns, questions and students with my colleagues in science and engineering.[3]

9.3.3 Fresh from the Field: SB5.0

My final example of a collaborative experience is the Synthetic Biology 5.0 conference which took place in Stanford, California in June 2011. An academic conference is perhaps not a classic fieldwork site, but synthetic biology conferences are where I do a great deal of my fieldwork.

Synthetic Biology 5.0 was the fifth in a series of conferences, the first of which was held at MIT in 2004 with 290 delegates. SB5.0 had approximately 700 delegates, making it the largest synthetic biology conference to date. Of the delegates, approximately 20 were social scientists and ethicists, broadly described (with perhaps another 20 from policy, government and industry), there were also approximately five artists and designers. This means there was a considerable sub-community of 'non-scientists' present. As members of this sub-community we found similar things interesting, and in the breaks we compared notes on the metaphors used, the commercial sponsors listed on the slides, and the issues that were *not* discussed.

I do not want to imply that this group of non-scientists only talked to each other, however, because all of us have built up collaborative relationships, often involving friendship, with the scientists in the field. My scientific colleagues would often come up to me after a talk saying something like "what did you think about that speaker's point about the different cultures of chemists and biologists?". And I was pleasantly surprised when an engineer that I have known for about 4 years explained to a college that I was "post-ELSI" (this is a recent and hard-fought shift in terminology).

One thing that is particularly interesting about synthetic biology conferences is that they engage explicitly with social, philosophical and ethical issues. For example, a Twitter feed from the final day of the conference includes comments about ontology (in the philosophical sense), open source, and design principles. The 'social' strand of synthetic biology was also clear in a paragraph on the cover page of the conference programme which says: "Our mission is to ensure that the engineering

---

[3] Of course social scientific involvement in science education stretches back to the 1970s at least, when 'contextual studies' were introduced into science education (see Edge 1995).

of biology is conducted in an open and ethical manner to benefit all people and the planet. We envision synthetic biology as a force for good in the world".[4] This is a strong statement for the front page of a scientific programme.

The incorporation of the social into the programme was also notable in the previous Synthetic Biology 4.0 conference which took place in Hong Kong in 2008, and which I also attended (thanks to a grant from the Synthetic Biology Network mentioned above). In Hong Kong there were sessions on global social impact, biosecurity, future scenarios, intellectual property and commercialization in the programme. One of the sessions was organised by an NGO (the ETC group), the NGO that is the most vocal in its opposition to synthetic biology.[5] For a session of this type to be organised at a scientific conference is itself worthy of note. In his closing comments to the conference, the conference organiser Drew Endy acknowledged that "We have new colleagues from social sciences, civil society organisations and industry".

But rather than building on these developments, the integration of the social was more problematic in SB5.0, which took place nearly 3 years later. This was partially because there were far fewer parallel sessions than there had been in Hong Kong, with the majority of the conference taking place in plenary. Although by March 2011 most of the invited speakers and sessions were detailed online, 1 week before the conference in June 2011 there was still one session in the programme for the second day that was completely lacking in speakers and titles. This was a session called 'Interacting with Society'. In the final printed version of the conference programme this reads "details to be announced".[6] But less than a week before the conference emails were sent to six social scientists (including me), an artist, an employee of the United Nations, someone from industry and a representative of a public/private research institution, all of whom were already registered to attend the conference at our own expense, asking us if we would be on a panel designed to "motivate and energize the scientific community to interact more about and with society".[7] This was incredibly short notice compared to the amount of time that was spent organizing the science sessions. Because there were ten people on the panel we only had a very limited amount of time each and the framing of the session was extremely problematic. Most pointedly, 'Interacting with society' implies that society is something 'out there' to be interacted with.

Were we being snubbed? Should we have been annoyed? That is what some of the people on the panel concluded. But there were other organizational and personnel issues going on behind the scenes which mean I do not think we should necessarily draw this conclusion. It could be argued that the fact that there *was* such a panel in a plenary session in a high-profile in international scientific conference is itself

---

[4] See http://sb5.biobricks.org/files/sb5-program-book-v3.pdf

[5] The conference programme is available at http://sb4.biobricks.org/agenda/sb4_agenda.pdf

[6] The printed version of the conference programme is available at http://sb5.biobricks.org/files/sb5-program-book-v3.pdf

[7] http://reconstructing-sciences.net/content/event-1-sb-50

positive. This incident is a good demonstration of the problems I often have in my engagements with synthetic biology, particularly: how do you study a field when you are to some extent involved in legitimising it?

One of the most challenging issues that arose for me when I found out I was going to be on the panel was that I had to decide what I wanted to say to this community that I have been studying for 4 years. I decided that there were three points I wanted to get across, grounded in the literature in science studies, but supported by my own experiences.

The first is that the technical is political, in other words, that epistemological and normative issues can't be clearly separated, that "every technological choice is potentially an ethical and political act" (Fisher 2005), meaning that it is illegitimate to separate the science from its social implications. I wanted to make the point that scientific and technological developments are the result of choices, such as funding decisions, and that these choices are based on values (Johnson and Wetmore 2007). The choices about what project a scientist chooses to do, what organism they choose to work on, what applications they choose to develop, what visions of the future they project, are all social and political choices. In other words, science is permeated with society already, and we can't separate knowledge-creation from its social, political, economic and historical context.

The second point was that 'things could be otherwise', that scientific and technological developments are not inevitable, but the result of "tacit understandings, choices, and conceptual frameworks that are later naturalized as part of the scientific process" (Doubleday and Viseu 2009, p.73). Once we become aware of the contingency of these choices, these choices can be changed.

The third was to argue for the 'opening up' of synthetic biology. This refers to the aspiration to broaden the range of voices that can contribute to the development of a technology, and to bring a wider range of actors into a field than would normally be involved (Schot and Rip 1997; Stirling 2005; Schuurbiers 2008). I wanted to make the positive point that synthetic biology is a good forum for 'opening up', because there are already philosophers, lawyers, bioethicists, sociologists, political scientists, anthropologists, artists and designers working in the field.

I hoped to get all these points across, but I did not achieve this in the very short amount of time I had. All I managed to say was "science is part of society".

9.4 Recurring Themes

So what points should be drawn out from these diverse experiences? There are a set of recurring themes that cut across the three narratives.

The first is the relationship between the social and the technical. This is something that we see very clearly in synthetic biology, in the attempt to build a community as well as a technology in the iGEM competition, in the integration of Human Practices into this competition, and in the prominent discussion of synthetic biology being 'a force for good in the world' on the front page of the conference programme. But the relationship between the social and technical can be problematic, because there is a tendency to see the social as an obstacle. Precisely this issue is identified by Schot and Rip (1997) who talk about how engineers “think their business is to focus on the technical, they redefine the social as barriers to be overcome and approach it as what they must do to make the technical successful” (p.264). What also emerges from these examples is the assumption that the social and technical can be easily separated, as exhibited by the title of the ‘Interacting with Society’ session at the conference. It is interesting that Doubleday and Viseu (2009) also note, reflecting on their experiences as social scientists engaged in nanotechnology, that the social and ethical dimensions of nanotechnology are normally placed *outside* the practices of the production of scientific knowledge.

We also saw that even though the integration of social and technical issues was laudable at a synthetic biology conference in 2008, it was more problematic in the 2011 conference, and that the integration of the social is not smooth, but needs constant vigilance. This raises issues about what we should make of a field that gives a voice to the social, but where this voice is often inadequate and fleeting. There is also the question of how we should react to these encounters; whether we should be disappointed in them, or take the view a token attempt to involve social perspectives is better than no attempt.

As this last comment indicates, another feature of these narratives is the mixture of positive and negative experiences. During the Sandpit my negative experiences involved being ‘homeless’ in an alien environment and on the outside of the scientific proceedings. On the positive side I managed to develop a project with two engineers that was interesting and important for all of us, where we generated a shared understanding of what we were doing and an agreement that this was not ‘public acceptance’. Similarly, being an advisor to the iGEM team provided an excellent opportunity to engage with science and engineering undergraduates in the early stages of their careers, but I found some aspects of being a non-scientific judge particularly uncomfortable.

After I presented some of this empirical material at a workshop one of my colleagues noted the ‘schizophrenic’ awkward negotiation of multiple identity positions that marked my experiences in synthetic biology. I think this idea of a negotiating multiple identity positions is a useful analytical angle on social scientific involvement in scientific and technological fields. It resonates with comments by other social scientists such as Forsythe (1999) who notes how in fieldwork “the collapsed roles of participant, observer, critic, employee and colleague collide with one another” (p.22), and Fortun (2005), who argues that social scientists studying the life sciences may have to hold several different conflicting positions simultaneously.

I have written previously about ‘contributor’ and ‘collaborator’ as possible roles for social scientists to take when engaging with synthetic biology (Calvert and Martin 2009), but recently I have found that the distinction into these different roles too detached from the realities of my engagements. They do not reflect the messiness of social scientific involvement in synthetic biology, the debts, obligations, concerns, loyalties, contradictions, hopes and fears, and the dynamic and affective nature of it all.

One role that is appealing, however, is that of the trickster or the jester. The ‘trickster’ is found in many different cultures, sometimes represented by the fox, or by the coyote in American Indian mythology.[8] Turnbull (2000) gives social scientists this role, saying that “We who purport to be historians, sociologists, or cultural critics, are also tricksters” (p.92). The idea here is that the social scientist is someone who asks critical questions, who provides an alternative perspective, and to some extent disturbs engrained ways of thinking. Turnbull explains that “The trickster is the spirit of disorder, the enemy of boundaries” (p.92). In this sense the role of the trickster fits the social scientist well, particularly when we want to show that the boundaries between the scientific and the technical do not stand. It also ties in with the aspiration to ‘open up’ synthetic biology to a range of (potentially destabilizing) influences. Furthermore, the jester or trickster will often use humour, which was the main response to the dance at the Sandpit. But this is something that has to be done with care. One of the dangers of using humour is that “we end up adopting the role of the court jester to the technocratic elite” (Croissant 1999, p.23).[9]

Although some features of the jester/trickster do fit well with the narratives discussed above, the metaphor has its limits, because there is a sense of knowingness and wileyness attached to these characters which belies the fluidity and lack of control that is often the dominant experience when being a social scientist in these situations. This lack of control is often due to the fact that interactions and engagements with synthetic biologists take place at many different sites, which are not necessarily familiar research sites. This issue of multi-sited ethnography is a methodological one, and I have found the methodological literature very helpful in conceptualising my involvement in synthetic biology. I focus on methodological challenges for the remainder of this chapter.

9.5 Methodological Challenges

The most pressing methodological challenge that has emerged from my encounters with synthetic biology has been that the collegial, peer-type relationships I am engaged in do not easily lend themselves to straightforward distinctions between social scientific researcher and scientific ‘informant’. As an investigator on an EPSRC project, as an advisor to an undergraduate iGEM team, and as a speaker

---

[8] Turnbull (2000) also cites the monkey god in India, the spider in Africa and the Loki in Scandinavia.

[9] There are different understandings of trickster in the STS literature. Vikkelsø (2007) says that the trickster “constantly shifts sides from one stakeholder to the other and explores the way their concerns can be practically reconciled in a politically sensitized way” (p.297). Bijker (1993) thinks of the reflexive STS researcher as jester, but not in a positive manner, because for him being a jester involves “not committing themselves to dirtying their hands by making necessary decisions” (p.116). And Haraway (1991) famously talks about tricksters, but for her it is nature or the world who is a trickster, because it/she does not fit into our existing categories.

at a synthetic biology conference, there is no clear water between me and the people I study.

I have also arrived at a situation where my research and teaching relationships with scientists and engineers are not notably different from those with my colleagues from the social sciences. I have found myself in an interdisciplinary network together with the scientists and engineers, attending the same seminars and conferences, interested in similar literature and academic questions, and even in the same gossip. In other words, I have found myself accepted as part of the community. But being part of this community is not always easy. As I noted above, as a social scientist one often feels that one is on the periphery, or outside the main scientific activity, sometimes struggling to find a 'home'.

Being a co-investigator on a project with scientists and engineers gives rise to particular methodological challenges. As a co-investigator one is a formal participant in a project with obligations to deliver the outputs and to manage the process in a way that means that social scientific distance is not an option. This raises important questions about how one should negotiate this type of peer-type relationship with scientists and engineers who are also one's object of study.

The role of teacher raises similar issues. My role as an advisor to the Edinburgh iGEM teams has simultaneously provided opportunities for research, for exploring the integration of the social and political into technical projects as they develop, and for seeing how young scientists respond to novel interdisciplinary interactions. Croissant (1999) has examined the tensions involved when teaching becomes a research space, and when one's students become one's informants. Questions about whether this is an appropriate use of a pedagogical situation arise. The power relationships that we often find in STS when 'studying up' are reversed when the social scientist is the one judging and evaluating the student.

In cases like this, when the situation is no longer one of social scientific researcher and scientific informant, traditional methodological approaches, such as the interview, also become less straightforward. This ties into broader discussions in sociology, most notably Savage and Burrows' (2007) work on 'The coming crisis of empirical sociology'. Savage and Burrows argue that interviews no longer gives sociologists distinctive access to the social, since interviews are used by many non-sociologists, such as chat show hosts and magazine journalists. Rather than doing still more in-depth interviews, Savage and Burrows (2007) argue that sociologists should make use of the social data gathered by others. Synthetic biology is a field that is rife with this kind of social data in the form of videos on YouTube, where numerous interviews with most of the key players can be found online. This wealth of alternative sources of data suggests that the traditional methodological approach of interviewing one's informants is not the only option. Savage and Burrows argue for more methodological innovation and imagination.

There is also the issue that interviewing people that one is currently working with can change the dynamic of the relationship. Pulling out a voice recorder in the flow of a conversation will often interrupt and constrain the discussion. When ongoing conversations are the norm it can seem false to attempt to capture this in an official 'interview' session. As another social scientist who works in synthetic biology put it, “hanging out and having drinks” is a more normal form of involvement with the field, and perhaps one that is likely to result in data that could not be captured in the traditional interview format.

The informant/researcher situation is complicated still further when friendship comes into play, which is normal when one has ongoing interactions with other researchers on shared projects. Friendships require trust, and they may also involve emotional ties, personal obligations and fears of upsetting and contradicting. Fortun (2005) argues that friendship is something to be embraced, rather than avoided, in interactions with scientists and engineers.

Some may see the similarities between my concerns about my involvements in synthetic biology and the familiar anthropological idea of ‘going native’, which is often the norm in fieldwork situations. This rather anachronous notion can be nuanced by thinking instead in terms of ‘rapport’, which Rabinow (2002) says is necessary to engage in fieldwork at all. Marcus (1997) argues for a more controversial relationship of ‘complicity’ with one’s informants. He draws on the definition of ‘complicity’ from the OED which is defined both as a “state of being complex or involved” (which I think summaries my engagements with synthetic biology well), but also, and more problematically, as “being an accomplice; partnership in an evil action” (Marcus 1997, p.85). The idea of complicity carries these more negative resonances, and draws attention to the possibility of being compromised in research situations, which reminds us of the negative interpretation of the term ‘Elsification’. I am compromised to the extent that I know my relationships with scientists and engineers have to be sustained in order for me to do my research. But perhaps more problematically, as I pointed out above, my willingness to be on an unsatisfactory panel at the SB5.0 conference shows that I am involved in legitimising the field. These are all further examples of the negotiation of multiple identity positions which results from social scientific involvement in fields such as synthetic biology. These situations are often uncomfortable, but this may be something it is necessary to accept. Being comfortable would suggest a lack of awareness of the compromises being made. Perhaps social scientists in these situations should embrace what might be called an ‘ethics of discomfort’.[10]

9.5.1 From Informant to Paraethnographer

I have argued that the idea of the ‘informant’ has become extremely problematic in my encounters with the synthetic biology community, to the extent that it is no longer feasible for me to adopt the role of the detached observer. But luckily there is a wealth of recent anthropological literature on this topic. In anthropology there has been a shift from thinking about data collection to thinking about collaborative

---

[10] This was a phrase used by Gaymon Bennett at the ESRC Seminar Series on Synthetic Biology and the Social Sciences, University of Edinburgh, 14–15th February 2011.

knowledge-making. Instead of talking about 'informant' anthropologists have started using terms like 'discussant' (Buergi 2010), 'interlocutor', and 'epistemic partner' (Holmes and Marcus 2008). Rather than describing themselves as doing interviews with key informants, it is now more common to talk about "co-puzzling with interlocutors" (Fischer 2011).

Epistemic partners can also be described as 'paraethnographers'. Paraethnographers have "a preexisting ethnographic consciousness or curiosity" (Holmes and Marcus 2008, p.82) that is epistemologically equivalent to the anthropologist's (Marcus 2005). In other words, our subjects are very good at doing their own ethnography. If our subjects are doing their own ethnography this clearly has implications for interdisciplinary collaboration. Holmes and Marcus (2008) argue that "the figure of the para-ethnographer changes fundamentally the rules of the game for collaboration" (p.86). And it is the case that many of the synthetic biologists I interact with are excellent paraethnographers. They are aware they are trying to build a new field and construct a new community, and regularly reflect on these efforts and their broader social, economic and political and historical context.

The kinds of entanglements with synthetic biologist-paraethnographers that I have described above are often difficult to negotiate, but Holmes and Marcus (2008) argue that such entanglements can open up spaces for building reflexivity into institutions. And I think there is evidence that we are starting to see this reflexivity emerge in synthetic biology. For example, in the iGEM competition reflexivity is built into synthetic biology through the mechanism of Human Practices, and the involvement of social scientists in synthetic biology projects, networks and conferences provides opportunities for reflexivity in many different contexts, even if this involvement is problematic and uneven.

Another characteristic of paraethnographers is that they are involved in multiple sites such as meetings, policy groups, grant giving panels, and public engagement events (Marcus 1997). These observations suggest that engagements with multi-sited paraethnographers should not be restricted to the laboratory, and they resonate with my experiences in engaging with synthetic biologists in various different settings such as scientific conferences and teaching committees. This raises questions about the place of the laboratory as a site of research collaboration, and suggests that if we are going to get the most out of our paraethnographers we should move beyond it.

9.5.2 Collaboration as a Research Method

The final, and most important, methodological point I want to make is the idea that we should start thinking of collaboration itself as a research method, and as a form of knowledge production. If we are indeed witnessing the emergence of a new way of doing fieldwork, which is characterised by a recognition of the reflexivity of the scientists we work with, who are best understood as our 'epistemic partners', then we should start to think in terms of producing new knowledge together. Fortun (2005) makes a similar point when he argues for "a different relationship among scientists and those of us who study science, a relationship that might produce new knowledge" (p.160).

So how do we go about producing new knowledge together? What if we have different objectives? One of the main objectives of synthetic biologists is to promote the field, whereas social scientists may be more interested in studying its development. But perhaps we do not have to have the same overarching objectives when we collaborate; instead we can have the same lower-scale, more pragmatic objectives, such as getting a grant or running a Masters programme.[11] In pursuing these pragmatic objectives together we will inevitably provoke each other to think in new ways. This point is made by Marcus (2008), who explains that "The basic trope of fieldwork encounter shifts from, say, apprentice, or basic learner of culture in community life, to working with subjects of various situations in *mutually interested concerns and projects*" (p.7, emphasis added). If we are training students, managing a grant, or putting in a research proposal together, we are involved in 'mutually interested concerns and projects'.

The idea of collaboration as a research method may sound ambitious, but based on my experience I think that it is best to think of it in terms of small steps, and ongoing 'soft' interventions, which are likely to exhibit themselves in the form of regular and unremarkable interactions and engagements. One of the features of these engagements may be that it is necessary for scholars of science and technology to "give up the claim to be the ultimate explainers of science", because by doing this "we can engage in inter-disciplinary conversation about what science is, what it does, and what it should do" (Hamlin 1992, p.534).

I have only roughly sketched out the idea of collaboration as a research method here. It requires further clarification and investigation, but I think it is a useful idea to draw on when trying to make sense of social scientists' involvement in scientific fields. I will give the last word on the topic to Fortun (2005), who says that it is his aim is to:

> construct new assemblages in which experimental practitioners from both the sciences and science studies can 'muddle through' together toward mutual understanding and even practical ends—uneasily, to be sure, but abetted by the same combination of laughter, dedication, forbearance born of sustained proximity, and mutual critique that characterizes the best friendships in the personal domain (p.170).

9.6 Conclusion: A New Way of Producing Knowledge?

In this chapter I have focused on the methodological issues that arise in the analysis of collaborations between natural and social scientists. I am collaborating across disciplines more intensively than ever before, and this gives rise to pressing questions about how to negotiate these collaborative relationships. I have given three

---

[11] These lower-scale objectives are often those that are shared between scientists in different disciplines, such as molecular biologists and computer scientists.

examples of my involvements in synthetic biology and I have drawn on literature from anthropology and science studies to make sense of these experiences. I suggested that it is necessary to juggle multiple identities in a way that could be considered somewhat schizophrenic, and I explored the idea that the social scientist could be thought of as a kind of jester or trickster in these collaborative relationships. I showed how, in the light of my experiences, the contrast between social scientific researcher and scientific 'informant' was problematic on many levels, and brought up issues of friendship, rapport, complicity and being compromised. Finally I turned to recent work in the anthropological literature about epistemic partners, paraethnographers and collaboration as a research method. My reason for drawing on these ideas was to show that methodological reflection is valuable in navigating situations where there is increasing collaboration between scientists, engineers and social scientists.

A question that remains is whether the issues I have discussed here are specific to synthetic biology, or whether they can be said to apply to more generally to promissory and potentially disruptive technoscientific fields where new relations between science, technology and society are being created. Earlier in this paper, I noted that synthetic biology possesses certain features which made it particularly well-suited to collaborations with social scientists. The first is that it is already interdisciplinary, but this is a feature shared with most emerging technologies, because they are likely to develop at the intersection of existing disciplines. Something that is perhaps more distinctive to synthetic biology is the attempt we see to build a community as well as a technology – in the integration of Human Practices into the iGEM competition, for example. There are, of course, attempts to build scientific communities around other new fields, but in synthetic biology there is the ambition that this community should have a broad scope, because of the aim of some branches of synthetic biology to democratise the technology. These aspirations are not so explicit in other new areas of science and technology, such as nanotechnology and stem cell research. For these reasons, synthetic biology does seem to lend itself particularly well to interdisciplinary interaction, and, I would argue, to collaborative knowledge-making. Whether these opportunities for collaborative knowledge-making will remain as the field becomes more mainstream and established is an open question, but one that that it will be extremely interesting to ask as synthetic biologists and social scientists muddle through together in the years ahead.

References

Bijker, W. E. (1993). Do not despair: There is life after constructivism. *Science, Technology & Human Values, 18*(1), 113–138.

Buergi, B. R. (2010, September 2nd–4th). *How do we collaborate? Scrutinizing the relationship between STS and biomedicine*. European Association for the Studies of Science and Technology, Trento.

Calvert, J., & Martin, P. (2009). The role of social scientists in synthetic biology. *EMBO Reports, 10*(3), 201–204.

Croissant, J. (1999). The view from the basement: The ethics and politics of teaching engineers while studying them. *Anthropology of Work Review, XX*(1), 22–27.

Doubleday, R., & Viseu, A. (2009). Questioning interdisciplinarity: What roles for laboratory based social science? In K. Kjolberg & F. Wickson (Eds.), *Nano meets macro: Social perspectives on nano sciences and technologies* (pp. 51–75). Singapore: Pan Stanford Publishing.

Dyson, F. (2007, July 19). Our biotech future. *The New York Review of Books*. http://www.nybooks.com/articles/20370/. Accessed 21 June 2012.

Edge, D. (1995). Reinventing the wheel. In S. Jasanoff, G. E. Markle, J. C. Petersen, & T. Pinch (Eds.), *Handbook of science and technology studies* (pp. 3–24). Thousand Oaks: Sage.

Fischer, M. (2011, January 6th–8th). *Biopolis: Model organism? Biopolis enters its third five years.* Asian Biopoleis: Biotechnology & biomedicine as emergent forms of life & practice, National University of Singapore, Singapore.

Fisher, E. (2005). Lessons learned from the Ethical, Legal and Social Implications programme (ELSI): Planning societal implications research for the National Nanotechnology Programme. *Technology in Society, 27*, 321–328.

Forsythe, D. E. (1999). Ethics and politics of studying up in technoscience. *Anthropology of Work Review, XX*, 6–11.

Fortun, M. (2005). For an ethics of promising. *New Genetics and Society, 24*(2), 157–173.

Frow, E., & Calvert, J. (2013). "Can simple biological systems be built from standardized interchangeable parts?" Negotiating biology and engineering in a synthetic biology competition. *Engineering Studies, 5*(1), 42–58.

Guston, D. H. (2006). Toward centres for responsible innovation in the commercialized university. In J. M. Porter & P. W. B. Phillips (Eds.), *Public science in liberal democracy: The challenge to science and democracy* (pp. 295–312). Toronto: University of Toronto Press.

Hamlin, C. (1992). Reflexivity in technology studies: Toward a technology of technology (and science)? *Social Studies of Science, 22*, 511–544.

Haraway, D. (1991). The actors are cyborg, nature is coyote and the geography is elsewhere. Postscript to "Cyborgs at large". In C. Penley & A. Ross (Eds.), *Technoculture* (pp. 21–26). Minneapolis: University of Minnesota Press.

Holmes, D. R., & Marcus, G. E. (2008). Collaboration today and the re-imagination of the classic scene of fieldwork encounter. *Collaborative Anthropologies, 1*, 81–101.

Illes, J., & Bird, S. (2006). Neuroethics: A modern context for ethics in neuroscience. *Trends in Neuroscience, 29*(9), 511–517.

Johnson, D., & Wetmore, J. (2007). STS and ethics: Implications for engineering ethics. In E. J. Hackett, O. Amsterdamska, M. Lynch, & J. Wajcman (Eds.), *The handbook of science and technology studies* (pp. 567–581). Cambridge, MA: MIT Press.

Kevles, D. J., & Hood, L. (1992). *The code of codes: Scientific and social issues in the human genome project*. Cambridge, MA: Harvard University Press.

Law, J., & Callon, M. (1988). Engineering and sociology in a military aircraft project. *Social Problems, 35*, 284–297.

Macnaghten, P., Kearnes, M. B., & Wynne, B. (2005). Nanotechnology, governance and public deliberation: What role for the social sciences? *Science Communication, 27*(2), 268–287.

Marcus, G. E. (1997). The uses of complicity in the changing mise-en-scene of anthropological field work. *Representations, 59*(13), 85–108.

Marcus, G. E. (2005, June 27–28). *Multi-sited ethnography: Five or six things I know about it now problems and possibilities*. Multi-sited Ethnography workshop, University of Sussex. Online at: http://eprints.ncrm.ac.uk/64/1/georgemarcus.pdf/. Accessed 21 June 2012.

Marcus, G. E. (2008). The end(s) of ethnography: Social/cultural anthropologys signature form of producing knowledge in transition. *Cultural Anthropology, 23*(1), 1–14.

Marris, C., & Rose, N. (2010). Open engagement: Exploring public participation in the biosciences. *PLoS Biology, 8*(11), e1000549. doi:10.1371/journal.pbio.1000549.

McGregor, J., & Wetmore, J. (2009). Research and teaching the ethics and social implications of emerging technologies in the laboratory. *Nanoethics, 3*(1), 17–30.

Paul, L., & Van den Belt, H. (2006). *The institutionalisation of ethics in science policy; Practices and impact*. Work Package 5: Ethics in food technologies. Online at: http://library.wur.nl/ebooks/1885868.pdf

Rabinow, P. (2002). *French DNA: Trouble in purgatory*. Chicago: University of Chicago Press.

Rip, A. (2005, June 29). *There is mainstreaming, loss of critical distance: Are STS scholars finally growing up?* Paper presented at workshop Does STS Mean Business Too?, Said Business School, Oxford.

Rip, A. (2006). Folk theories of nanotechnologists. *Science as Culture, 15*(4), 349–365.

Robertson, J. A. (2001). Human embryonic stem cell research: Ethical and legal issues. *Nature Reviews Genetics, 2*, 74–78.

Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology, 41*(5), 885–899.

Schot, J., & Rip, A. (1997). The past and future of constructive technology assessment. *Technological Forecasting and Social Change, 54*, 251–268.

Schuurbiers, D. (2008). *Ethics in action. Winning essay of the Mekelprize 2008 for PhD students*. Platform for ethics. Online at: http://www.cspo.org/library/other/?action=getfile&file=183&section=lib/. Accessed 21 June 2012.

Selin, C. (2007). Expectations and the emergence of nanotechnology. *Science, Technology and Human Values, 32*(2), 1–25.

Smolke, C. D. (2009). Building outside of the box: iGEM and the BioBricks Foundation. *Nature Biotechnology, 27*(12), 1099–1102.

Stirling, A. (2005). Opening up or closing down? Analysis, participation and power in the social appraisal of technology. In M. Leach, I. Scoones, & B. Wynne (Eds.), *Science and citizens. Globalization and the challenge of engagement* (pp. 218–231). London: Zed.

Turnbull, D. (2000). *Masons, Tricksters and Cartographers: Comparative studies in the sociology of scientific and indigenous knowledge*. Amsterdam: Harwood Academic Publishers.

Vikkelsø, S. (2007). Description as intervention: Engagement and resistance in actor-network analyses. *Science as Culture, 16*(3), 297–309.

Webster, A. (2007). Crossing boundaries: Social science in the policy room. *Science, Technology & Human Values, 32*(4), 458–478.

Zhang, J. Y, Marris, C., & Rose, N. (2011). *The transnational governance of synthetic biology: Scientific uncertainty, cross-borderness and the art of governance* (BIOS Working Paper No. 4). Online at: http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/publications/2011/4294977685.pdf

Chapter 10
Ethicists in the Laboratory: Reflecting About Non-existent Objects

Simone van der Burg

Abstract It has often been questioned whether ethics on the laboratory floor is useful, because there is not yet a technology to evaluate in the earlier phases of research. In this article it is argued that ethics does not need the existence of the object it discusses, for its assessments to be meaningful. In discussion with Peter-Paul Verbeek's ethics of design, and Arie Rip's prospective ontology, this chapter defends an intensionalist approach to technology which is inspired by Alexius Meinong. This approach allows to distinguish between technologies that are part of reality, and those that are not, without making the realm of the non-existent meaningless. Just like scientific talk about possible capacities of technologies is meaningful, for it leads to assumptions that can be researched, ethics is also able to evaluate those capacities. Both scientists and ethicists are concerned with characteristic capacities of something, before that 'something' exists. If we accept that scientists do that, there seems to be no reason why extra arguments should be provided to prove that ethics is a meaningful activity in the laboratory too, and could assess a technology that is still 'in the making'.

In the last 5–10 years, a number of proposals have been put forward to locate ethics on the laboratory floor. The emergence of this type of ethics can be interpreted as an attempt to provide ethical feedback timely, when it is still able to inform the development process of new technologies. This laboratory approach is often contrasted with an ethics that provides assessments after a technology finished developing, for at that moment in time it is too late to change anything fundamental about it:

Simone van der Burg acknowledges CSG – Centre for Society and the Life Sciences, which provided funding for research that lead to this paper.

S. van der Burg (✉)
IQ Healthcare, Radboud University Nijmegen Medical Center,
114 IQ Healthcare, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands
e-mail: S.vanderBurg@iq.umcn.nl

N. Doorn et al. (eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*, Philosophy of Engineering and Technology 16, DOI 10.1007/978-94-007-7844-3_10,

too many parties – researchers, producers and funding agencies – have invested time, energy and money in it and have an interest in selling it (Collingridge 1980). An earlier engagement of an ethicist, which is sometimes called an upstream or midstream engagement, is thought to enable scientists and developers of technology to integrate ethical perspectives into their deliberation about research and development choices, which influences the course of the research and development process, as well as the eventually resulting technological product. The purpose of such engagements is to co-shape the technology in ethically more favorable ways.

The early involvement of ethicists is sometimes referred to as an 'ethics on the laboratory floor'. This does not necessarily mean that the work of ethicists is confined to the space of the laboratory, and that they solely accompany scientists who carry out their experiments at the laboratory bench. Laboratory work also includes meetings that take place in other rooms within or outside of the (academic) institute where the laboratory is located (Boenink 2013). These meetings may include researchers and technicians who carry out the experiments, but also their more senior supervisors and external partners, such as users or representatives of the industry and/or (public) research funding institution. Writing research project proposals – even though it does not take place inside a laboratory – is also part of laboratory work, since it shapes the purposes of the activities performed at the laboratory bench, as well as the research consortium that is involved in realizing it.

What is characteristic of a laboratory ethics is that it is involved in- and accompanies research and development of new technologies, and aims to enhance ethical reflection about them while they are being created. Laboratory ethicists are of course not the ones who introduce morality into the laboratory. Values and norms are already part of the everyday way in which scientists work and interact. Just like in any other lasting cooperation or cohabitation of people, these values and norms can be made explicit in research protocols and guidelines, in systems of praise and blame, but they may also remain implicit in the ways in which participants in a research project treat each other, in their mutual expectations and in the reactions and emotions that result when expectations are not met. While scientists themselves may engage in ethical reflection about the morality that guides their interaction, ethicists may clarify and enhance that reflection by means of an articulation and clarification of the (tacit) values and norms that are at stake, and a broadening of the imagination of the future in which the technology may become widely used.

In this volume several examples are provided of methodologies that fit these general constraints of an ethics on the laboratory floor, most notably in the contributions by Erik Fisher and Daan Schuurbiers (Chap. 5) and by Ibo van de Poel and Neelke Doorn (Chap. 6). These contributions articulate methods that are designed to shape the reflection of scientists and developers by means of close collaboration with a humanist or an ethicist in a specific R&D project. Their approaches differ. Van de Poel and Doorn aim, for example, to identify the ethical issues and the fair distribution of responsibilities, by means of interactive workshops that function to broaden the reflection of each participant by means of an exchange with other actors in the R&D project as well as other stakeholders, such as representatives of the users. The ethicist in this context, figures as a facilitator of the workshop, and a guide in it. The STIR method as described by Fisher and Schuurbiers, is a more dialogical approach for it proposes that the humanist in the laboratory probes the capacity of technical experts to reflect on the societal relevance of their research by means of asking responses to questions of clarification and justification.

In comparison, the two other methodologies of CTA and value-sensitive design are certainly effective to ‘open up the lab’, but they do not fit within the constraints of a laboratory ethics. Reflective workshops which adopt a Constructive Technology Assessment approach, usually take place outside of the laboratory and they take place once (so they do not accompany a research project). Furthermore, they have a broader aim of bringing about a co-evolution of technology and society, which addresses also the macro (science policy) and meso level (institutions). Batya Friedman et al. and Alina Huldtgren, alternatively, focus not on laboratory research but on design. While designers deliberate about possible ways to realize something, scientists are – especially in the earlier phases of research – often uncertain about what this ‘something’ is going to become and whether it can be realized at all.

Adding to the variety of activities that are already represented in this book and which mainly aim to shape an ethics of participants in a research project, this contribution will focus on ethics on the laboratory floor and its ambition is to contribute to the shaping of the object that is being researched. Next to a focus on the morality of the ‘makers’, it also looks at *what* is being made. This object of research – the technology – may still have a very uncertain character during research, which is a challenge for ethicists in the laboratory. Ethicists in the laboratory are therefore often confronted with the question whether it is useful to reflect on the moral value of the ‘something’ even if it is not yet a ‘something’.

This chapter will explore possible answers to this question which have been put forwards by Peter-Paul Verbeek who is one of the rare philosophers who has attempted to look at ‘things’ as embodying a morality, and Arie Rip’s inspiring ‘prospective ontology’. Building on these predecessors, this chapter will however argue that a clearer distinction is needed between the ‘real’ and the not (yet) real which can be sought in an intensionalist approach.

10.1 Evolving Shapes

The evolving nature of new technology is sometimes captured vividly in metaphors, such as Lorraine Daston’s metaphor of a ‘biography’ to refer to descriptions of evolving objects of science (2000). This metaphor suggests an analogy between the development of a technology and a human life, which offers an alternative to the common understanding of technology as an unchanging object with features and a function. Technologies that are being researched may acquire different shapes; they may occur as an informal research idea, a promise in a research project, a laboratory set-up, a prototype or an instrument that can be used in tests on users.

It is already quite common to compare the existence of technologies with the dynamics of a human life. In the work of Bruno Latour and Don Ihde, for example, technologies do not figure as passive material objects, but as material artifacts that offer active contributions to the world of human interactions and experiences (Latour 1992, 1993; Ihde 1990). Both authors abandon the metaphysical divide between subject and object, which has often been ascribed to the Enlightenment: they show how objects shape the actions, experiences and identities of subjects, or even how they can be understood as actors themselves. Peter-Paul Verbeek builds on this abandonment of the Enlightenment subject-object distinction in *What things do* (2005) to form his own philosophy of technological mediation, describing and analyzing how understandings of human life – including understandings of the good life – always generate in contexts in which artifacts are also present. This analysis was the basis of his more recent book *Moralizing technology* (2011) in which Verbeek calls for the development of a 'posthumanist' ethics which not only focuses on the decisions that human beings take, but also on the moral significance of technologies. Part of *Moralizing technology* is devoted to the development of an ethics of design, whose purpose is to materialize morality in the eventually resulting artifact, hence, the title of the book 'moralizing technology' (p. 113, Verbeek 2011).

This original book offers a very insightful and exciting perspective to the ways in which ethics can be relevant for technology. Verbeek offers suggestions as to how to anticipate the ways in which a new technology may shape human interactions when it will become used, which contribute to the realization of the more desirable/ethically attractive possibilities that a technology is capable of offering. The suggestions that Verbeek brings forwards include: (1) a study of the actual moral imagination of designers about the future roles that their technology could play in diverse contexts of use, which can be written down in the form of scenarios, (2) the use of a method of Constructive Technology Assessment which broadens the imaginations of designers by means of stakeholder engagements during which intended users express their perspectives. The approach to CTA that Verbeek proposes is augmented with a view of technological mediation, as he proposed it in his book *What things do*. Thirdly, (3) the building of simulations that allow designers to experiment with different possible user interactions with the design.

A laboratory ethics shares with Verbeek's approach the interest in shaping technology, and some of his suggestions are useful in a laboratory context too. A study of the future imaginations of the scientists, for example, may be needed to come to an understanding of the technology that they are working on, since – especially in the earlier phases of research – this technology is not yet 'there' as a tangible or visible device, artifact or system. CTA is most certainly a helpful tool to broaden these imaginations, and may enhance reflection about contexts of use – and about the quality of life it that the technology may help to realize for users. But not all of Verbeeks suggestions are equally useful for a laboratory context. It is for example a bit early to experiment with different technological futures in simulations. Such experimentation becomes a possibility when technologies are already able to 'do' something, and different possible interactions with users can be imagined. Verbeeks examples all fit into this category: they already have known capacities, and simulations help to reflect on the question to what kind of interactions they could give rise. A couch, or a diswasher are already able to do something – to offer people a seat, to wash dishes – and simulations allow to imagine how different designs may invite other engagements with it. A couch can, for example, be imagined which is made of material that becomes prettier with use, thus encouraging people to keep it and stimulate longevity; a dishwasher with a separate button for rinsing communicates to users that they needn't pre-rinse their plates before putting them into the machine, which saves water. Novel technologies that Verbeek discusses – such as the foodphone or ambient intelligence – also already have a clear capacity. While *what these technologies become* can still be altered in important ways during design, the *existence* of the capacities of these technologies is already established: they have capacities, whose meaning is explored in different contexts of interaction with people.

For Verbeek, interaction in a social context is constitutive for what a technology is. In agreement with Latour, Verbeek argues that technologies become what they are in interaction with users in a specific context, and there is no identity prior to that interaction. But simulations presuppose that technologies are able to *do* things, even though the character of their actions may change with the context of interaction. Based on this presupposition it is possible to ask: how should a new technology shape human interactions? What behavior should it enhance? What behavior should it help to avoid? These are questions that are important for ethicists who aim to materialize morality during design. But ethicists who consider technologies in an earlier phase, in the laboratory, cannot always ask the same questions. Borrowing a phrase by Deborah Johnson, it is characteristic for a laboratory ethics that it investigates the ethically relevant aspects of something that is still 'in the making' and which is for that reason not yet a 'something' (Johnson 2007). Especially during the earlier phases of research the capacities of a technology are only imagined possibilities, and it is the purpose of research to point out which one of them will eventually materialize, if any at all.

Ethics on the laboratory floor is therefore an effort to reflect on the moral aspects of technologies which are not yet 'existent' objects; they are not yet part of our reality, and it is not sure if they ever will. Hence the understandable critical question whether it is useful to think about such non-existent objects at all, or whether such reflection is better postponed to a context of design.

10.2 Material Expectations

Scientists are often not convinced that ethical reflection about the object of their research is a useful activity at all. Frequently, the advice is given to 'come back later', when research is more advanced, and more reliable predictions can be done about whether or not the technology that is being researched has a chance of being realized. The presupposition behind this advice is that there first should be 'something', before any reflection about how it should be shaped and what features it should get could be useful at all.

Such friendly advice generates from well known concerns. Arie Rip addressed these concerns in his discussion of expectations in ‘Technology as prospective ontology’ (Rip 2009). The notion ‘ontology’ is here interpreted broadly as the ‘furniture of the world’, and Rip brings forwards a variety of examples to show that expectations about the future are already part of that furniture, and are therefore part of reality: such as, expectations in research proposals or in demos of future products (cars, kitchens, military equipment), websites that envision future applications of cloning (clone a pet) or nanotechnology uses inside familiar artefacts such as suncream or in coatings. These expectations are ‘real’ in the sense that they inspire (research) activity, investment of time and money, emotion and (political) debate. Nanotechnology is, according to Rip, a domain where ‘prospective ontologies materialize’ (p. 416, Rip 2009), meaning that special visualizing technologies allow access to their materiality at the nanoscale, but they also materialize in narratives that articulate their possible future. Narratives, according to Rip, are also ‘material’, for they occupy space and time in the sense that they may be voiced at a location, or materialize in a printed text, a book or a website (See note 4, p. 407, Rip 2009).

By calling them ‘material’ parts of reality, Rip seems to want to point out that promises and expectations are not just subjective visions or phantasies that people project on the world and that will simply alter when people change their minds; expectations resist such inconstancy, for people invest time, energy and money into the further development of a perspective to the future, therewith adding to the credibility that they will eventually be things, devices or systems that invite other types of interaction. Arguing for the materiality of these narratives, visions and expectations, Rip shows that the future is already part of our reality now, and that it therefore deserves attention, including political (or ethical!) attention.

This appealing thesis offers arguments to support the value of an early debate on emerging technology, which may include ethical topics. By pointing out that being ‘real’ does not require the presence of a technology, Rip opens space to think broadly about what our world is like: this world may include technologies, but also imaginations which elicit desire or repulsion, interest, emotion, money and (social) activity. Since they are part of the world, they may also be an object of political or ethical reflection and debate.

At the same time, however, Rip’s perspective also blurs the distinction between different types of materialities that furnish the world. What seems not to matter, in Rip’s prospective ontology, is whether a vision of the future is informally uttered during a meeting, or written down in a memo, as a promise in a research project or media-publication, or if it becomes a test set-up, demo or prototype. The future can be experienced and/or talked about sensibly in a variety of ways, and all of them are part of the furniture of our world, because they figure as objects of our attention, activities, investments etc. But in these anticipations of the future, technologies are already being imagined as capable devices with which it is possible to interact, regardless of the research stage that they are in. But if expectations and promises are ‘material’ parts of the world, just like a prototype is or an artefact that we have used for decades, then distinctions disappear that usually play an important role in how we think and talk about what it is that furnishes the world, as well as how we tend to interact with it.

Rips' prospective ontology elicits a view of reality on the basis of interactions. Promises and expectations are part of reality, because they invite action and interaction. On the basis of this approach it could be argued that research objects invite a different type of action or reaction from researchers, investors or members of the public than the objects of design do, or technologies that are ready for use. Such a separation between the reality of technologies on the basis of characteristic interactions with it, is however a difficult basis to say something about where ethics fits. Is it an appropriate reaction to technology during research? Or is ethical reflection a characteristic activity for technologies that are already in use? If there is nothing about the technology that justifies an ethical interaction with it, we seem to be left with no reason to go for an earlier or later ethical engagement.

10.3 Ethics of 'Non-existent Objects'

In daily conversation as well as in a lot of philosophical and sociological thought on technology, the tendency is to speak about objects, artefacts, things or systems that are being researched, rather than about the invention of types of interactions. The history of this focus on objects is a long one, and it reached a climax in the debate over non-existent objects between Bertrand Russell and the Austrian philosopher Alexius Meinong who was famous during the turn from the nineteenth to the twentieth century for his seemingly paradoxical claim that 'there are nonexistent objects'. The potential impact of this claim on the present discussion about the appropriate timing of ethical thinking about technology justifies elaborating it to some extent.

Meinongs work, which roots in a phenomenological tradition, was much admired for some time, but it was largely forgotten for many decades after Russell's criticism of this paradoxical claim in *On Denoting*. Russell did not discard Meinongs claim right away, for he was intrigued and corresponded with Meinong about it for some time. In *On Denoting*, however, he strongly argues that there are no objects that don't exist, and defends a form of what has become known as the 'extensionalist theory'. This extensionalist theory holds that subjects and predicates have a different role in the meaning of a sentence. According to this theory, every basic statement (proposition) can be analyzed as involving a subject concept and a predicate concept. For example, in the sentence 'The Dutch prime minister is honest', the subject (the Dutch prime minister) identifies the topic of the proposition, while the predicate (honest) says something about it. Once the subject of this proposition has been found (for example, we can see the Dutch prime minister), the proposition is ready to convey its information about it and we can decide whether the proposition is true or false. But if the subject in the proposition does not refer to anything (for example, the Dutch government is demissionary and there is no Dutch prime minister), the sentence loses its meaning. This means that it is not enough that there is a subject in the sentence: in order for the proposition to have meaning, it must also refer to something in the world (it must be *about something*).

According to the extensionalist theory, meaningful speech and thought depends on its relation to objects, or its *aboutness*. Alexius Meinong, however, builds on a phenomenological background and defends an intensionalist theory. He accuses the philosophy of his time, which includes also the perspective of this extensionalist approach, of being guilty of *the prejudice in favour of the real*.[1] This prejudice rules out the possibility that we speak meaningfully about something that does not (yet) refer to any object that is present in the actual world, but according to Meinong we are actually doing that a lot in our everyday speech. Examples of 'nonexistent objects' are objects such as the giant dwarf, the golden mountain, the winged horse Pegasus and the round square, which may occupy the imagination, while they are not actually part of the real. According to Meinong, speech about them is meaningful, because we can talk about their features, quite apart from their reality. Even though there is no object in reality to whom characteristics such as 'round' *and* 'square' can be assigned at the same time, it is possible to think about it. Or in Meinongs words: it is possible to speak of predicates quite independently of the existence of the object. This claim by Meinong has become known as the principle of the independence of so-being (Sosein) from being (Sein).[2] Meinong holds that we can attribute any property whatsoever to thought objects, independent of whether they exist.

While Meinong himself is not concerned with technologies 'in the making', this principle of the independence of so-being from being could be used to for objects of scientific research – such as technologies – that can meaningfully occupy the mind, without actually being part of reality. Scientists in the laboratory are more concerned with the so-being of an object, than with the object itself. In the earlier phases of research into a technology, scientists are not actually imagining an object as if it were an existing thing: they are imaging features, characteristic actions or functions, quite apart from the existence of the object. They are concerned with characteristics, quite apart from their being an object, a device or thing to ascribe them to.

Experience as an ethicist in the laboratory this fits very well with this picture. I was involved in research into photoacoustics, which was investigated for its capacity to detect breast cancer (Van der Burg 2009, 2010). The technology was based on angiogenesis, which is the idea that cancer needs blood to spread and grow, and that therefore malignant tumours can be recognized by the excessive vessel growth around them. Photoacoustic technology combines light and sound with the purpose to detect the presence of extra blood vessels. But in the laboratory set-up there is not yet a technological device present. Rather, there are several technologies: there is a laser, an ultrasound transducer, a computer, and a substitute for the human

---

[1] Meinong (1988, p. 3).

[2] Meinong (1988, pp. 7–9). See also Albertazzi et al. (2001), Jacquette (2001), Routley (1980), Schubert Kalsi (1987) for further explanations of Meinongs approach.

breast – interestingly called a phantom – with a tube with ink in it that represents the blood vessel. During the experiments it is established that light is absorbed by blood (which is colourful). This laser light beam is pulsed – it goes on and off – which causes short temperature rises and falls at the locations where it is absorbed. When the temperature rises, the tissue expands, when it falls, the temperature drops. This expanding and shrinking makes a sound which can be pictured with ultrasound. The eventual product that the scientists attempt to realize is a picture of the presence of blood vessels in the breast, which indicates that there is a malignant tumour in the breast.

Scientific thought about a sound-light technology such as the photoacoustic mammography brings to mind considerations about ‘non-existent objects’ such as the round square or the giant dwarf. It combines and mixes aspects of our experience, which are hard to imagine together. Technologies such as these could be called ‘synesthetic technologies’, referring to a neurological disorder that affects that sounds also produce visual experiences such as colour, or the other way around, that experiences of colour produce sounds. Likewise, photoacoustic mammography translates light (associated with sight) into temperature (tactile senses) into sound (auditive senses) into an image (sight). Photoacoustic mammography makes use of normal human experiences, but combines them in such a way that they go beyond the experiences that most people are capable of having; thus making possible visible sounds and audible vision. Just like giant dwarfs and round squares demand to imagine combinations of opposites which are therefore called ‘nonexistent’, audible vision and visible sound transgresses the limitations of what is considered to be the usual furniture of ‘our world’.

In a laboratory set-up there is often not yet a device, but a collection of things. In the study into photoacoustic mammography, for example, there is a set of devices (a laser, computer, acoustic transducer) which are put together to study the association or interaction between them. It is the characteristics of this interaction that interests the scientists. Quite apart from the existence of photoacoustic mammography as a device, these seemingly incombinable capacities are researched which define what these devices are supposed to ‘do’. Scientists talk, think, write and work on the characteristic capacities of this technology, before there is a device to whom they can be ascribed. Before there is a ‘something’, scientists talk sensibly about the predicates of this ‘something’; such as its efficacy, speed, adequacy, resolution etc.… Nobody considers this kind of talk ‘insignificant’ or ‘meaningless’, as long as these predicates are backed up by sensible experiments, and theoretical knowledge. But just like we grant that scientists can think about and discuss characteristics of their technology prior to its existence, ethicists can do that too. Ethicists are able to talk meaningfully about predicates such as good, bad, equal, just, honest, autonomous, generous, related to capacities that are being researched, prior to the existence of a device whose capacities are thus predicated. There is therefore no reason why scientists can do their work, and ethicists should wait until there is a device to evaluate. Just like scientists deal with a non-existent object during research, so do ethicists. Both are able to speak meaningfully about it, prior to its existence.

10.4 Concluding Remarks

In this article it is argued that ethics does not need the existence of the object it discusses, for its assessments to be meaningful. It builds this claim on discussions with Peter-Paul Verbeek's ethics of design, and Arie Rip's prespective ontology. Verbeek looks at technology in a later stage, which is therefore only partly useful for an ethics on the laboratory floor. In supplement to Verbeek, Arie Rip shows that technologies in an earlier phase of research can also take a meaningful part in human (political) interaction with them. Prior to their existence as technologies, they are already present in the human world as expectations, narratives, promises which inspire thought and action of scientists, research funders, producers, investors etc. Ethics can be one of the ways to interact with technologies during research.

Rip's approach, however, raises the question whether a distinction should be made at all between technologies at different stages of research and development. An expectation in a research proposal and a prototype are likewise part of the material world we live in, although in daily life we are used to making distinctions between them. While Verbeek and Rip choose a dynamic interactive approach to technology – and rarely talk about them as 'objects' – an extentionalist presupposition seems to play a role in the background of their approach. The discussion about what it means to be 'there', and if it is also possible to speak meaningfully about technologies which are not there, is interestingly solved by showing the 'materiality' of expectations – thus pulling non-material parts of the world into materiality, in an effort to prove how 'real' they are. While it is a very valuable effort to show that expectations are already part of reality, and should therefore be taken seriously as objects of ethical or political thought and action, it is questionable whether they should be imported in this way into the realm of the real.

An intentionalist approach allows to distinguish between technologies that are part of reality, and those that are not, without making the realm of the non-existent meaningless. This approach allows the possibility to talk meaningfully about technology, prior to its existence. Scientific talk about technologies that are to be researched, after all, is also about technologies that are not yet 'there'. There seems to be no reason why ethical thought could be only meaningful after there is an object, while scientific thought can be meaningful before that object exists. Both scientists and ethicists are concerned with characteristic capacities of something, before that 'something' exists. If we accept that scientists do that, there seems to be no reason why ethicists should provide extra arguments to prove that they can think and talk meaningfully about a technology while it is still 'in the making'.

References

Albertazzi, L., Jacquette, D., & Poli, R. (Eds.). (2001). *The school of Alexius Meinong*. Aldershot/Burlington/Singapore/Sydney: Ashgate; Schubert Kalsi, 1987.

Boenink, M. (2013). The multiple practices of doing 'ethics in the laboratory': A mid-level perspective. In S. van der Burg & T. Swierstra (Eds.), *Ethics on the laboratory book*. Hampshire/New York: Palgrave/Macmillan.

Collingridge, D. (1980). *The social control of technology*. London: Pinter Publishers.

Daston, L. (2000). *Biographies of scientific objects*. Chicago: The University of Chicago Press.

Ihde, D. (1990). *Technology and the lifeworld*. Bloomington: Indiana University Press.

Jacquette, D. (2001). Außersein of the pure object. In L. Albertazzi, D. Jacquette, & R. Poli (Eds.), *The school of Alexius Meinong*. Aldershot/Burlington/Singapore/Sydney: Ashgate.

Johnson, D. (2007). Ethics and technology 'in the making': An essay on the challenge of nanoethics. *Nanoethics, 1*, 21–30.

Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In W. E. Bijker & J. Law (Eds.), *Shaping technology/building society*. Cambridge, MA: MIT Press.

Latour, B. (1993). *We have never been modern*. Cambridge, MA: Harvard University Press.

Meinong, A. (1988). *Uber Gegenstandstheorie; Selbstdarstellung*. Hamburg: Felix Meiner Verlag.

Rip, A. (2009). Technology as prospective ontology. *Synthese, 168*, 405–422.

Routley, R. (1980). *Exploring Meinong's jungle and beyond; An investigation of noneism and the theory of items*. Canberra: Australian National University.

Schubert Kalsi, M.-L. (1987). *Meinong's theory of knowledge*. Dordrecht: Martinus Nijhoff Publishers.

Van der Burg, S. (2009). Imagining the future of photoacoustic mammography. *Science and Engineering Ethics, 15*(1), 97–111.

Van der Burg, S. (2010). Ethical imagination: Broadening laboratory deliberations. In S. Roeser (Ed.), *Emotions about risky technologies* (International library of ethics, law and technology). Dordrecht: Springer.

Verbeek, P.-P. (2005). *What things do: Philosophical reflections on technology, agency and design*. University Park: The Pennsylvania State University Press.

Verbeek, P.-P. (2011). *Moralizing technology; Understanding and designing the morality of things*. Chicago/London: The University of Chicago Press.

Chapter 11
Metaphors and Cohabitation Within and Beyond the Walls of Life Sciences

Eleonore Pauwels

Abstract This contribution will first describe some of the contexts for cross-field collaborations within the life sciences, and then will highlight relevant theoretical reflections, including the concept of "insertion," "modulation," and "trading zones." Based on what can be learned from trading zones, we will use synthetic biology as a case-study to explore one of the inherent difficulties in evaluating the promises of our biotechnological futures: the role and faith of engineering concepts and metaphors within and beyond the walls of life sciences. We will then reflect on the importance of mutual learning between the two cultures of natural sciences and humanities to unpack what might be lost in translation through the use of engineering concepts and metaphors. We conclude with a range of institutional challenges to be tackled when it comes to promoting cohabitations within and beyond the walls of life sciences.

11.1 Introduction: An Introductory Digression Around the Concept of *Cohabitation*

> Can we cohabitate with you? Is there a way for all of us to survive together while none of our contradictory claims, interests and passions can be eliminated?
> (Bruno Latour (2005), *From Realpolitik to Dingpolitik- or How to Make Things Public*)

This quote from Bruno Latour suggests alternative ways of doing what we have been used to call "collaboration" between fields, between sectors, between cultures, and between publics. It calls for alternative ways to assemble and dissemble around

---

E. Pauwels (✉)
Science and Technology Innovation Program, Woodrow Wilson International Center for Scholars, 1300 Pennsylvania Avenue, Washington, DC 20004, USA
e-mail: Eleonore.Pauwels@wilsoncenter.org

N. Doorn et al. (eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*, Philosophy of Engineering and Technology 16, DOI 10.1007/978-94-007-7844-3_11,

the issues we care for. It calls for improving and renovating our "techniques of representation," meaning the different techniques that contribute to make "public" the issues we care for, to unveil what we consider being a "matter of concern." These *matters of concern* are as diverse as the issues that assemble a concerned public around them: just think about the 2008 financial meltdown and its economic and political ramifications, the revolutions erupting in Maghreb and Machreq, nuclear proliferation, the spread of genetically engineered mosquitoes to fight dengue, research around bio-energy including the development of synthetic engineered algae. Around every one of these areas of concern we see growing entanglements of passions, indignations, and controversies within a complex web of stakeholders and opponents. *Matters of concern* create an "agora;" they create political conditions for dissenting imaginations.

Matters of concern move us from what has been called *Mode 2*[1] of knowledge production to the *Agora* – "where science and innovation interact with societies"[2] – and provide a role for modes of collaborations of a more complex kind. In this case, scientists, engineers, policy-makers and diverse layers of societal actors, sensitised through engagement to wider social imaginations, might decide for themselves to approach science and innovation differently. As explained by Stirling about current discourses on sustainability (2009:5):

> Often, the position is expressed as if there were 'no alternatives.' The questions asked are thus typically restricted to 'yes or no?'; 'how much?'; 'how fast?' and 'who leads?' If we move instead to more plural understandings of progress, then the quality of debate – and of the ensuing choices – thereby stands to be enriched. Instead of fixating on some contingently-privileged path, we might ask deeper, more balanced and searching questions about 'which way?'; 'what alternatives?'; 'who says?' and 'why?' This is the essence of a normative, analytic, epistemic, ontological – and consequently intrinsically political – project of 'pluralising progress'.

The above excerpt eloquently demonstrated the importance of being politically receptive to dissenting imaginations. Instead of designing endogenous modes of collaborations, the prelude is intended to more reflexively understand the political background within which actors from different fields of social practices will be invited to interrogate particular framings of socio-technological regimes and their potential transition pathways, and to re-open them for debate (Stirling 2008; Smith and Stirling 2008). In this journey involving research and policy actors capable of questioning the status quo, there is a necessary need for "daring to imagine" (Wynne 2009), for reflexivity and for empowerment as suggested by Jamison (2010:13): "change-oriented research is about empowerment, by which the researcher applies

---

[1] Several STS academics have suggested that the traditional "Republic of Science" is being replaced by a new "Mode 2" of knowledge production (Gibbons et al. 1994). Two properties linked to this new "Mode" – transdisciplinarity and an orientation toward problem-solving – are particularly relevant for our discussion.

[2] This concept of the "Agora" was introduced by Andy Stirling in the Session "Sustainability and Emerging Technologies" at the 2009 Conference of the Society for Social Studies of Science (4S), October 29, 2009.

knowledge gained from experience to processes of social learning, carried out together with those being "studied"."

This specific notion of empowerment requires to be attentive to what Wynne calls an "epistemic other" (2009:13): "it is difference manifesting itself as an unknown set of realities, acting themselves as unknowns and beyond our control (but not beyond our responsibility), into a world we thought we controlled." On the surface of this epistemic variety, a democratically-committed knowledge-society is supposed to have the scientific and political imaginations to work out how a plurality of social actors could share knowledges, practices, and experiences with diverse scientific, policy and economic actors (Jasanoff 2009). It is the unveiling of the conditions prevailing to these improved forms of collaborations and the main challenges they are facing – in particular, the use of group-specific language and metaphors – that I wish to explore in this contribution. And eventually the term "cohabitation" should be preferred to "collaboration." Indeed, the concept of collaboration itself is matter to be discussed. Cross-field and -sector collaborations have too often been considered as "fusion" – where actors converge towards a premeditate vision or goal, suppressing *ipso facto* the room for a diversity of knowledges, practices and experiences; too often, collaborations are experienced as an attempt to co-optation – meaning that the instrumental support of a field, such as ethics, philosophy or sociology, is required to make up for an interdisciplinarity of "façade." The term "cohabitation" entails more: it presupposes that we leave enough room for different frameworks of thinking to seat together, exchange and ultimately develop visions that are based on a true diversity of claims, knowledges and imaginations. That kind of diversity, institutional, legal, epistemic, disciplinary and discursive, is part of what it takes to be more reflexive. Diversity is not synonymous with reflexivity or even conducive to it, but it's a necessary prerequisite for a healthy reflexive discourse on new technologies and must be preserved and appreciated.

Now, what does "reflexivity" mean? How does it differ from "reflection?" Reflection is thinking broadly about all the possible facets of a phenomenon as if there were as far as possible in an environment. This thought process is a crucial ingredient in reflexivity. But reflexivity is also situating yourself as an observer in the system, dissociating, from this system, your own interests, and thinking about your position amidst the things you are looking at. The law, for example, is reflective and, even, reflexive as it regards precedents. How can we correctly implement reflexivity? It is about creating arenas and spaces where *matters of concern* can be unveiled, where cohabitation between fields does not suppress dissenting voices or anyone's concerns. As said by Latour, "is there a way for all of us to survive together while none of our contradictory claims, interests and passions can be eliminated?" Such arenas are the intellectual and collective spaces where it becomes possible to question the status quo by questioning concepts such as "progress," "power" and "ownership." For example, it is necessary to directly ask what progress means for society as we move forward in science and technology innovation. The question of what constitutes true progress in synthetic biology is paramount right now for a field that is busy designing synthetic organisms. As we are progressively expanding the engineering of biological organisms, which humans have been doing with agriculture for thousands of years, but today, with more control, we must ask what purpose and for whose purpose we are doing these forms of engineering.

This contribution constitutes the premises of a thought experiment (Gedanken-experiment) around the concept of *collaboration*, or more exactly, *cohabitation* between fields, sectors, cultures, and ways of approaching regimes of technoscientific innovation. Ultimately, how do you create the infrastructures so that complex ways of thinking from different fields, sectors and cultures can meet somewhere and learn from each other? How can we think about forms of "cohabitation," where researchers from different fields could reflect together on research design, research questions and trajectories? Is it possible for different socio-technical imaginations to cohabit? What are the necessary conditions (institutional, epistemic, political and cultural) to develop different forms and places for reflexivity in different contexts such as the educational systems, the policy systems, or the laboratories?

This contribution will first describe some of the contexts for cross-field collaborations within the life sciences, and then will highlight relevant theoretical reflections, including the concept of trading zones. Based on what can be learned from trading zones, we will use synthetic biology as a case-study to explore one of the inherent difficulties in evaluating the promises of our biotechnological futures: the role and faith of engineering concepts and metaphors within and beyond the walls of life sciences. We will then reflect on the importance of mutual learning between the two cultures of natural sciences and humanities to unpack what might be lost in translation through the use of engineering concepts and metaphors. We conclude with a range of institutional challenges to be tackled when it comes to promoting cohabitations within and beyond the walls of life sciences.

11.2 CONTEXT – The Rise of the New Biology: What Does It Mean for Cross-Field Collaborations?

More than 50 years ago, in the Senate House in Cambridge, the celebrated novelist C. P. Snow delivered his now famous intervention, "The Two Cultures and the Scientific Revolution."[3] He depicted the gap that had opened up between scientists and "literary intellectuals" and argued that practitioners in both areas should build bridges, to further the progress of human knowledge and to benefit society.

---

[3] The Two Cultures is the title of an important 1959 Rede Lecture by British scientist and novelist C. P. Snow. It explores how the lack of interactions and knowledge-sharing between the "two cultures" of modern society – the sciences and the humanities – was a significant obstacle to solving the world's problems. Several influential thinkers within the field of Science and Technology Studies have successfully begun to revisit C.P. Snow's divide. Jasanoff (2004, 2005), for example, explains through the analytical framework of co-production how the objects and practices of scientific research are embedded in larger moral, legal, and social environments, and vice-versa.

Not surprisingly, Snow's vision has gone unrealized. Instead, as analyzed by Steven Shapin, the scientific persona itself is progressively evolving into one of entrepreneurship. For instance, it is not unusual to see life scientists increasingly and successfully becoming part of a web of private laboratories, start-up companies, and ventures developing around the promises of the New Biology. Some of them are not only brilliant minds; they become expert in communicating directly with stakeholders about their work using media and targeting large readership.[4] These dynamics not only reinforce the notion that engineering of life has value (bio-value), but also nurtures the related regime of techno-scientific promises supposed to advance societal goals. How these common goals and other domains of public good are actually defined and negotiated is a Pandora's box that has only occasionally been opened to public scrutiny.

Regimes of techno-scientific promises, which combine general societal "progress" with technological advances, have been building stones of our polities for decades now. The life sciences are an integral part of these regimes of promises. Synthetic biology, with its aim to engineer biological pathways, lies at the heart of what the U.S. National Research Council has called A New Biology for the twenty-first century (NRC 2009). This report recommends that a "New Biology" approach – one that depends on greater integration within biology and closer collaboration with physical, computational, and earth scientists, mathematicians, and engineers – be used to find solutions to four key societal needs. These societal needs are sustainable food production, ecosystem restoration, optimized biofuel production, and improvement in human health.

Particularly, the New Biology is considered a paradigm shift by some of its proponents in the sense that it destroys some of the walls methodically built around disciplines such as physics and biology. In their piece "Calculating Life," Calvert and Fujimura (2009) eloquently analyze how the New Biology is increasingly tied up to the exigencies of a presupposed social contract. They take the example of systems biology to show the new interactions happening at the science-society border:

> […] systems biologists articulate a division between the 'social' and 'scientific' elements of their field, and then openly discuss the value of the social aspects. This is in part because they work in a situation that differs significantly from the more common way of doing science in the past century. Instead of working within disciplinary boundaries, systems biologists see their research as transgressing these boundaries. They attempt to integrate not only data and technologies, but also disciplines and people. Scientists talk about how "the development of systems biology depends on the sociology" and how it is important to cultivate a social environment in which scientists with different expertise can work together productively.

---

[4] In (2007), J. Craig Venter – the renowned scientist who plaid a significant role in the race to deciphering the Human Genome – published his biography "A Life Decoded – My Genome: My Life." In (2010), Rob Carlson, another active proponent of the development of the new biology, published a book on synthetic biology written for a large readership "Biology is Technology – The Promise, Peril, and New Business of Engineering Life."

Calvert and Fujimura then continue their analysis by referring to the "wall" metaphor which, interestingly, brings us back to this notion of *cohabitation* we have been alluding to earlier:

> Another demonstration of the interconnectedness of the social and the scientific is the commonly expressed idea that systems biology has 'no walls'. This point is made in a metaphorical sense: systems biologists maintain that the field draws on expertise from whichever area is most useful or appropriate at the time because "ideas are everywhere", as one interviewee noted. Two senior researchers even thought that the disciplinary spread of systems biology could extend to the social sciences and humanities. The idea of 'no walls' also has currency in a literal sense because there are no walls between the laboratories at many systems biology institutes, in order to facilitate communication between researchers. New buildings have interdisciplinarity purposely built into the design, with social spaces where the 'wet' experimental people and the 'dry' computational people can easily come across one another.

Such testimonies show that modes of integrating life sciences are slowly emerging. But whether those are punctual attempts or successes to come in terms of promoting non-endogenous collaborations remains to be seen since dissensions already exist within centuries-old biology. On this specific argument, Calvert and Fujimura bring extremely valuable testimonies from inside molecular biology and systems biology:

> The issue of whether systems biology constitutes a paradigm shift comes up most often in comparisons with molecular biology. In fact, Fred Boogerd and colleagues note that "practicing systems biologists are often hindered by paradigm battles with molecular biologists" (Boogerd et al. 2007). A computer scientist highlighted the antagonism between the two fields by saying "it's still very much an 'us and them' thing between the molecular and the systems people". This is in a context where, until recently, "the mainstream was dominated by the reductionist molecular biology agenda". Against this background, it is perhaps not surprising that systems biologists often experience resistance from the 'old school' of molecular biology. Indeed, some interviewees argued that the antagonism towards systems biology from the proponents of previous paradigms is itself a sign of a paradigm shift.

In its "Vision of the Future," the NRC Report envisages a drastic integration of several fields that are thought to be key in solving sustainability challenges confronting our societies (NRC 2009, viii):

> Given the fundamental unity of biology, it is our hope and our expectation that the New Biology will contribute to advances across the life sciences [...] [T]he life sciences have the potential to provide a set of tools and solutions that can significantly increase the options available to society for dealing with problems. Integration of the biological sciences with physical and computational sciences, mathematics, and engineering promises to build a wider biological enterprise with the scope and expertise to address a broad range of scientific and societal problems.

Such a vision postulates a form of unity within biology, which, as shown in the above excerpts, is contested and might therefore create a potential for fragmentation and disillusion along the road. When it comes to designing fruitful, balanced collaborations, the gap is even more difficult to bridge between the "two cultures" of life sciences and humanities.[5] Yet, not only collaborations within the life sciences but also beyond, including the social sciences and humanities, is needed. The revolution that the "New biology" (NRC 2009) imposes to the life sciences, its nature and goals, preferably would require parallel adaptations in societal governance, but despite the efforts of visionary researchers to overcome the divisions between the two cultures of humanities and natural sciences,[6] the New Biology has been imagined mainly in the minds of biologists, other natural scientists, mathematicians, and engineers. A comprehensive understanding of the epistemic, ontological, and normative changes induced by this "New Biology" paradigm would benefit from the involvement of researchers from humanities and social sciences.

11.3 Theoretical Input: Crossing the Line "In and Out" the Laboratory

Recent research in Science and Technology Studies (STS) has been conducted with a goal to develop new theoretical frameworks based on constructive technology assessment, midstream modulation, trading zones, shared expertise, moral imaginations and epistemic cultures that show promise for understanding and facilitating interdisciplinary collaborations (see Chaps. 3, 5 and 7 of this book; Schot and Rip 1997; Fisher 2007; Fisher and Mahajan 2006; Fisher and Schuurbiers 2009; Gorman 2004; Gorman et al. 2004, 2009; Gorman and Mehalik 2002; Knorr-Cetina 1999).

11.3.1 Insertion and Modulation

Two approaches, within CTA, are of particular interest to this contribution: the notion of "insertion" and "moving about."[7] These two ethos have been coined as crucial methodologies within CTA experiments and refer to CTA practitioners as "knowledgeable visitors." They "insert" themselves upstream into the lab or the R&D ecosystems and question, probe, and interrogate practices and decisions leading to specific

---

[5] The field of bioethics has been quite successful at collaborating with teams of natural scientists and might be a source of learning for what mode of collaboration works and which does not work. But we also may want to be "reflexive" and examine the cases where ethicists might have been used as "token ethicists" and lacked room for questioning the research design, questions and trajectories.

[6] Sheila Jasanoff (2004) is noted for her work on co-production: the analytical framework of co-production directly pertains to governance issues by exploring how the objects and practices of scientific research are embedded in larger moral, legal, and social environments, and vice versa.

[7] See Chap. 3 by Rip and Robinson.

research pathways. They are not only "inserted" or "embedded;" they are also "moving about," meaning they are following objects or practices under investigation through multiple "floors," including research laboratories, conferences, planning events and public debates. Through this insertion and "moving about" journey, they do not only acquire the necessary empirical knowledge to understand the visions, expectations, and practices part of the scientific enterprise, they also bring different or dissenting questions and connections to places where science "happens."

Similarly, within midstream modulation, the "embedded" humanist finds him or herself confronted with the exploration and questioning of the everyday decisions taken with the laboratory. If a form of dialogue becomes possible (some experiments seem to be unsuccessful, Cf. Rabinow and Bennett 2012), the "embedded" humanist might be able to develop and share a reflexive point of view and therefore exert a modulation within the process of laboratory life. The extent to which this reflexive point of view is limited by factors as multiple as funding, authority, and inability to set a dialogue across different disciplinary paradigms is well explained in Chap. 5 by Fisher and Schuurbiers.

How do these two ethos differ or come close to what this contribution describes as cohabitation? Cohabitation is about a house, an arena, or a journey where imaginations are confronted and shared to think about what societal progress is and how technology contributes to it. It is a journey where "insertion" and "modulation" are made possible upstream enough for a discussion about societal progress to happen before R&D choices are set in stone. It is a journey that allows for imaginations to be rather focused on large technological systems rather than a punctual technological choice. In fine, these notions are different but close trajectories to think about how dissenting imaginations can coexist into innovation systems.

11.3.2 Trading Zone, Interactional Expertise and Cross-Field Collaborations

The metaphor of a "trading zone" has been first developed by Peter Galison to explain how scientists and engineers from different disciplinary cultures manage to collaborate across apparently incommensurable paradigms (Galison 1997). Through case-studies in physics, Galison found that different epistemic communities had to develop first jargons, then pidgins, and finally full-scale creoles to be able to share perspectives across their own scientific paradigms. He noticed that, despite coming from contrasting scientific paradigms, experts were able to develop communication processes which can be seen as a trading zone: in this trading zone, experts use what Bromme calls a group-specific language which usually relies heavily on the development of metaphors as a bridge between different epistemic paradigms (Bromme 2000). The concept of trading zones has been significant to better understand interdisciplinary collaborations. The "trade" metaphors adequately portray the way academic experts are increasingly used to meeting, exchanging ideas, learning from each other, and then returning to their epistemic community with concrete "goods" in the form of improved research practices (Morreale and Howery 2002).

In parallel with the reflection on trading zones, Collins and Evans (2002) have described the different levels of expertise that play a role in interdisciplinary collaborations: one of special interest to our research is the interactional form of expertise. The interactional expert corresponds to an agent who understands enough of the language and norms of the different epistemic cultures involved in the trading zone to facilitate the trade. For example, early in the development of MRI, surgeons interpreted as a lesion what an engineer would have recognized as an artefact of the way the device was being used; this breakdown in the creole between these communities was recognized and solved by an interactional expert who had a background in both physics and medicine (Baird and Cohen 1999).

Trading zones might also incorporate collective overarching goals such as sustainability. Indeed, sustainability has emerged as the ascendant policy issue of the twenty-first century. While we continue to argue about the true definition of "sustainability" – particularly since it has become a fashionable buzzword for the policy community and related funding agencies – the challenge of converting our present socio-technical system to a "sustainable" system has developed as a new master narrative, inspiring policy discourse both in Europe and the United States. Trading zones that build around these collective overarching narratives need to apply what has been defined as "moral imagination" (Werhane 1999). The concept of moral imagination assumes that human beings learn practical ethics from deep stories, collective archetypes, and unconscious, or often emotive dimensions of a problem or paradox that become models for ethical behaviors (Johnson 1993). These stories are usually invisible, unquestioned, and progressively adopted as simple accounts of the reality (Sethi and Briggle 2011). Moral imagination reasserts that these stories are contingent views that need to be confronted with the "epistemic other" (Wynne 2009) when a serious dialogue about different worldviews is at stake.

11.3.3 Epistemic Cultures and Negative Knowledge

The framework of epistemic cultures analyzes the dynamics of knowledge-production – their amalgams of arrangements and mechanisms bonded through affinity, necessity, and historical coincidence (Knorr-Cetina 1999). This framework might be useful to explore the potential for new modes of cross-sector (public/private) and cross-disciplinary collaboration between the life sciences and social sciences to develop reflexivity in scientific practices. Synthetic biology, in particular, has witnessed the development of these "lab-scale interventions" (Rabinow and Bennett 2009; Fisher 2007). The rationale behind these collaborative ventures is to identify social and ethical controversies further upstream in the R&D process.

These collaborations also promote bridges that enable the communication of ethical and regulatory insights from the social sciences and bioethics component back to the laboratory. At the same time, such collaborative practices would benefit from being anchored in trading zones involving outsiders to the lab and non-institutional networks such as DIYBio or private conglomerates.

Part of the dynamics within epistemic cultures, the notion of "negative knowledge" (Knorr-Cetina 1999) and the problematization of the non-production of knowledge (Proctor and Schiebinger 2008), might illuminate some of the controversies around the safety/societal implications of synthetic biology. Indeed, a thorough understanding of the factors at stake in the non-production of knowledge might help understand what knowledge is "lost in translation" through the use of metaphors in life sciences. The concept of negative knowledge seeks to analyze the limits of knowing, the mistakes we make in trying to know, the things that interfere with our knowing, what we are interested in, and what we do not really want to know. This notion of negative knowledge might give insight into the way research agendas are built and delimited, especially when it comes to research oriented toward solving societal ills.

In this contribution, we postulate that trading zones have the potential to improve technology governance by integrating interdisciplinary assessments through gradual co-production of methodologies, analyses and concepts. Trading zones could act as spaces for the articulation of plural scenarios about technological innovations and, *ipso facto*, promote the transmission of social, ethical, and regulatory controversies from social sciences to the lab and *vice versa*. However, a critical mass of the trading zone participants need to have enough knowledge and understanding of the language and norms of the different cultures involved in the zone to be able to facilitate the trade. This *sine qua non* condition invites us to think that the use of metaphors in science is far more complex than the ends it serves in literature and politics. For example, when synthetic biologists use metaphors such as "software of life," an analogy is drawn between biological and computational systems that allows synthetic biology to borrow important concepts like 'feedback loops,' 'information,' 'robustness,' 'noise,' from its engineering and computation counterpart. Interestingly, these concepts are not static. They are refined through scientific practices. These concepts sometimes even have different meanings to different researchers in different contexts. For example, for some researchers the concept of noise relates to a nuisance and complexity that should avoided; for others, it constitutes a source of uncertainty that can be both negative or positive – leading to unexpected discoveries. This "plasticity" of metaphors and related engineering concepts should be explored through the practices and the discourses that make them evolve.

Due to their interpretative force and potential influence on public visions, we need a reflective and more critical use of metaphors when it comes to scientific practice. This chapter could function as a trigger for more in-depth discussions on the various functions of metaphors or contribute to already existing discussions.

11.4 Case Study: Engineering Metaphors as a Challenge to Cohabitation Within and Beyond the Walls of Life Sciences

> Books of life, junk DNA, DNA barcodes: all these images can and have distorted the picture, not least because scientists themselves forget that they are metaphors. And when the science moves on, when we discover that the genome is nothing like a book or a blueprint – the metaphors tend, nonetheless, to stick. The more vivid the image, the more dangerously seductive and resistant to change it is.
>
> (Philip Ball, Nature, February 23, 2011)

To explore the potential of using trading zones upstream in technology assessment and, *ipso facto*, the challenge of developing a group-specific language, this chapter will focus on the problematic use of engineering metaphors within the emerging field of synthetic biology. Decisive questions to be asked are: What is lost in translation through engineering and mechanistic metaphors? What is the role of metaphors in "keeping biology complex or not?" How are ignorance and uncertainty constructed, imposed and manipulated through metaphors? What spaces are left for epistemic openings in an attempt to explore diverse meanings of synthetic biology?

11.4.1 Metaphors as an "Inside-Out" Border

Arrays of metaphors and concepts used in synthetic biology have their origin in information theory (computation) and engineering. A good example is the use of "chassis" – sometimes "safe chassis" – to define the basic functionalities of a bacterial genome on top of which forward-engineered biological systems can be implanted. And indeed, most scientists view metaphors as an essential teaching tool. Chemist Theodore Brown's book (2003:14), "Making Truth: Metaphors in Science," opens with the following central thesis and tone:

> […] metaphorical reasoning is at the very core of what scientists do when they design experiments, make discoveries, formulate theories and models, and describe their results to others–in short, when they do science and communicate about it. Metaphor is a tool of great conceptual power. It enables the scientists to interpret the natural world in wonderful and productive ways. At the same time, the metaphorical reasoning that lies at the heart of scientific thought and imagination is constrained in ways that go toward defining the range and character of science.

The power and effects of metaphors might be even more pervasive when it comes to new scientific developments where metaphors usually are supposed to unveil the esoteric. Interestingly, metaphors in science often act as "boundary-objects," residing in the "in-between" spaces characterized by epistemic aspirations, tensions and uncertainties. In the 1990s, Star and Bowker considered boundary objects as locations to study individual practices in institutional arrangements: “These objects may be abstract or concrete…but their structure is common enough to more than one world to make them recognizable, a means of translation. The creation and management of boundary objects is a key process in developing and maintaining coherence across intersecting communities” (Bowker and Star 1999:297). What is of interest to us is the ability of boundary objects to exist in multiple communities which allows them to serve as translational devices between disparate articulations of practice. This translational function is essential within the walls of the life sciences when cross-field fertilization happens to build a new field like synthetic biology. But this function goes beyond the walls of life sciences. The communicative spaces metaphors represent also allow for the translation and extrapolation of synthetic biology’s potential from scientific into policy and public discourses. This is why, beyond their uses as boundary objects in communities of practices across life sciences, metaphors also play a key role in the production of knowledge about the New biology and in the construction of our common biotechnological futures.

11.4.2 What Is Lost in Translation?

This section introduces a few empirical reflections which arise from a 5-months interdisciplinary study undertaken by researchers who present the form of interactional expertise we have been alluding to earlier. From January to May 2012, a physicist and philosopher of science as well as a linguist teamed up to study the role and faith of engineering metaphors within synthetic biology.[8] Through the use of data mining software, these researchers explored the peer-reviewed literature on synthetic biology[9] and identify the metaphors used as well as their related conceptual categories. As a next step, through laboratory observation, interviews and qualitative analyses, they conducted an in-depth investigation as to where the metaphors come from and with what meaning they are used in scientific and public communication.

When used within the walls of life sciences, in dialogues between biologists, engineers, computer scientists and physicists, metaphors serve as inspiration but are robust enough to progressively create an identity among merging communities of practices. As explained by one of the synthetic biologist interviewed, the majority of metaphors in synthetic biology have their origin in information theory (computation) and engineering:

> In synthetic biology I found a very different reception, from molecular biology, microbiology and the broader genome community in that, in the late ‘90s, I think each of those communities recognised they were now facing, complex systems that were very difficult to

---

[8] Andrea Loettgers is currently a researcher at the California Institute of Technology and has a background in physics and philosophy of science; Eleonore Pauwels is a public policy scholar at the Wilson Center in Washington DC and has a background in linguistics and public policy.

[9] Source: Web of Science

> understand. […]Those communities recognised that they needed to now better understand the circuits, the networks the pathways in which, these parts operated and I think they turned to physicists and engineers and said "what can you do?" and.. synthetic biology grew in part out of the response of those communities to that […] where, I think we saw that there was an opportunity to reverse engineer and better understand actual systems. But that the data was still quite sparse and that we could simultaneously, or make better progress by, taking a forward engineering approach where we could, build up small systems out of these components de novo, and study their behaviours.

He goes further by explaining how the use of engineering concepts and metaphors constitute a differentiating factor in the construction of the synthetic biology field:

> I think they (engineering concepts and metaphors) are fundamental, to our work and… I think it's a differentiating factor of our lab, and other synthetic biology labs from more traditional cell biology approaches to these systems. And we use inspiration from engineering, both with the design but even more critical we use engineering principles, both to guide the design of our synthetic constructs as well as to aid in our understanding of how natural systems function.

This last excerpt shows the extent to which, for synthetic biologists, engineering concepts and metaphors are not translated literally into biology but submitted to a lot of tinkering, what Andrew Pickering (1995) called the "Mangle of Practice." Engineering concepts and metaphors are used as inspiration but they have to accommodate biology:

> We had to accommodate the biology. I think here, it's probably more an analogy than a metaphor, but it's where, from the circuit design standpoint, many in the field will typically take as inspiration, a circuit design or a functional circuit goal from engineering, but then we of course are constrained by what you can do with the biological systems and as a result, the creative comes in then thinking about what design will allow us to accomplish the functional goal, accounting for the biology. […] The constraining aspect of the analogy then, comes in where you get the differences between say biology and engineering systems. And a good synthetic biologist will work hard to try to account for those differences and either, accommodate by trying to compensate for them or, to accommodate them by embracing them and actually working them into the function.

Yet, some of the practitioners are reflexive enough to show the limitations of engineering concepts and metaphors in conveying the inherent complexity characterizing life sciences. They raise attention to the meanings inevitably "lost in translation" when using engineering concepts and metaphors such as "legos," "biobricks," "building blocks:"

> I think the biggest misconception that arises is, you get the sense that, what we do is as easy as building a building out of Legos. And, that we're in a position now to.. re-engineer organisms, in.. very prescribed ways, quite readily and, this is where I think the misconception arises is that by, invoking these concepts which are not yet practice. By giving the sense that synthetic biology is the engineering of biology and we now are in a position where we can engineer biology and.. standardisation, BioBricks, leads to this misconception that, this is now readily easy for us to come up with complete new organisms, or change things. […] The danger is that if you limit yourself to thinking about the engineering analogies or what we' already built, you may actually not open yourself up to, seeing new biological principles that do not yet have engineering analogues. […] Are you missing how the system actually works by trying to fit it, into your electrical engineering analogies or, mechanical analogies?

By trying to implement logical digital circuits into biochemical networks, synthetic biologists could run into what Fujimura (2005:220) described in one of her articles as forcing conceptual framework onto biological systems which could lead to a loss of important information. She goes further by emphasizing how the metaphor might provoke some distortion:

> The reason for disentangling the materialities of engineered and biological systems is to delineate the various border-crossings in order to understand what is lost in the translation. By ‘lost’, I mean more than ‘loss’. Translations can distort, transform, delete, and add. For example, although Kitano, Doyle, and Hood understand that the Boeing 747 and the automobile are too simple to emulate biological systems, they nevertheless use principles from these systems to model biological systems. In the process, their models still may excise whatever cannot be translated into the instrumental and technical terms of control engineering as they calibrate between biological organisms and virtual, artificially created advanced technologies. This excision does not make their productions any less material or real. Instead, one of my purposes for studying systems biology is to ask whether there are other productions that could have been made in their place.

There are thus a few research questions we may want to keep in mind when analyzing the use of engineering concepts and metaphors in scientific communication about synthetic biology: What is lost in translation through engineering and mechanistic metaphors? What alternative use of metaphors could have been made? What kind of knowledge and ontologies are produced?

Fujimura (2005:211) outlines an important paradox emerging from the engineering-control approach which is likely to raise questions at the science-society frontier:

> The control engineering approach appears to be a top down, engineered systems approach to biological organisms that begins with particular design requirements and principles. In contrast, biological organisms are ostensibly the results of evolution, which means that the organism and the species as well as the evolving environments are historically contingent products. Is the engineered systems approach too mechanistic and naive, given the historical and contingent production of biological organisms?

This excerpt alludes to the importance of dealing with broad assessments of the impact of the engineering community and its collective practices on social or biological systems: If engineering and command-control principles continue to dominate systems biology’s modeling of living organisms, what will be the resulting products and social consequences? What is implied in terms of safety and security through engineering-control metaphors? In which terms do synthetic biologists define what a success would be in their research? What are the implications of these control-engineering metaphors for our socio-ecological and socio-technical systems? Are there alternative models?

11.4.3 Hype and Promissory Futures

Beyond the walls of life sciences, when synthetic biology is portrayed into public discourse, reports of the current success and future potential of synthetic biology are frequently illustrated with images depicting speculative and futuristic visions.

As related in *The New Yorker* of September 2009 (Specter 2009), "synthetic biologists are convinced that, with enough knowledge, they will be able to write programs to control those genetic components, programs, that would let them not only alter nature but guide human evolution as well." The communicative spaces enabled by metaphors and the visions they produce present "life" not only as it is, but also "life as it could be" or "life as we could make it be." A good example of futuristic use of a metaphor is the Synberc "tumor-killing bacteria." Despite its weak reference to current achievement in synthetic biology, this image serves as a means of communication for the "exchange" of expectations between the discourses of science, economy and the mass media. This should encourage us to ask a few additional questions: What biotechnological futures are anticipated through metaphors? Are there other ways, other metaphors to be used while thinking about the future of synthetic biology? How should techno-scientific promises be distinguished from hype?

Another eye-catching image of synthetic biology is the one depicted in a recent article of *The New York Times* Magazine titled "The God of Small Things – Craig Venter's bugs might save the world:" "In the menagerie of Craig Venter's imagination, tiny bugs will save the world. They will be custom bugs, designer bugs – bugs that only Venter can create. He will mix them up in his private laboratory from bits and pieces of DNA, and then he will release them into the air and the water, into smokestacks and oil spills, hospitals and factories and your house. Each of the bugs will have a mission. Some will be designed to devour things, like pollution. Others will generate food and fuel. There will be bugs to fight global warming, bugs to clean up toxic waste, bugs to manufacture medicine and diagnose disease, and they will all be driven to complete these tasks by the very fibers of their synthetic DNA. Right now, Venter is thinking of a bug. He is thinking of a bug that could swim in a pond and soak up sunlight and urinate automotive fuel. He is thinking of a bug that could live in a factory and gobble exhaust and fart fresh air."[10]

This vision is simultaneously futuristic and foreseeable, reminding us that synthetic biology is ultimately part of a technological continuum anchored in the Enlightenment and constantly progressing through techno-scientific breakthroughs, such as recombinant DNA technologies.

Behind this impression of a continuum, however, there is something salient in the visions populating synthetic biology; through intentional biological design and manufacturing, engineered life forms – from engineered yeast to Venter's "synthetic cell" – are becoming "factories" on their own. In short, while laboratories have grown into "factories" through twentieth century's collective imaginaries, today synthetic biology design turns the living cell itself into a factory. To this effect, Peter Galison (1999) remarkably analyzed how scientific practices and understandings have evolved through the nineteenth century from an Enlightenment culture seeking to unveil nature's true face, to a regime of mechanical objectivity. Scientific practices have progressed from those of intervening genial individuals to ones at ease building

---

[10] http://www.nytimes.com/2012/06/03/magazine/craig-venters-bugs-might-save-the-world.html?_r=1&pagewanted=all

and supervising precise machines. The below excerpt depicts the transformations occurring within the sanctuary of the laboratory (Galison 1999:33–34):

> Many features of the laboratory and factory coincide; they are deeply linked, and often co-produced. One can point, for example, to worker discipline, centralized power sources, and architecture – as well as shared political economic ideals of maximizing work and minimizing waste. But for our purposes here, the key commonality is the joint fascination with the reduction of individual variability through the use of machines: the production of regularity as a positive virtue that was simultaneously moral and epistemic. It was here that the quieting of the will met the discipline and self-restraint of the factory. […] Scientific laboratory workers had long taken on the mantle of self-disciplined supervisors of machine. When scientists announced with pride in objectivity that they would do nothing to impose individual variation on the regular, uniform, and reliable output of their machines, they were testifying not only to the power of science in industry, but to the conjoint understanding of laboratory and factory.

The vision of a future inhabited by "living factories" constitutes a significant and symbolic pace on the road to the molecular economy. It epitomizes and reinforces what some have called the production of "biovalue" within a "moral economy of hope" (Rose and Novas 2005:442, 452):

> Biology is no longer blind destiny, or even a foreseen but implacable fate. It is knowable, mutable, improvable, eminently manipulable. Of course, the other side of hope is undoubtedly anxiety, fear, and even dread at what one's biological future, or that of those one cares for, might hold. But whilst this may engender despair or fortitude, it frequently also generates a moral economy of hope, in which ignorance, resignation, and hopelessness in the face of the future is deprecated. This is simultaneously an economy in the more traditional sense, for the hope for the innovation that will treat or cure stimulates the circuits of investment and the creation of biovalue. […] It also tries to encapsulate the ways in which life itself is increasingly locked into an economy for the generation of wealth, the production of health and vitality, and the creation of social norms and values.

This transition toward increasing reliance on the production of biovalue and the techno-scientific promises that surface in the aftermath presents a kaleidoscope of interesting epistemological and ontological claims. The extensive use of engineering concepts and metaphors in the emergence of synthetic biology portrays the field as one easy to grasp and, at the same time, a very appealing and promising endeavour (Specter 2009). These mechanistic representations are anything but new in biotechnology and genetic engineering, where metaphors or images constructed to represent new processes, products, and their potential effects have widely adopted mechanistic models. Beyond the need to sketch the functioning of biological systems, these models also convey the implicit reassurance that these systems can be optimized and that they are reliable and under control; their behavior is predictable. This reassuring concept has also affected the design of regulation; mechanistic metaphors have been used as examples of mitigating uncertainties and managing safety aspects (OTA 1989).[11] Additionally, the effects of these images and metaphors are

---

[11] In 1989, almost coincidentally with the release of the first U.S. patent on a complex organism, the Oncomouse, the Office of Technology Assessment ("OTA") published the report entitled *Patenting Life*. In order to stress the analogy between mechanical and biological inventions, and thus the inevitable patentability of organisms, the OTA showed, side by side, the two drawings accompanying, respectively, the Mousetrap (patented in 1900) and the Oncomouse.

amplified by the fact that, as with most emerging sciences, the practitioners in charge of mapping synthetic biology are also concurrently inventing it (Sethi and Briggle 2011).

Above all, in the scientific and public spheres, synthetic biology fits into a regime of innovation based on techno-scientific promises and therefore is epitomized through metaphors and narratives that involve the articulation of a vision (Wynne et al. 2007). Often this articulation takes the form of hype. Vision and hype are both types of discourse that look toward the future. The vision of synthetic biologists is a future where humans engage in the large-scale design and creation of new life forms that are exquisitely tailored for human purposes. The genetic engineering of organisms and the extensive design and manufacture of living things from virtual genetic sequences blurs the line between machine and organism, life and non-life, and the natural and the artificial, and thus transforms the relationship between human kind and nature in ways that are exciting to some people but troubling for others (Bedau et al. 2009; Pauwels 2009).

In the near future, there might be a need to explore the readiness of the engineering profession to address the ethical and social issues associated with our biotechnical futures. The possibility of error, human and otherwise, is why history is important when we think about future technologies. How well have we managed the introduction of other technologies? Have we, as a society, learned anything?

11.5 Questions: To Researchers and Funding Agencies

> Transgressing disciplinary boundaries… [is] a subversive undertaking since it is likely to violate the sanctuaries of accepted ways of perceiving. Among the most fortified boundaries have been those between the natural sciences and the humanities.
>
> (Valerie Greenberg (1990) – *From Transgressive readings: The text of Franz Kafka and Max Planck*)

Some recent research initiatives have started to revisit what C.P. Snow called the “two cultures” gap (see the chapter by Fisher and Schuurbiers (Chap. 5) and the chapter by Gorman et al. (Chap. 8) in this volume). They intend to promote different ways in which the cultures of science – far from standing apart from the rest of the academic disciplines – are in timely conversations with the cultures of the humanities, the social sciences, the arts, and the law.

One of these initiatives is called “lab-scale intervention.” Nanotechnology – and to a limited extent, synthetic biology – has witnessed the development of these new modes of cross-disciplinary collaboration between natural sciences and humanities that help develop reflective scientific practices. The rationale behind these collaborations is to recognize issues of concern, from ethical uncertainty to social controversies, high upstream in the research and innovation process. These collaborations are also supposed to promote more rapid communication on ethical and regulatory matters from the law and bioethics component back to the laboratory. Encouragingly, recent studies show that it is possible to form an interdisciplinary trading zone in which a scientist and a humanist jointly explore a cutting-edge topic in nanotechnology. Concretely, engineers and humanists become actively involved in the process of knowledge-exchange, better described as “knowledge-trading,” with the consequent result that some engineers and humanists develop long-term interactions, building trust and enabling mutual learning by working together in hybrid collectives.

In the future, these tandems of researchers from different disciplines could develop plural narratives and metaphors that promote the transmission of scientific, ethical, and regulatory controversies from the social sciences to the lab and vice versa. These tandems would function as a mirror or a “reflexivity tool” for the life sciences involved in synthetic biology design and the social sciences interested in the related implications. In a “knowledge-society,” these cross-boundaries tandems could also be opened to the policy and public sphere by including policymakers, NGOs, investors, and science journalists.

Indeed, such collaborative experiments will be enriched by continual conversations with those outside the lab, including policymaking communities and non-institutional networks such as Do-It-Yourelf-Biology and private conglomerates. Such an early dialogue between researchers and policymakers, for example, would help identify moments of safety or regulatory uncertainties in synthetic biology trajectories, or what Brian Wynne calls “epistemic other.” Indeed, policymaking communities do not need only a clear perspective on the challenges posed by synthetic biology to ethics and politics but must also promote, inside public policy communities, more reflexive thinking on the social and normative dimensions of synthetic biology design.

Though these cross-disciplinary attempts are still nascent, they already raise questions and require us to be critical: to what extent do these lab-scale studies lead to better capacity to critically analyze the relevance of synthetic biology promises to societal goals? To what extent do they allow us to collectively experiment with possible alternatives within synthetic biology? To what extent will they succeed in developing co-production among multiple disciplines and perspectives from the outset as opposed to downstream reflection upon the ethical, legal, and social implications of synthetic biology?

With the above interrogations populating our imagination, we feel the need to continue our thought experiment around the concept of cohabitation and make clearer to the reader what the virtual border is between traditional modes of *collaboration* and *cohabitation*. This thought experiment interestingly echoes recent observations made by a group of researchers in “Towards a Manifesto for Experimental Collaborations between Social and Natural Scientists” (Balmer et al. 2012). The writing process of this manifesto was purposely open and pluralist: “On 19th June at Kings College London, natural scientists, engineers and social scientists interested in synthetic biology gathered to discuss some proposed principles for negotiating and practicing collaborations that we had initially drafted to stimulate discussion at the event. Everyone was fruitfully engaged and it became clear that to move ahead we need more discussion, more comments and more engagement. We now put these principles forward as an invitation to further dialogue.” Among these principles, several, including the notions of reflexivity, pluralism, collective experiment and hospitality, converge towards what we think are the predicaments for building cohabitation. Cohabitation could be a term to name these improved modes of collaboration the manifesto is inviting us to develop.

The manifesto opens with a background observation similar to what Fujimura and Calvert (2009) and the author of this contribution have attempted to shed light on: researchers from social sciences are too often allowed solely in one room of the house, limited to establishing ethical and social implications of technological choices already built in stone. In brief, researchers from natural and social sciences are not sharing the same space but are separated by rigid walls slowly built around "funding arrangements, disciplinary and institutional boundaries, governance regimes and local politics" (Balmer et al. 2012). Yet, years of research (e.g. Jasanoff 2004) have showed the extent to which the scientific enterprise is primary a social construction, how intertwined choices of scientific and social relevance are, and, *ipso facto*, how crucial it is to be able to question and confront these choices with wider, sometimes dissenting, imaginations. As eloquently said in the room of King's College, "this means both natural and social scientists can work together to produce a critical and human understanding of how design, development and application of new technologies are accomplished. We can make technologies and science more socially responsive and relevant by understanding the dynamics of lab work and innovation to identify where choices are made and where alternative routes could be taken."

To avoid building sidewalls and promote different imaginations to co-habit, we need a revolution in the modes of production of knowledge. We shall also be ready to question the reasons for non-production of knowledge. Here, we mean we need to consider the non-production of knowledge as a subject of scientific, historical and social inquiry. If we dare to echo Latour's question "Can we cohabitate with you? Is there a way for all of us to survive together while none of our contradictory claims, interests and passions can be eliminated?," we need to design spaces where interactions between the two cultures are pluralist, reflexive, experimental and promote mutual learning. Pluralism invites us to be open to diverse forms of knowledge and experiences, to "epistemic other," to be open to different perspectives on what "progress" means for a human society in a specific context and culture. Reflexivity leads us to question our modes of production (and non-production) of knowledge. For example, sustainability has been erected as a buzzword in some policy arenas and neglected in others. But, beyond the term, some questions are rarely asked: what is it that we want to sustain? When it comes to sustainability, we know very little. Why? Researchers in Stanford (Proctor and Schiebinger 2008) have pursued a genuine research path to show how, around some phenomenon and object such as the tobacco industry or climate change issues, our societies, consciously or unconsciously, do not produce knowledge but instead promote ignorance. A disposition to pluralist reflexivity would resonate with Stirling's interrogations (2009:5): "'which way?'; 'what alternatives?'; 'who says?' and 'why?' This is the essence of a normative, analytic, epistemic, ontological – and consequently intrinsically political – project of 'pluralising progress'."

Such pluralist and reflexive modes of sharing knowledge are difficult, demanding and, above all, experimental. They require, despite institutional pressures, being equally ready for successes and failures, having the ability to learn from failures and non-production of knowledge. They thus have to rely on processes of mutual learning where prevail openness and understanding – "hospitality" as mentioned by the manifesto. This notion of "hospitality" brings us back to the *cohabitation* metaphor and the significance of developing a language that can be shared by researchers from different fields without being left on the margins. We have shown through our exploration of engineering concepts and metaphors in synthetic biology how and why a nuanced understanding of assumptions behind concepts and metaphors might be a starting point to better promote mutual learning. The interdisciplinary experiment which led to questioning these concepts and metaphors is only a first exploratory step and should not be considered in itself as a cohabitation experiment.

Initiatives such as trading zones, lab-scale interventions, and the manifesto are seminal steps towards being able to cohabitate without eliminating each other's claims, interests and passions. However, numerous challenges remain in trying to promote *cohabitation* within and beyond the walls of life sciences. As noticed by Erik Fisher and Daan Schuurbiers (2009:426), "colleagues and peers might greet boundary-crossing attempts with suspicion, and laboratory researchers will question the value of considering the social dimensions of their work if it does not allow for more immediate practical insights. Social researchers, by contrast, might worry about becoming co-opted, being limited to more benign forms of critique or becoming 'token ethicists' used to deflect societal concerns."

Indeed, there are real costs to taking the time and effort to share knowledge across disciplinary and sectoral borders. One of the barriers is actually the world of academia itself: sharing credit and having that credit count for a career and tenure is a competitive business. Effective collaborations come from sharing and building of trust. The real issue becomes how do we step back and build trust between researchers? It is an issue of networking and cooperative behavior, which requires shifting the balance to cooperative behavior while still keeping a balance of creativity.

There are some questions that policy makers need to address on a federal level. Number one of course is how we best support science at the intersection of natural sciences and humanities, as an emerging discussion to be fostered. The second question is how the federal government should facilitate these boundary-crossing collaborations at the appropriate scales. The federal government is used to the individual investigator model. But it has to become more and more used to large scale collaborations, finding that space in-between whether it is midscale research collaboration or midscale infrastructure development. This movement should be further encouraged. Furthermore, how should the federal government, along with the rest of us, link science toward policies and progress for addressing some grand challenges that face not only the nation but also the world? Ultimately, we want the insights gained from science to help us in solving sustainability and other grand challenges from policy advances in conjunction with technological advances.

So, concretely, what can funding agencies do? They have to find ways to allow the collaborations to occur across disciplinary borders. Agencies' missions are not only to fund good science, to advance knowledge and discovery, but also to look at the vitality of the scientific enterprise in this country for the future. Science and science communities do not always act in their best interest for the long-term. They act in the best interest of the individual players at the time. It is very important that agencies step back at times and ask what is in the best long term interest of the field, the nation, and innovation. In serving that goal, governmental entities must therefore think about the kind of research activities that will be needed down the road to solve problems we are facing as a society, especially those that will require collaborative and interdisciplinary research to be solved. We must also ensure a diverse research portfolio and consider the minority interests not fully represented in the process as well as the individual researcher. The scale and scope of many problems demand an established mechanism for allowing larger collaborative projects to be developed. There is a need for thinking creatively as an organization about budget allocation, the peer review process, and how to bring people together that might not normally come together, towards fostering interdisciplinary science. There is a role for federal agencies in encouraging and sustaining research coordination networks not just within their own disciplines but particularly across many fields.

11.6 Conclusion: The Price of Metaphors

This contribution briefly depicted how, under the heading "new biology," life sciences have begun to target social problem-solving as an explicit purpose of research, thus producing imagined visions of our bio-technical futures and new challenges for governance. Eventually, through this diagnosis, we aim at unveiling the dynamics that promote the constant weaving of the life sciences with a political regime of techno-scientific promises. The overall objective is to reflect critically on who gets to imagine, anticipate, and configure human futures, as well as to reflect critically on the matters of concern that emerge in the aftermath.

The effort to think about the work of scientists and engineers in the broader social, ethical and political contexts they inhabit is often linked to the normative objective of producing more socially robust and responsible technologies. In this chapter, we suggest to focus on one approach to building collaborative practices for contextualizing science, engineering work and the engineering laboratory – *cohabitation* – with the long-term goal of assisting scientists and engineers to become more reflexively aware of the contexts within which their work take place. Ultimately, the author's main premise is about the significant role of engineering metaphors and concepts within attempts to collaborate within and beyond the walls of life sciences. Due to their interpretative force and potential influence on public visions, we need a reflective and more critical use of metaphors when it comes to scientific practice. As eloquently stated by pioneering cyberneticists Arturo Rosenblueth and Norbert Wiener, "the price of metaphor is eternal vigilance" (Lewontin 2001).

References

Baird, D., & Cohen, M. (1999). Why trade? *Perspectives on Science, 7*, 231–254.

Balmer, A., Bulpin, K., Calvert, J., Kearnes, M., Mackenzie, A., Marris, C., Martin, P., Molyneux-Hodgson, S., & Schyfter, P. (2012). *Towards a manifesto for experimental collaborations between social and natural scientists*. http://experimentalcollaborations.wordpress.com. Accessed 3 July 2012.

Bedau, M., Parke, E. C., Tangen, U., & Hantsche-Tangen, B. (2009). Social and ethical checkpoints for bottom-up synthetic biology or protocells. *Systems and Synthetic Biology, 3*(1–4), 65–75. Springer Netherlands, Dordrecht.

Boogerd, F., Bruggeman, F. J., Hofmeyr, J.-H. S., & Westerhoff, H. V. (Eds.). (2007). *Systems biology: Philosophical foundations*. Amsterdam: Elsevier.

Bowker, G., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Bromme, R. (2000). Beyond one's own perspective – The psychology of cognitive interdisciplinarity. In P. Weingart & N. Stehr (Eds.), *Practising interdisciplinarity* (pp. 115–133). Toronto: University of Toronto Press.

Brown, T. (2003). *Making truth: Metaphors in science*. Urbana-Champaign: University of Illinois.

Calvert, J., & Fujimura, J. (2009). Calculating life? A sociological perspective on systems biology. *EMBO Reports, 10*(1), 46–49. doi:10.1038/embor.2009.151

Carlson, R. (2010). *Biology is technology – The promise, peril, and new business of engineering live*. Cambridge, MA: Harvard University Press.

Collins, H. M., & Evans, R. (2002). The third wave of science studies. *Social Studies of Science, 32*, 235–296.

Fisher, E. (2007). Ethnographic invention: Probing the capacity of laboratory decisions. *NanoEthics*. doi:10.1007/s11569-007-0016-5. Springer.

Fisher, E., & Mahajan, R. L. (2006). *Midstream modulation of nanotechnology research in academic laboratory*. ASME International Mechanical Engineering Congress and Exposition (IMECE2006), Chicago, pp. 1–7.

Fisher, E., & Schuubiers, D. (2009). Lab-scale intervention. *EMBO Reports Science & Society, 10*(5), 424–427.

Fujimura, J. (2005). Postgenomic futures: Translations across the machine-nature border in systems biology. *New Genetics and Society, 24*(2), 195–226.

Galison, P. (1997). *Image and logic: A material culture of microphysics*. Chicago: The University of Chicago Press.

Galison, P. (1999). *Objectivity is romantic* (ACLS occasional paper – The humanities and the sciences, No 47, pp. 15–43). Philadelphia: ACLS.

Gibbons, M., Nowotny, H., Limoges, C., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary society*. London: Sage.

Gorman, M. E. (2004). Collaborating on convergent technologies – Education and practice. *Annals of the New York Academy of Sciences, 1013*, 1–13.

Gorman, M. E., & Mehalik, M. M. (2002). Turning good into gold: A comparative study of two environmental invention networks. *Science, Technology & Human Values, 27*, 499–529.

Gorman, M. E., Groves, J. F., & Shrager, J. (2004). Societal dimensions of nanotechnology as a trading zone: Results from a pilot project. In D. Baird, A. Nordmann, & J. Schummer (Eds.), *Discovering the nanoscale*. Amsterdam: Ios Press.

Gorman, M., Werhane, P., & Swami, N. (2009). Moral imagination, trading zones, and the role of the ethicist in nanotechnology. *NanoEthics, 3*(3), 185–195.

Greenberg, V. (1990). *Transgressive readings: The texts of Franz Kafka and Max Planck*. Ann Arbor: University of Michigan Press.

Jamison, A. (2010). In search of green knowledge: A cognitive approach to sustainable development. In S. Moore (Ed.), *Pragmatic sustainability: Theoretical and practical tools* (pp. 68–80). New York: Routledge, forthcoming. http://people.plan.aau.dk/~andy/In%20Search%20of%20Green%20Knowledge.doc. Accessed 16 June 2012.

Jasanoff, S. (2004). *States of knowledge: The co-production of science and social order*. London: Routledge Press.

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

Jasanoff, S. (2009). *Governing innovation*. Paper presented at the symposium knowledge in question – A symposium on interrogating knowledge and questioning science # 597. http://www.india-seminar.com/2009/597.htm. Accessed 16 Dec 2011.

Johnson, M. (1993). *Moral imagination: Implications of cognitive science for ethics*. Chicago: University of Chicago Press.

Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Latour, B. (2005). From Realpolitik to Dingpolitik or how to make things public. In B. Latour & P. Weibel (Eds.), *Making things public – Atmospheres of democracy* (pp. 14–43). Cambridge: MIT Press and Karlsruhe: ZKM.

Lewontin, R. C. (2001). *Science, 291*, 1263–1264.

Morreale, S. P., & Howery, C. B. (2002). Interdisciplinary collaboration: Down with the silos and up with engagement. Ohio Learning Network. http://www.oln.org/teachingandlearning/lci/lcarchive/lcresources.php. Accessed 15 June 2012.

National Research Council, Board on Life Sciences: Division on Earth and Life Studies. (2009). *A new biology for the 21st century*. Washington DC: The National Academies Press.

OTA. (1989, April). *New developments in biotechnology: Patenting life* (Special Report OTA-BA-370). Washington DC: U.S. Government Printing Office.

Pauwels, E. (2009). Review of quantitative and qualitative studies on U.S. public perceptions of synthetic biology. *Systems and Synthetic Biology, 3*(1–4), 37–46. Springer Netherlands, Dordrecht.

Pickering, A. (1995). *The mangle of practice: Time agency and science*. Chicago: University of Chicago Press.

Proctor, R. N., & Schiebinger, L. (Eds.). (2008). *Agnotology: The making and unmaking of ignorance*. Palo Alto: Stanford University Press.

Rabinow, P., & Bennett, G. (2009). Synthetic biology: Ethical ramifications. *Systems and Synthetic Biology, 3*, 99–108. Springer.

Rabinow, P., & Bennett, G. (2012). *Designing human practices*. Chicago: The University of Chicago Press.

Rose, N., & Novas, C. (2005). Biological citizenship. In A. Ong & S. Collier (Eds.), *Global assemblages: Technology, politics and ethics as anthropological problems* (pp. 439–463). Oxford: Blackwell.

Schot, J. W., & Rip, A. (1997). The past and future of constructive technology assessment. *Technological Forecasting and Social Change, 54*(2/3), 251–268.

Sethi, L. M., & Briggle, A. (2011). Making stories visible: The task for bioethics commissions. *Issues in Science and Technology, 28*, 29–44.

Smith, A., & Stirling, A. (2008). *Socio-ecological resilience and socio-technical transitions: Critical issues for sustainability governance* (STEPS Working Paper 8). Brighton: STEPS Centre.

Specter, M. (2009, September 28). A life of its own – Where will synthetic biology lead us? *The New Yorker*. http://www.newyorker.com/reporting/2009/09/28/090928fa_fact_specter. Accessed 10 June 2012

Stirling, A. (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.

Stirling, A. (2009). *Direction, distribution and diversity! Pluralising progress in innovation, sustainability and development* (STEPS Working Paper 32). Brighton: STEPS Centre.

Venter, J. C. (2007). *A life decoded – My genome: My life*. New York: Viking Penguin.

Werhane, P. H. (1999). Justice and trust. *Journal of Business Ethics, 21*(2–3), 237–249.

Wynne, B. (2009). *Daring to imagine*. Paper presented at the symposium knowledge in question – A symposium on interrogating knowledge and questioning science # 597. http://www.india-seminar.com/2009/597/597_brian_wynne.htm. Accessed 15 Jan 2012.

Wynne B., Callon, M., Eduarda Gonçalves, M., Jasanoff, S., Jespen, M., Joly, P.-B., Konopasek, Z., May, S., Neubauer, C., Rip, A., Siune, K., Stirling, A., & Tallachini, M. (2007). *Taking European knowledge society seriously. European Commission Report of the Independent Expert Group on Science and Governance* (EUR 22700). Luxembourg: European Commission.

Part IV
Conclusions

Chapter 12
Early Engagement and New Technologies: Towards Comprehensive Technology Engagement?

Neelke Doorn, Daan Schuurbiers, Ibo van de Poel, and Michael E. Gorman

Abstract In this concluding chapter of the volume "Early engagement and new technologies: Opening up the laboratory," the editors take stock of the different engagement methods that have been described in this volume, i.e., Constructive Technology Assessment (CTA), Value Sensitive Design (VSD), Socio-Technical Integration Research (STIR), Network Approach for Moral Evaluation (NAME), and Political Technology Assessment (PTA). The methods are compared with respect to their aims and impacts, their type of intervention, the level and phase of technological development, and their normativity. The editors develop a tentative framework for Comprehensive Technology Engagement, identifying the major questions and challenges to advancing interdisciplinary engagements at early stages

---

N. Doorn (✉) • I.R. van de Poel
Department of Technology, Policy and Management, Delft University of Technology,
P.O. Box 5015, 2600 GA Delft, The Netherlands
e-mail: n.doorn@tudelft.nl; i.r.vandepoel@tudelft.nl

D. Schuurbiers
De Proeffabriek (The Pilot Plant), Lookwatering 36, 2614 KA Delft, The Netherlands
e-mail: daan@proeffabriek.nl

M.E. Gorman
Department of Science, Technology & Society, University of Virginia,
351 McCormick Road, Charlottesville, VA 22904–4744, USA
e-mail: meg3c@virginia.edu

N. Doorn et al. (eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*,
Philosophy of Engineering and Technology 16, DOI 10.1007/978-94-007-7844-3_12,

of technological development. Three theoretical and methodological challenges are addressed: (1) issues of normativity and representation, (2) moving across levels and time, and (3) dealing with uncertainty and indeterminacy.

12.1 Introduction

The ultimate aim of technology engagement might be formulated as the better attuning of technological development to societal needs and values. What social needs are and which values are paramount, and how R&D is to be best attuned to these is not beyond dispute. Nevertheless, in this concluding chapter we will take up these broader questions.

We start with evaluating and comparing the various approaches to technology engagement that were discussed in Part II: Constructive Technology Assessment (CTA), Value Sensitive Design (VSD), Socio-Technical Integration Research (STIR), Network Approach for Moral Evaluation (NAME), and Political Technology Assessment (PTA). We will compare these approaches with respect to their aims and reported impacts, the type of intervention, the level and phase of technological development in which they operate, and their normative background.

After comparing the approaches, we will reflect on the question in what ways these approaches collectively contribute to elucidating or enhancing the role of science and technology in society. We discuss theoretical and methodological challenges for technology engagement: how to move across levels and time, issues of normativity and representation, and the challenge of dealing with ignorance and indeterminacy.

In reflecting on these broader questions, we develop a tentative framework for Comprehensive Technology Engagement (CTE). We end with discussing practical challenges for CTE and a proposal for a center for CTE. The views in the conclusions are those of the editors and do not necessarily reflect those of the contributing authors. In order to be able to critically compare the different approaches, we have stressed the differences between the methods more than their similarities. This may prompt further debate, which will hopefully contribute to the advancement of the field.

12.2 A Comparison Between Approaches

Although there are many similarities between the approaches to early engagement discussed in Part II, there are also distinct differences. To bring these similarities and differences to the fore, we start this concluding chapter with a comparison between the methods. Our aim is not to judge which method is best. Rather, we want to show the diversity of methods and their complementariness. What approach is most appropriate will in general depend on one's aims and the specific contexts. We would like to suggest that what is called for often is a combination of the approaches discussed in Part II of this volume. This raises the broader question of how to organize comprehensive forms of technology engagement, to which we turn in the second part of this chapter.

We will compare the approaches presented in Part II of the volume in four respects. First, we will look at the aims mentioned by the authors and the reported impacts of the approaches in the cases in which they have been applied (Sect. 12.2.1). Then we will look at the intervention prototype in each of the methods (Sect. 12.2.2), the level and phase of technological development in which the approaches are usually applied (Sect. 12.2.3), and, finally, the normative stance taken in each of the methods (Sect. 12.2.4).

12.2.1 Aims and Impacts

To start with CTA (Chap. 3), its general aim partly stems from the drawbacks of traditional TA approaches and is usually formulated in terms of bridging the gap between innovation and traditional measures to "control" technology development. The "constructive" aspect of CTA lies in its aim to broaden technological design and development and make it more reflexive.

Regarding the impact of their intervention activities, Rip and Robinson mention, with qualification, increased reflexivity in the co-evolution of nanotechnology in society, one of the technologies they focus on. They also mention some institutionalization of scenario and strategy workshops. In that sense, there may be reason for (modest) optimism in terms of its achievements. But the question remains whether insertion itself is sufficient to address all three layers they discuss, and whether insertion is the only suitable method. Rip and Robinson present insertion as an integral part of the CTA activities: without insertion CTA would not be effective.

The goal of VSD (Chap. 4) is to build values into the design of technology itself. The case studies presented by Friedman et al. indicate that VSD can be quite successful in doing so. It seems likely that the success of VSD will depend on the willingness of the designers of a system themselves to take values into account. Moreover, VSD might be more difficult to apply to technologies that are still in earlier phases of development than the design phase.

The goal of STIR (Chap. 5) is to elucidate and possibly enhance modulations at the midstream by using a protocol that focuses on opportunities, considerations, alternatives, and outcomes. This protocol opens up new possibilities for technical research as well as a deeper consideration of ethics in the laboratory and of the impact the research will have on users and stakeholders. Once STIR has passed from reflexive to deliberate, routine decisions may be altered so as to respond to a broader array of concerns, which may ideally enhance the responsiveness of science to its societal context. The number of STIR studies is growing and includes various interventions that have documented the step from *de facto* to reflexive and deliberate modulation.

One provocative finding is that scientists who work with STIR researchers have reported on at least two occasions that the engagement proved valuable for their work. Gorman's chapter (Chap. 8) follows several of these STIR engagement experiences in detail and provides a framework and methods for documenting them at greater depth. Fisher and Schuurbiers themselves are modest in their claims. They argue that it is hard to present the success of STIR interventions in terms of a dichotomous pair of metrics. Success may depend on a range of conditions, including the voluntary engagement of the scientific and engineering practitioners to participate in the STIR studies in the first place and their own perceptions of the appropriate course of action in different situations.

The goal of NAME (Chap. 6), as described by Van de Poel and Doorn, is formulated as a threefold objective: (1) identifying moral issues in R&D networks, (2) ethical reflection and judgment on these issues, and (3) distributing the responsibilities for addressing these issues. Van de Poel and Doorn discuss two applications of the NAME approach. The tangible outcomes of the interventions in the studies is a (re)distribution of responsibilities related to the societal impact of the technology at hand and the inclusion of some moral issues that were at first overlooked by the technological researchers. The strength of the NAME approach seems its focus on distributing responsibilities. A possible limitation of the method is the commitment required from technological researchers. As the technological researchers allow intervening actors in their midst on a voluntary basis, these "visitors" do not have any means to secure the cooperation of the technological researchers, nor do they have a commitment that the effects of their interventions will be carried out or sustained. This is a challenge that NAME shares with STIR (and to a lesser degree with CTA/insertion).

PTA (Chap. 7), finally, aims at engaging political actors with R&D and research activities. The target group of PTA is thus different than those of the other four methods, which primarily aim at engaging technological researchers, scientists and designers. As a consequence, PTA, unlike the other methods, does not intervene in research activities (cf. Sect. 12.2.2). Van Est's description of political TA suggests that, in the end, communicative skills and the capacity to create an atmosphere of trust are more important than any method in securing positive impacts, though methods are important in documenting results and improving engagement practices. These are lessons that probably hold for any engagement approach.

12.2.2 Types of Intervention

We will now look at the types of intervention undertaken in the various approaches. Interventions vary in two respects: with respect to the object of intervention and with respect to the roles of the engagement actors. Possible objects of intervention are (1) the technology being researched, developed, or designed; (2) the network of actors in which the technology is researched, developed, or designed; and (3) the researcher or designer.

Although all approaches in the end aim at influencing the technology at hand, only in VSD the technology itself is the main target. Still, VSD also aims at influencing the network by involving more stakeholders, and it is clearly also targeted at changing the mindset of the designer by making her aware of the importance of values. Both CTA and NAME are primarily aimed at influencing the network, although in both cases the idea is that this will in the end change the technology. The individual gets somewhat less attention in these approaches. Conversely, STIR is primarily aimed at the individual level, although it may also be applied to networks and organizations, and, again, the idea is that intervention eventually will change the technology developed.

With respect to the role of the engagement actors, one might distinguish between the following modes: (1) embedded, (2) parallel, (3) temporarily inserted, and (4) joint (cooperative). Embedding the engagement actors typically happens in STIR, where the engagement actors participate on the lab floor for a considerable period of time and more or less become one of the participants. In NAME, the role of the engagement actors should be seen as parallel. Although the engagement actors feed the results back to the researchers in the R&D network and get their information from them, the NAME engagement actors for a large part do their own work parallel to the technological research. CTA, as we have seen in the chapter by Rip and Robinson, is characterized by insertion. This seems to be a mode of intervention that is more temporary and short-lived than the embedding of the STIR engagement actors; yet the engagement actors seem to become more a participant than in the parallel approach in NAME. In VSD finally, the engagement actors becomes more or less part of the design team in a cooperative effort. This might look a bit like the role of the embedded STIR engagement actors; however, the VSD engagement actors have more a specific role that is from the beginning recognized as important by the technological researchers or designers. Moreover, the VSD approach could in principle also be applied by technological researchers themselves, which allows for the absence of the engagement actor because the technological researchers themselves take on board values and involve stakeholders.

12.2.3 Levels and Phases of Technological Development

As Rip and Robinson emphasize in their contribution, TA activities take place on different "floors" or levels of technological development as we will call them. Although, again, the various approaches might be applied at different levels, they are typically more aimed or more appropriate at one level than another. We propose to distinguish here between the following five levels:

1. The individual
2. The project
3. The organization

4. The technological sector
5. Society

With the project we mean the temporary network in which a technology is designed, developed, or researched within a limited time frame and, usually, with specific aims.[1] The organization refers to a unit that has a longer duration than a project and is formally and often hierarchically organized. Typical examples that are relevant here are companies, universities, and research laboratories. With the technological sector, we mean the amalgam of organizations, projects, and individuals that is working on a particular technology like cars or packaging. Depending on how technological sectors are precisely delineated, various technological sectors may be embedded in each other or intersect.

If we look at the various methods discussed in Part II, the following picture arises. As VSD is primarily aimed at the technology at hand, it is usually applied in specific design projects. Although VSD might also be adopted by companies (organizations) or in an entire technological sector, currently it seems to be mainly applied at the project level.

The strength of CTA lies in its comprehensive view on technology development: from the ongoing practices in the laboratory to the broad range of activities related to public policy, regulation, and societal debate. Compared to the other approaches, CTA seems especially apt for addressing issues at the level of the technological sector, rather than the project or organization. As Rip and Robinson speak about "moving about" in the world of the technology at stake, they transcend the world of the individual researcher or project. Given the underlying assumption of CTA (viz. that science, technology, and society co-evolve), this focus is not surprising.

STIR inserts participant observers into one CTA floor: the R&D laboratory, therewith combining the observer/assessor aspect of insertion with the Vision Assessment goal of getting the researchers themselves involved in pushing their lab work in the right direction—including more sustainable procedures, considerations of the values of users, and opening up new possibilities for research directions.

Given that the practitioners engage in this voluntary STIR endeavor, the STIR approach may be an effective means to achieve change on the individual and laboratory levels. In line with Rip and Robinson, Fisher and Schuurbiers argue that the method probably requires further expansion into policy and public engagement domains to sustain longer-term capacities for reflexive co-evolution.

The NAME approach is currently applied at the project level. Since the approach focuses on networks, it could in principle also be applied to broader networks, such as technological sectors (cf. Van de Poel 2008).

The level of society as a whole is partly addressed in Political TA, which focuses on the policy and political level. Since the results of "TA projects need to be communicated across the border, from the scientific towards the political sphere" (Van Est, Chap. 7), PTA forms an indispensable element of a comprehensive

---

[1] Projects might be inter-organizational, so that the levels distinguished here are not strictly embedded in each other.

approach. Political TA should therefore not be seen as an alternative to one of the other engagement methods, but rather as a necessary complement.

12.2.3.1 Phases

In addition to various levels, we might also distinguish between various phases of technological development. All methods discussed in this volume address "technology in the making" (as opposed to "ready-made technologies" embodied in marketable products or processes). Still, we might distinguish between differences phases of making technology and we propose a distinction between the following phases: (1) Fundamental research (2) Research and development (R&D), (3) Innovation, and (4) Design.[2]

VSD is clearly most relevant in design, while STIR is directed at R&D and partly also fundamental research as it focuses on work in research laboratories and at universities. Both CTA and NAME seem to primarily focus on innovation, i.e. new technology, although NAME does so more on the project level and CTA at the level of an entire technological sector. PTA, finally, does not seem to focus on one specific phase of technological development

Figure 12.1 summarizes how the different methods each have a different focus, although most of them can also be applied at other levels or in other phases.

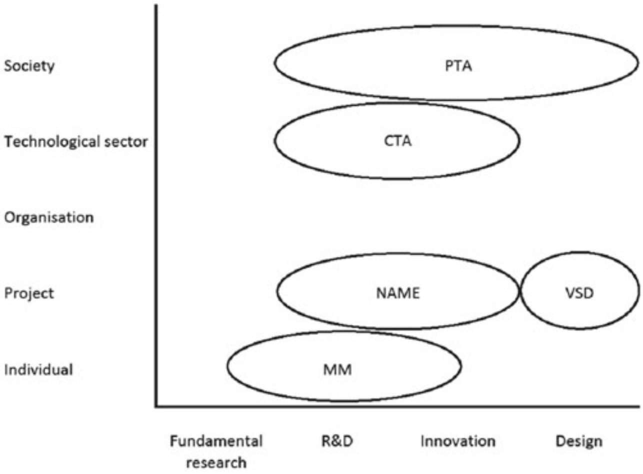


Fig. 12.1 Overview of different approaches in terms of levels and phases

---

[2] Although the term "phase" suggests a certain sequence, we would like to stress that these phases are not necessarily sequential or linear. Still, these four types seem to describe the main sets of activities that characterize technology in the making (although one might distinguish additional ones like testing or experimenting).

12.2.4 Normativity

We will now look at how the approaches deal with issues of normativity. Normativity is probably one of the most controversial topics with respect to engagement methods. A first question is whether engagement are, or should be, normative or descriptive. This issue is complicated by the fact that one might disagree about what is exactly "normative" and "descriptive." Is a method that describes what happens and feeds this back to researchers, so possibly enlarging their reflexivity, only descriptive or also normative because it contributes to reflexivity? And if one is to accept that the methods have a normative component, the question is what this component is or should be.

From the description in Part II, it seems that PTA and STIR are not explicitly normative. Unlike the other methods, PTA is not aimed at intervening in technological development. If PTA has a normative goal, it appears to be something like better informing politicians and engaging them more with technological issues. PTA might thus be said to contribute to enhancing the reflexivity of policy actors.

Similarly, Fisher and Schuurbiers do not mention any evaluative norms other than making technological researchers and engineers become "more reflexively aware of socio-ethical contexts, to alter their decision processes, and to change the nature and direction of their work in light of the collaborative inquiry" (Fisher and Schuurbiers, Chap. 5).

In an earlier publication on CTA, Schot and Rip (1997) distinguished three normative goals for CTA: (1) anticipation, (2) learning, and (3) reflexivity. These norms seem to lie on the network level.

The normative aim of VSD might be understood as building values of ethical importance into the technology designed. This focuses on the technology rather than the network. This does not mean, however, that VSD does not have normative aims at the network level. In fact, it aims at including more stakeholders, or at least their views, in the network, and also at more reflexivity of individual designers.

NAME explicitly mentions reflexivity and "openness and inclusiveness" as procedural norms. These norm also seem to play an implicit role in the other approaches. VSD, for example, stresses the importance of taking into account the values of various stakeholders; CTA speaks about broadening technological design and innovation, and STIR wants to make technological researchers more aware of appraisals of technology in the rest of society. The NAME approach further builds on the idea of moral pluralism. In the NAME framework, recognizing the plurality of moral views may be a way to avoid a moralizing stance and to secure a cooperative attitude by the technological researchers. Given that NAME is primarily procedural in nature, the method may be most suitable in situations where moral pluralism is indeed an issue. This way, a balance can be struck between a top-down approach in which the engineers and technological researchers have little to say about which "moral issues" to address in the project and an approach in which the individual people only address those issues which happen to attract their individual interest. As Van de Poel and Doorn indicate, their cases suggest that confrontation with dissenting opinions on relevant moral issues may already prompt sufficient discussion to raise awareness of these issues. It is then, within the bounds of reasonable pluralism, up to the relevant actors and stakeholders to assess which issues fall within the scope of the project and consequently, how to address these issues (Doorn 2012).

12.3 Towards Comprehensive Technology Engagement (CTE)

The previous section compared various approaches to technology engagement. The question that remains is to what extent this collection of approaches as a whole speaks to the challenges and motivations for engagement identified in the introduction. In a way, each of the approaches identifies *necessary* conditions for addressing those broader challenges (insertion of CTA actors on different floors; integration of values in design processes; modulating the decisions made at the midstream of R&D; encouraging reflection on research; etc.). Are they collectively *sufficient* conditions for early engagement with science and technology? And if not, what would still be needed to address broader visions of "better technology in a better society?" (Schot and Rip 1997). Of course, what "better" means in this case is at least partly a matter of political taste and preference, and the different approaches within this volume themselves have different normative stances. Does "better" imply a more open or inclusive R&D process? More reflexive practitioners? A more transparent decision-making process? Or simply a faster rate of innovation? Visions of the role of science and technology are characterized by normative plurality, even among engagement actors.

Still, we might ask: given the variety of roles (Calvert and Martin 2009) and objectives of engagement practitioners, and the differences between approaches and intended and unintended outcomes, what binds them? We introduce the notion of *comprehensive technology engagement* (CTE) as a locus for theoretical and practical reflections on the diversity of engagement activities. CTE is not another method, nor is it intended to be an overarching concept in itself; it merely aims to bring under one header the range of questions on the practice and theory of engagement as a whole. This includes the approaches discussed in this volume, but also those that operate on broader policy or public levels (cf. Swierstra et al. 2009; Verbeek 2011). CTE thus embraces diversity with respect to the motivations for, stances towards and intended outcomes of intervention and engagement, but identifies that the commitment to *engage* itself at some level implies a shared normative commitment to enhance the quality of the relations between science and technology in society. Note that the focus is on *engagement*; assessment implies evaluation, whereas engagement emphasizes an equal relationship where the parties work together. Engagement does not replace the need for assessment in certain situations. Instead, engagement is a strategy that promotes the parties involved to reflect on the emergent implications of the work they are doing. The goal of engagement is co-creation of solutions and ideas.

The core assumption of CTE is that engagement across representations, floors, and times will not only produce a better understanding of how technologies emerge but will also lead to more responsible development because the CTE process will encourage democratic reflection and dialogue. The ongoing dialogues will probably have to include ethicists as a matter of course, who can maintain a focus on normative implications. The end result is unlikely to be a consensus; instead, the dialogue will be continuous and often heated as different values and assumptions clash. One of the benefits of CTE will be a clearer picture of how technologies become socio-technical systems, and where the tipping points occur that make these systems harder to change.

The concept of CTE thus brings forth a new set of questions and challenges both on a theoretical and practical level. The following sections go into these questions in more detail and suggest tentative answers.

12.4 Theoretical and Methodological Challenges for CTE

On a methodological level, the central question for CTE is how to enable comparisons between the variety of approaches with respect to both the motivations for engagement and the methods employed to achieve these objectives. Can we map the varying normative commitments, and possibly identify common denominators between sets of approaches? Comprehensive Technology Engagement enquires into the comparability and potential compatibility of data collection methods. Multiple approaches to technology engagement are good—diversity is a real strength—but it is important that the differences be made explicit and that methods for collecting comparative data be employed. Gorman (Chap. 8) provides some suggestions like having technology assessors use an evolving standardized set of prompts to keep diaries that can be disclosed to others doing TAs in the same system; keep a kind of chronological flowchart of activities they observe; note reactions to the TA process and have independent researchers who can conduct occasional interviews and administer surveys to determine how participants are responding; and feeding results of the on-going assessment back into workshop discussions with members of the laboratory or network or agency, making it a tool for continuous improvement. Participants in the workshops can also help think of other important indicators to track; evolve descriptive, visual models of the way the multi-floor assessment evolves, focusing especially on the roles of the different Technology Assessors and how they inform each other. These models could be shared among the assessors and revised based on feedback; they have heuristic value in provoking deeper reflection on processes.

Having established methods for comparison, a follow-up question is whether we can establish metrics for success, comparing which methods are most appropriate to achieving certain objectives. When do we conclude an engagement was successful? Can we be more specific in advance about what the engagement is to achieve? Should the technical work be more reflexive or transparent? Should the team be more inclusive or show signs of higher order reflection? Or should the success be measured in terms of the uptake of methods or, less ambitious, on the question whether or not the embedded researcher is "not being kicked out." These are different measures for success but they are hard to operationalize *a priori*. Once again, the focus should be on the ability to compare and share—not only methods, but also ways of measuring success.

12.4.1 Moving Across Levels and Across Time

One major question with respect to methodology is how to move findings across levels of technological development. Rip and Robinson make an important observation in Chap. 3: while engagement at the bottom layer is important, the other layers are still needed. The CTA model might in fact provide the overview capability on a technological frontier because CTA activities work on multiple "floors," including research laboratories, conferences, planning events and public debates. While this perspective sits well with the notion of CTE in theory, there is much methodological ground to be covered: we have not even begun to connect findings at the level of laboratory interventions with those, say, in the policy room, the science cafe, or the production floor. How to build synergies across engagements on multiple levels?

Figure 12.1 also points out that the methods hitherto discussed have shown relatively little attention to the level of organizations. As a consequence, a significant driver of socio-technical change is also partly ignored by the approaches: entrepreneurial innovations and how they lead to industrial production. STS scholars have engaged with businesses (Woolgar et al. 2009) though not with a CTE focus. There have been some STIR forays into the production floor (e.g. Flipse et al. 2013). Also, much work has been done in commercial contexts under the broader header of Corporate Social Responsibility (CSR). As a central stage in the innovation cycle, this area deserves further attention.

One of the major problems is intellectual property (IP): whoever engages with entrepreneurs, for example, will have to sign non-disclosure agreements, inhibiting dialogue with other CTE engagement activities. Science laboratories are becoming the catalysts for new companies, so the commercial and the scientific are increasingly intertwined. CTE could provide an actor's perspective on this process without revealing details of a specific technology. Eventually, after the IP issues have been clarified, the actors might agree to have the technological details disclosed.

It is essential to insert ethicists and social scientists into multiple floors. A few would move across and among floors, making contact with those engagement actors who were working on a single floor—in a particular research network, at a regulatory or funding[3] agency, or with parliament (like the Rathenau Institute). It would be essential to embed some engagement actors in businesses navigating this new frontier. There would be modulators working in specific R&D laboratories.

It would be productive to have the NAME approach applied to laboratories where a STIR participant was already present, and have someone doing a multi-level CTA dropping in from time-to-time to put the work in context with what was going on in governance and oversight. Similarly, a VSD could be done without embedding or insertion. Combining methods could serve as a check against one of the dangers of becoming an embedded technology assessor "going native." CTE actors would adopt different roles in these floors, and would be working at different stages of the research and development process. The key would be for these actors to be able to communicate with one another, sharing ideas, frustrations, and findings.

12.4.1.1 Moving Across Time

The general question of how to meaningfully connect various initiatives also applies to those studies that investigate the possible futures of technological developments as well as their past. Much work has been done on futures. Grunwald and Achternbosh (Chap. 2), for example, place especial emphasis on upstream TA, using techniques like Vision Assessment (VA), which focuses on eliciting and assessing visions of a new technological frontier and can include prospective life-cycle and market analyses. Similar work has been done at CNS-ASU (e.g., Selin 2007) and elsewhere (e.g., Lucivero et al. 2011). VA includes a values component. Instead of opening up the lab, this upstream TA expects the researchers themselves to push the technology in the right direction by creating guidelines working with the inserted TA actor.

A CTE could be used to take a snapshot of the current state of a technological frontier like nanotechnology or the convergence of nano, bio and ICT technologies. An important first step would be to conduct a biography of developments that led to the current state, for example, how and why did the National Nanotechnology Initiative come into being? This biography should pay particular attention to the ethical issues likely to emerge on this frontier. Initial scenarios could be generated to anticipate possible outcomes. As also emphasized by Van der Burg (Chap. 10), new technologies are partly dependent on their technological ancestors. A necessary component of Comprehensive Technology Engagement would be the inclusion of a historian of technology who could write a biography of the emergent technologies, drawing on the relevant STS literature. This historical aspect runs across the different floors and should also be coordinated with TA activities at the other levels, not the least the policy and political domain (e.g., Political TA, Chap. 7). The historical

---

[3] Gorman worked for 2 years as a Program Director at the US NSF. He was an actor, not a modulator, but he reflected on his experience in a way consistent with a modulator (see (Gorman 2011) for some reflections on this work).

approach would be particularly important for the entrepreneurial and industrial aspects of CTE, where details of the technologies that had to be legally protected initially can become part of the public record after patents are in place and some trade secrets are no longer relevant because they have been reverse-engineered.

12.4.2 Questions of Normativity and Representation

It is obvious that engagement approaches have a descriptive component. In addition, they may be characterized by certain implicit and explicit normative aims. As we have seen in Sect. 12.2.4, this is more controversial. Our stance is that some form of normativity is unavoidable and indeed desirable, although one might disagree about the exact normative aims of engagement.

Although the emphasis in engagement activities is primarily on enhancing reflexivity, we believe the normative aims of CTE might be broader than this. One important mandate for CTE is offered by responsible innovation (RI). As stressed by for example Von Schomberg (2012), RI has a product and a process dimension. The product dimension relates to certain normative standards a product should meet. This includes addressing risks, but also taking into account certain values (Directorate-General for Research and Innovation EU 2012). One might, for example, think of the basic values mentioned in the EU charter and the EU Treaty like dignity, justice, freedom, equality, solidarity, and citizens' rights (ibid.). The process dimension relates to normative standards that the process of R&D should meet. Procedural norms that have been mentioned in relation to RI include inclusion of stakeholders, accountability and transparency (Von Schomberg 2011, 2012; Directorate-General for Research and Innovation EU 2012).

The current approaches to CTE go some way in meeting these normative challenges. VSD may be especially appropriate to address the product dimension of CTE and RI. The other approaches (CTA, STIR, NAME) are useful for addressing the procedural dimension, especially for norms like reflexivity and inclusiveness. However, procedural norms like transparency and accountability are not addressed, at least not explicitly, in the currently available methods. Here lies an interesting challenge for CTE.

In relation to the norm of inclusiveness, one particularly important topic for CTE is the question of representation: who do engagement agents represent? There is considerable diversity within the background knowledge that different disciplines bring in (ethicists, philosophers, sociologists, anthropologists), but more importantly: what is the democratic legitimacy of the claims of engagement agents? Are they in fact the types of concerns that most deserve to be brought to the attention? Do they in any way represent the concerns of broader public? It is essential to integrate input from relevant stakeholders before a new system is locked in. The challenge is to find ways of integrating the findings of public perceptions studies, focus groups, citizen's cafes, surveys coupled with deliberations, public dialogues (such as is being done by the two Centers for Nanotechnology in Society at Arizona State

University and the University of California Santa Barbara and in various other places throughout Europe) with the engagement activities that are taking place at different floors. Public value mapping (Bozeman 2003) could provide another important clue to addressing this challenge.

One important development over the past few years with respect to technology engagement is the increased involvement of ethicists. As Gorman et al. (2009) note: “The role of an ethicist is to ask normative questions: What is the value of this project? What are the possible positive and negative outcomes? Are there any possibilities that have not been explored? How will the success of the project contribute to human or even planetary well-being, and what are the possible dangers involved? Moreover, well-trained ethicists can often step back from the particular context to take a more disengaged perspective on the technology or technologies in question” (Gorman et al. 2009, 186). As the contribution by Van de Poel and Doorn (Chap. 6) shows, the role of the ethicist in relation to such normative questions is not to answer them on behalf of society but rather to organize a process that does justice to the plurality of moral opinions.

12.4.3 Dealing with Uncertainty and Indeterminacy

Schot and Rip have described the overall TA philosophy as: “to reduce the human costs of trial and error learning in society’s handling of new technologies, and to do so by anticipating potential impacts and feeding these insights back into decision making, and into actors’ strategies” (Schot and Rip 1997, 251). This quote stresses the importance of anticipation, which has indeed been a main goal of TA, and plays an important role in technology engagement as well.

As we saw in Chaps. 1 and 2, the initial idea in TA was that it would be possible to objectively assess the social impacts of technology. This idea has now long been left behind and anticipation can take many other forms than prediction; it can involve promises and expectations, scenarios, technology roadmapping, vision assessment, deliberation, and the like. Still the idea of anticipation is not entirely unproblematic. The focus on anticipation may itself undermine the ability to deal with unexpected developments and surprises (Wildavsky 1988).

It is important to understand that the uncertainty that is inherent in anticipating the future is not only that we do not know what the chance (probability) is that a certain scenario will materialize, but that it also concerns what might be called ignorance, the possibility that something quite unexpected, a surprise, might occur.[4] Developments about which we are ignorant might be very hard, if not impossible, to anticipate. The uncertainty inherent in anticipating the future may also take the form of indeterminacy. We speak of indeterminacy if the causal chains towards the future are open. This is usually the case with technology-in-the-making, as the eventual shape and social consequences will depend on actions and decisions of

---

[4] For discussions of different types of uncertainty, see, for example Wynne (1992) or Renn (2005).

actors downstream. Indeterminacy also has important normative implications, as it places part of the responsibility for the future use and effects on the shoulders of actors downstream.

Dealing with ignorance and indeterminacy is thus a challenge for CTE. The procedural emphasis in many engagement methods might be helpful in dealing with this challenge. It helps to avoid a reification of technology, and maintains openness to the future. What is often called implementation of technology is in fact a process of mutual adaption of a technology and its social environment (Leonard-Barton 1988). Technology and society co-evolve (e.g. Rip and Kemp 1998). Technologies are thus not finished when they leave the lab or a company; they are still in the making.

What is important here is the awareness that engagement activities need to continue after products have left the lab or design studio. In the midstream terminology, engagement needs to continue downstream. Indeed many engagement activities originally started downstream. However, if one takes seriously the co-evolution of technology and society, the distinction between midstream and downstream itself becomes an object of study. For example, a midstream laboratory project may be an effort to create a new technology, which would be upstream in the product cycle.

One way to overcome the distinction between midstream and downstream is to view the introduction of technology into society as a (social) experiment (Krohn and Weyer 1994; Van de Poel 2009). In this view, learning about technology, and its shaping, continues after it has left the lab and entered society. The technology as social experiment approach might thus be a way to deal with the ignorance and indeterminacy that is inherent in technological development. It also raises interesting new normative questions, especially about the conditions under which experiments with a certain technology in society are acceptable.

12.5 Practical Challenges for CTE

Addressing the theoretical and methodological questions identified above will be crucial to advance our understanding of what engagement means and what it might achieve. But there are also important considerations of a more practical nature: first of all, how to enable discussion of these broader themes among the variety of engagement agents? How to prevent compartmentalization of these discussions and enable cross-disciplinary dialogues? And, perhaps more importantly, how to engage broader technological and policy actors in these discussions?

One important consideration is how to ensure that the outcomes of research projects find their way towards larger audiences of practitioners. No matter what our objective is, a fundamental question for each engagement activity is: how will it help shape the technoscientific frontier? It is clear that engagement initiatives should reach large audiences if they are to have a lasting impact on the way R&D is carried out. To address a broader range of actors, perhaps some of the work should move from "research-mode" into "application-mode." Currently, research findings are published largely in peer-reviewed journals; not necessarily the right outlet for reaching practitioners beyond one's own research community. An academic tendency is to distinguish, demarcate, separate, reinvent – but the outside world is looking for an overarching theme to relate to. How to work towards achieving this?

Another consideration is the asymmetry in power relations between intervening researchers and the technical researchers, which may have a hampering effect on the engagement interventions. As Calvert showed in her contribution, it may be difficult for an intervening researcher to find the right in the project or community (to feel "at home"). Engagement actors, whether from the social sciences or the humanities, are often dependent on the willingness of the technical researchers to allow them in their midst. Especially when involved in projects with industrial partners, intervening researchers may have to obey strict reporting rules related to confidentiality.

With engagement (be it as embedded, parallel, inserted, or cooperative researcher) comes a tension between keeping one's independence and becoming part of the team, with the ultimate risk of being compromised. This point, addressed by several of the contributions, is actually not new. Already in the eighteenth century, Immanuel Kant (1795) argued that philosophers should not become kings because of the danger of becoming corrupted by power and it has ever since been a point of concern in interdisciplinary engagements. This suggests that the intervening researcher should on the one hand become an insider, but that she should maintain her critical stance (Van de Poel 2008: 36; see also Schuurbiers 2011; Doorn and Nihlén Fahlquist 2010). Van de Poel argues that it requires a combination of personal skills and institutional safeguards to deal with this challenge. Dependent on the level of intervention but also the aim of the intervention, the appropriate organizational structure has to be developed (as also discussed by Van Est in Chap. 8). At the personal level, the skills and expertise of the intervening researcher may create room to be critical. Rip and Robinson (Chap. 3) argue that the intervening researcher should be recognized as knowledgeable, both in her own field (i.e., the social sciences or humanities) and with respect to the technology at stake. Only then will the interaction between the intervening researcher and the technical researchers be a genuine two-way exchange (Cf. Fisher and Schuurbiers, Chap. 5).

Still another concern is the lasting effect on the longer term. Many of the projects described in this volume are still dependent on personal relationships between intervening researchers and members of particular research groups. One way to deal with this challenge would be the establishment of a center for CTE that may institutionalize the interventions, and therewith establish a formal commitment and give the intervening researchers a more profound place in a team or a project. However, too much formalization may also create obstacles for engagement studies. The trading zones, where all partners benefit mutually, may be a fruitful starting place for establishing more symmetrical relationships.

12.5.1 A Center for CTE?

Centers for Nanotechnology in Society have been created by the United States National Science Foundation at Arizona State University and the University of

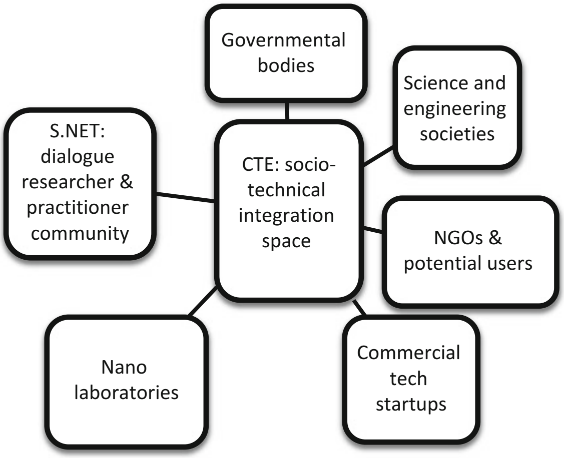


Fig. 12.2 A hypothetical CTE center with connections to several of the relevant locations and floors

California at Santa Barbara to do research and provide guidance on topics like anticipatory governance of emerging technologies and public engagement—not just in the US, but world-wide. To achieve the kind of Comprehensive Technology Engagement described here, perhaps a Center should be established that could coordinate multiple floor assessments, convene workshops, identify best practices, connect with experts on research methods and consult with governments, agencies and research networks that want to anticipate, monitor and adaptively manage emerging socio-technical systems (Gorman et al. 2008). Eventually, such a Center could possibly obtain funding not only from governments but also from clients seeking this kind of assessment and engagement because it creates more opportunities for breakthroughs that will be beneficial to users and stakeholders. Here the Center would have to be aware of one of the dangers of ELSIfication—of being assigned a marginal role. Technology Assessors would have to avoid the "impact" model and demonstrate that science and technology are in society, and the Assessors are part of that society, working closely on multiple floors to engage participants and stakeholders in deep, constructive reflection. Such a Center would not have to start from scratch. This volume illustrates the diversity of engagement activities already under way, and there are many more not covered here. So a Center for CTE might begin as a trading zone, convening a wide range of TA actors from multiple stages and levels, with the goal of establishing a creole among them so they could exchange best practices. The Center could then provide information and contacts for those looking to incorporate an aspect of TA into a project.

Figure 12.2 shows a simplified version of a CTE center and some of the floors and locations it would connect to. S.NET is included as a good space for the STS scholars doing engagement to convene and discuss what they are learning with colleagues—which should include at least some scientific and engineering and policy actors.

References

Bozeman, B. (2003). *Public value mapping of science outcomes: Theory and method*. Washington, DC: Center for Science, Policy, and Outcomes.

Calvert, J., & Martin, P. (2009). The role of social scientists in synthetic biology. *EMBO Reports, 10*, 201–204.

Directorate-General for Research and Innovation EU (2012). *Ethical and regulatory challenges to science and research policy at the global level*. Brussel: European Commission.

Doorn, N., & Nihlén Fahlquist, J. A. (2010). Responsibility in engineering. Towards a new role for engineering ethicists. *Bulletin of Science, Technology & Society, 30*, 222–230.

Doorn, N. (2012). Exploring responsibility rationales in Research and Development (R&D). *Science, Technology & Human Values, 37*, 180–209.

Flipse, S. M., Van der Sanden, M. C. A., & Osseweijer, P. (2013). Midstream modulation in biotechnology industry: Redefining what is 'part of the job' of researchers in industry. *Science and Engineering Ethics, 19*, 1141–1164.

Gorman, M. E. (2011). Doing science, technology and society in the national science foundation. Commentary on: "Engaged, embedded, enjoined: Science and technology studies in the national science foundation". *Science and Engineering Ethics, 17*, 839–849.

Gorman, M. E., Wardak, A., Fauss, E., & Swami, N. (2008). A framework for using nanotechnology to improve water quality. In N. Savage et al. (Eds.), *Nanotechnology applications for clean water*. Norwich/New York: William Andrew Applied Science Publishers.

Gorman, M. E., Werhane, P. H., & Swami, N. (2009). Moral imagination, trading zones, and the role of the ethicist in nanotechnology. *NanoEthics, 3*, 185–195.

Kant, I. (1795). *To perpetual peace: A philosophical sketch* [Zum ewigen Frieden: Ein philosophischer Entwurf]. Königsberg: Friedrich Nicolovius.

Krohn, W., & Weyer, J. (1994). Society as a laboratory. The social risks of experimental research. *Science and Public Policy, 21*, 173–183.

Leonard-Barton, D. (1988). Implementation as mutual adaption of technology and organization. *Research Policy, 17*, 251–267.

Lucivero, F., Swierstra, T., & Boenink, M. (2011). Assessing expectations: Towards a toolbox for an ethics of emerging technologies. *NanoEthics, 5*, 129–141.

Renn, O. (2005). *White paper on risk governance. Towards an integrative approach*. Geneve: International Risk Governance Council.

Rip, A., & Kemp, R. (1998). Technological change. In S. Rayner & E. L. Malone (Eds.), *Human choice and climate change* (pp. 327–399). Columbus: Battelle.

Schot, J. W., & Rip, A. (1997). The past and future of constructive technology assessment. *Technological Forecasting and Social Change, 54*, 251–268.

Schuurbiers, D. (2011). What happens in the lab: Applying midstream modulation to enhance critical reflection. *Science and Engineering Ethics, 17*, 769–788.

Selin, C. (2007). Expectations and the emergence of nanotechnology. *Science, Technology & Human Values, 32*, 196–220.

Swierstra, T., Boenink, M., & Van Est, R. (2009). Converging technologies, shifting boundaries. *NanoEthics, 3*, 213–216.

Van de Poel, I. R. (2008). How should we do nanoethics? A network approach to discerning ethical issues in nanotechnology. *NanoEthics, 2*, 25–38.

Van de Poel, I. R. (2009). The introduction of nanotechnology as a societal experiment. In S. Arnaldi, A. Lorenzet, & F. Russo (Eds.), *Technoscience in progress. Managing the uncertainty of nanotechnology* (pp. 129–142). Amsterdam: Ios Press.

Verbeek, P. P. (2011). *Moralizing technology: Understanding and designing the morality of things*. Chicago: University of Chicago Press.

Von Schomberg, R. (Ed.). (2011). *Towards responsible research and innovation in the information and communication technologies and security technologies fields*. Brussels: EU Directorate General for Research and Innovation.

Von Schomberg, R. (2012). Prospects for technology assessment in a framework of responsible research and innovation. In M. Dusseldorp & R. Beecroft (Eds.), *Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden* (pp. 39–61). Wiesbaden: Springer.

Wildavsky, A. B. (1988). *Searching for safety*. New Brunswick: Transaction Books.

Woolgar, S., Coopmans, C., & Neyland, D. (2009). Does STS mean business? *Organization, 16*(1), 5–30.

Wynne, B. (1992). Uncertainty and environmental learning. Reconceiving science and policy in the preventive paradigm. *Global Environmental Change, 2*, 111–127.