



Emerging Technologies, Ethics and International Affairs

NEW PERSPECTIVES ON TECHNOLOGY IN SOCIETY

EXPERIMENTATION BEYOND THE LABORATORY

Edited by
Ibo van de Poel, Lotte Asveld,
and Donna C. Mehos



ROUTLEDGE



“Technologies released into the public domain are expected to be safe and reliable. Experiments, by definition, are ventures into the unknown. How, thus, can technologies in society be responsibly conceptualized as social experiments? *New Perspectives on Technology in Society* is a major step to help discern ethical issues and possibilities of moral learning during the experimental introduction of new technologies into society. The contributions in this excellent volume show on how carefully planned social experiments can denote processes in which ignorance and surprise are creatively used for deliberate and systematic learning. Given the unavoidable character of the experimental nature of novel technologies in the twenty-first century (from self-driving cars to synthetic biology or new medical tests) the studies in this volume could not be timelier.”

— *Matthias Gross, Helmholtz Centre for Environmental Research
and University of Jena, Germany*

“What happens when technologies are introduced? Some describe this as the social construction of a new world; others refer to it as a collective design process. This volume offers a new perspective on technology in society: Much is gained by conceiving of technological innovation as a form of experimentation—new technologies are bold interventions that set into motion a process of social learning and political negotiation. Experiments in the laboratory, in the field, and in the real world raise questions of method and the limits of control that are taken up in this wide-ranging collection of papers.”

— *Alfred Nordmann, Institut für Philosophie,
TU Darmstadt, Germany*

This page intentionally left blank

New Perspectives on Technology in Society

The development and introduction of a new technology into society can be viewed as an experimental process, full of uncertainties which are only gradually reduced as the technology is employed. Unexpected developments may trigger an experimental process in which society must find new ways to deal with the uncertainties posed.

This book explores how the experimental perspective determines what ethical issues new technologies raise and how it helps morally evaluate their introduction. Expert contributors highlight the uncertainties that accompany the process, identify the social and ethical challenges they give rise to, and propose strategies to manage them.

A key theme of the book is how to approach the moral issues raised by new technology and how to understand the role of experimentation in exploring these matters.

Ibo van de Poel is Antoni van Leeuwenhoek Professor in Ethics and Technology and head of the Department of Values, Technology & Innovation at the Faculty Technology, Policy and Management of the Technical University Delft in the Netherlands. He has published on engineering ethics, the moral acceptability of technological risks, design for values, responsible innovation, moral responsibility in research networks, ethics of new emerging technologies, and the idea of new technology as social experiment.

Lotte Asveld is an Assistant Professor at Delft University of Technology studying the societal aspects of biotechnology. Her main research interest concerns responsible innovation in the field of biotechnology and synthetic biology: how can the societal debate on biotechnology and synthetic biology be integrated in innovation trajectories? Lotte has worked as a researcher in the department of Philosophy at DUT, where she also received her Ph.D. Her Ph.D. concerned societal decision making on technological risks. Lotte also worked as a researcher at the Rathenau Institute, focusing on the bioeconomy, and as a freelance researcher in China.

Donna C. Mehos is an independent scholar who has studied historical and sociological aspects of science and technology. In her earlier work, she examined nineteenth-century science in European cultural life and technological expertise in the colonial and postcolonial world, including the technopolitics of the Cold War. Her recent work explores current infrastructure development, including decentralization of infrastructure networks, social acceptability and policy implications of wind energy, and the future of gas in energy infrastructures.

Emerging Technologies, Ethics and International Affairs

Series Editors: Steven Barela, Jai C. Galliot, Avery Plaw,
Katina Michael

This series examines the crucial ethical, legal and public policy questions arising from or exacerbated by the design, development and eventual adoption of new technologies across all related fields, from education and engineering to medicine and military affairs. The books revolve around two key themes:

- Moral issues in research, engineering and design
- Ethical, legal and political/policy issues in the use and regulation of technology

This series encourages submission of cutting-edge research monographs and edited collections with a particular focus on forward-looking ideas concerning innovative or as yet undeveloped technologies. Whilst there is an expectation that authors will be well grounded in philosophy, law or political science, consideration will be given to future-orientated works that cross these disciplinary boundaries. The interdisciplinary nature of the series editorial team offers the best possible examination of works that address the 'ethical, legal and social' implications of emerging technologies.

For a full list of titles, please see our website: [www.routledge.com/
Emerging-Technologies-Ethics-and-International-Affairs/book-series/
ASHSER-1408](http://www.routledge.com/Emerging-Technologies-Ethics-and-International-Affairs/book-series/ASHSER-1408)

Healthcare Robots

Ethics, Design and Implementation

Aimee van Wynsberghe

Ethics and Security Automata

Policy and Technical Challenges of the Robotic Use of Force

Sean Welsh

New Perspectives on Technology in Society

Experimentation beyond the Laboratory

Edited by Ibo van de Poel, Lotte Asveld, and Donna C. Mehos

New Perspectives on Technology in Society

Experimentation beyond
the Laboratory

**Edited by Ibo van de Poel,
Lotte Asveld, and Donna C. Mehos**

First published 2018
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2018 selection and editorial matter, Ibo van de Poel, Lotte Asveld, and Donna C. Mehos; individual chapters, the contributors

The right of Ibo van de Poel, Lotte Asveld, and Donna C. Mehos to be identified as the authors of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

Names: Poel, Ibo van de, 1966– editor. | Asveld, Lotte, editor. | Mehos, Donna C. editor.

Title: New perspectives on technology in society: experimentation beyond the laboratory / edited by Ibo van de Poel, Lotte Asveld, and Donna C. Mehos.

Description: Abingdon, Oxon; New York, NY: Routledge, 2018. |

Includes bibliographical references and index.

Identifiers: LCCN 2017030216 |

Subjects: LCSH: Social sciences—Experiments. |

Technological innovations—Social aspects.

Classification: LCC HM846.E97 2018 | DDC 300—dc23

LC record available at <https://lccn.loc.gov/2017030216>

ISBN: 978-1-138-20401-0 (hbk)

ISBN: 978-1-315-46825-9 (ebk)

Typeset in Times New Roman
by codeMantra

Contents

<i>List of illustrations</i>	ix
<i>Notes on contributors</i>	xi
<i>Acknowledgements</i>	xv
Introduction	1
IBO VAN DE POEL, DONNA C. MEHOS, AND LOTTE ASVELD	
1 Control in scientific and practical experiments	16
PETER KROES	
2 The diversity of experimentation in the experimenting society	36
CHRISTOPHER ANSELL AND MARTIN BARTENBERGER	
3 Moral experimentation with new technology	59
IBO VAN DE POEL	
4 The theatrical debate: Experimenting with technologies on stage	80
FRANK KUPPER	
5 Social learning in the bioeconomy: The Ecover case	103
LOTTE ASVELD AND DIRK STEMERDING	
6 Cognitive enhancement: A social experiment with technology	125
NICOLE A. VINCENT AND EMMA A. JANE	
7 Living a real-world experiment: Post-Fukushima imaginaries and spatial practices of “containing the nuclear”	149
ULRIKE FELT	

8 “Dormant parasites”: Testing beyond the laboratory in Uganda’s malaria control program	179
RENÉ UMLAUF	
9 Experimenting with ICT technologies in youth care: <i>Jeugdzorg</i> in the Netherlands	199
BEN KOKKELER AND BERTIL BRANDSE	
10 Adversarial risks in social experiments with new technologies	222
WOLTER PIETERS AND FRANCIEN DECHESNE	
 <i>Index</i>	 241

List of illustrations

Figures

5.1	Relationship between worldviews, frameworks and perceptions	109
5.2	'Four prevalent views on Nature'	110
5.3	Frameworks/perspectives	114
7.1	Typical map published after the Fukushima accident	159
7.2	Radiation map	160
7.3	Map identifying the status of the territory	161
7.4	Public map in Minamisoma (showing leisure activities in the region) © Ulrike Felt 2013	163
7.5	Cross-road closure on the way to the evacuation zone (corridor road). © Ulrike Felt 2013	165
7.6	A boat in the middle of the plain in front of a sign "schoolchildren crossing". © Ulrike Felt 2014	166
7.7	Physical demarcations: Border control at the evacuation zone. © Ulrike Felt 2013	167
7.8	Physical demarcations: Radiation control barriers. © Ulrike Felt 2014	167
7.9	Namie City – abandoned to "the nuclear": One of the empty main streets. © Ulrike Felt 2014	168
7.10	Namie City – abandoned to "the nuclear": Abandoned bike rack at the train station. © Ulrike Felt 2014	169
7.11	Public Geiger counter within an inhabited zone showing 0.52 microSievert/h. © Ulrike Felt 2013	170
7.12	Radioactive soil collected in plastic bags. © Ulrike Felt 2014	172
8.1	RDT training material. Source: www.finddiagnostics.org	186
8.2	Suspected cases, positive cases, and ACTs. Source: PMI/USAID Kampala May, 2013	187

Tables

4.1	The tasks for pragmatist ethics	87
4.2	The "Doctor inside!" scenario	91

This page intentionally left blank

Notes on contributors

Christopher Ansell is Professor of Political Science at the University of California, Berkeley. He received his B.A. in Environmental Science from the University of Virginia and his Ph.D. in Political Science from the University of Chicago. He is the author of *Pragmatist Democracy: Evolutionary Learning as Public Philosophy* (Oxford University Press 2011) and the co-editor of *Public Innovation through Collaboration and Design* (Routledge 2014), the *Handbook of Theories of Governance* (Edward Elgar), and *Governance in Turbulent Times* (Oxford University Press 2017).

Lotte Asveld is an Assistant Professor at Delft University of Technology studying the societal aspects of biotechnology. Her main research interests concern responsible innovation in the field of biotechnology and synthetic biology: how can the societal debate on biotechnology and synthetic biology be integrated in innovation trajectories? Lotte has worked as a researcher in the department of Philosophy at DUT, where she also received her PhD. Her PhD concerned societal decision making on technological risks. Lotte also worked as a researcher at the Rathenau Institute, focusing on the bioeconomy, and as a freelance researcher in China.

Martin Bartenberger earned a Master degree in Political Science from the University of Vienna and is currently finishing his PhD studies at the Institute of Political Science at Leiden University. He has taught at the University of Vienna and the Vienna University of Economics and Business and has worked as a research manager, lecturer and web programmer. His research focuses on crisis management and experimentalist policy making and has been published in *Ecological Economics* and *Public Administration* as well as in several other journals and anthologies.

Bertil Brandse is senior expert and researcher of applied sciences on alignment of information systems with objectives and strategies of organizations and of user communities. In his master degree granted by the department of Business Computer Science at the University of Twente,

he specialized in information planning and management. Bertil Brandse founded his own firm in 1992, of which he is CEO since 1998, THORAX Information Projects & Consultancy (www.thorax.nl). His team at Thorax focuses on the social services sector and health sector in the Netherlands and is mainly involved in programs and projects wherein the citizen is at the center of innovative IT-solutions. Bertil Brandse is member of the Avans UAS research group on Social resilience in the digital society.

Francien Dechesne is researcher at the Center for Law and Digital Technologies (eLaw) of the Leiden Law School. She holds a PhD in mathematical logic and has postdoctoral research experience in the field of verification of computer security protocols. She is particularly interested in bridging between the formal sciences and ethical and societal aspects of ICT. Her current research within the SCALES project of the NWO Responsible Innovation program looks into the balancing of public, commercial and individual interests and risks in applications of data analytics.

Ulrike Felt is Professor of Science and Technology Studies (STS) and Head of the interfaculty research platform “Responsible research and innovation in academic practice”. Her research focuses on governance, democracy and public participation as well as on shifting research cultures. Her main areas of study cover the life science/(bio)medicine, nanotechnology, nuclear energy and sustainability research. From 2002–2007 she was editor-in-chief of *Science, Technology, & Human Values*. She was leading the editorial team of the most recent Handbook of Science and Technology Studies (MIT Press, 2017). Since 2017 she is president of European Association for the Study of Science and Technology (EASST).

Emma A. Jane (formerly Emma Tom) is a Senior Research Fellow at the University of New South Wales in Sydney, Australia. Misogyny online is the focus of her ongoing research into the social and ethical implications of emerging technologies. In 2016, Emma received the Anne Dunn Scholar Award for excellence in research about communication and journalism. This followed her receipt, in 2014, of a three-year government grant to study gendered cyberhate and digital citizenship. Prior to her career in academia, Emma spent nearly 25 years working in the print, broadcast, and electronic media during which time she won multiple awards for her writing and investigative reporting. Her ninth book – *Misogyny Online: A Short (and Brutish) History* – was published by Sage in 2017.

Ben Kokkeler is a senior expert and researcher on Open and Responsible Innovation, focusing on smart cities, smart health, and smart public safety. As visiting researcher he works at STePS (Science, Technology and Policy Studies), the department at the University of Twente where he prepared his PhD thesis on “distributed academic leadership in emergent research organisations”, supervised by professors Rip and Kuhlmann, and

Fisscher (Organisation Studies and Business Ethics). At the European Technopolis Group he is senior consultant, at Avans University of Applied Sciences professor on Social resilience in the digital society (Centre for public safety and criminal justice).

Peter Kroes is emeritus Professor in Philosophy of Technology at Delft University of Technology, The Netherlands. He has an engineering degree in physics (1974) and wrote a PhD thesis on the notion of time in physical theories (University of Nijmegen, 1982). He has been teaching courses in Philosophy of Science and Technology and Ethics of Technology, mainly for engineering students. His research in Philosophy of Technology focuses on technical artifacts, engineering design, socio-technical systems and technological knowledge.

Frank Kupper is Assistant Professor in science communication and public engagement at VU Amsterdam. Trained as a biologist, philosopher and theater maker, he received a PhD in Science & Technology Studies at VU Amsterdam. His research and teaching revolve around creating a better understanding of science-society interactions, designing and facilitating reflexive public engagement and dialogue. Specifically, he focuses on the potential role of playful and creative methods to nurture processes of reflection, learning and change at the interface of science and society. Ultimately, he aims to contribute to creative democracy as a way to shape the conversation about science in society.

Donna C. Mehos is an independent scholar who has studied historical and sociological aspects of science and technology. In her earlier work, she examined nineteenth-century science in European cultural life and technological expertise in the colonial and postcolonial world, including the technopolitics of the Cold War. Her recent work explores current infrastructure development, including decentralization of infrastructure networks, social acceptability and policy implications of wind energy, and the future of gas in energy infrastructures.

Wolter Pieters is an Assistant Professor in cyber risk at Delft University of Technology, faculty of Technology, Policy and Management. He has MSc degrees in computer science and philosophy of science, technology and society from the University of Twente, and a PhD in information security from Radboud University Nijmegen, focused on the controversy on electronic voting in elections. His research interests include cyber risk management, cyber security decision making, and cyber ethics. He was technical leader of the TRESPASS European project on socio-technical cyber risk management, and will be part of the new CYBECO project on behavioural models for cyber insurance.

Dirk Stemerding has been working as a senior researcher Technology Assessment at the Dutch Rathenau Instituut. He was one of the co-authors

of the Rathenau study *Getting to the core of the bio-economy: a perspective on the sustainable promise of biomass* (2011). He has been leading a work package on synthetic biology in the European project *Global Ethics in Science & Technology* (GEST 2011-2014) and was one of the editors of the volume *Science and Technology Governance and Ethics: a global perspective from Europe, India and China* (Springer 2015). He was also involved as work package leader in a four-year European Mobilisation and Mutual Learning Action Plan aiming at responsible research and innovation in synthetic biology (SYNENERGENE 2013-2017). Since his retirement he is working as an independent researcher on issues relating to biotechnology & society.

René Umlauf received his Ph.D. in Sociology from the University of Bayreuth. His research focus is on the relation between science, technology and organizations in the context of Global Health. He is particularly interested in the role tests and testing procedures of different scale and scope play for the production of multiple forms of evidence. In his current post-doc work at University of Halle René aims to expand the notion of testing. He looks at the complex dynamics between changing disease perceptions, shifting Global Health agendas and the role of data production and how this affects standardization of therapeutic services.

Ibo van de Poel is Antoni van Leeuwenhoek Professor in Ethics and Technology and head of the Department of Values, Technology & Innovation at the Faculty Technology, Policy and Management at the Technical University Delft in the Netherlands. He has published on engineering ethics, the moral acceptability of technological risks, design for values, responsible innovation, moral responsibility in research networks, ethics of new emerging technologies, and the idea of new technology as social experiment.

Nicole A. Vincent is a philosopher at Macquarie University and University of New South Wales in Australia, and at Technische Universiteit Delft in The Netherlands. Supported by over \$1 million external grants, she has over forty scholarly publications in neurolaw, neuroethics, bioethics, ethics, philosophy of tort and criminal law, political philosophy, and philosophy of technology. With close to 100 talks at international scholarly meetings, as well as popular media coverage via TED talks, television and radio interviews, newspaper articles, and public debates, she strives to bring scholars into dialogue with the public. She reviews for a range of scholarly journals and funders, and serves on the neuroethics sub-panel of the Australian Brain Alliance.

Acknowledgements

This edited volume is based on papers that were first presented at the International Conference on Experimenting with New Technologies in Society, which was held August 20–22, 2015, in Delft, The Netherlands. This book was written as part of, and made possible by, the research program ‘New Technologies as Social Experiments’, which was supported by the Netherlands Organization for Scientific Research (NWO) under grant number 277-20-003.

This page intentionally left blank

Introduction

*Ibo van de Poel, Donna C. Mehos,
and Lotte Asveld*

In this volume, we explore the perspective that the development of a technology and its introduction to society should be seen as an experimental process because of the myriad of uncertainties which will only be reduced gradually after the technology is actually employed. Furthermore, unexpected technological developments, such as the Fukushima disaster, may trigger an experimental process in which society must contend with new uncertainties posed by a technology and its social embedding and effects. The aim of this edited volume is to understand better the role of experimentation in the introduction of new technologies into society. This concerns what has been called real-world or societal experimentation with new technologies (Martin and Schinzinger 1983; Krohn and Weyer 1994; Felt et al. 2007; Van de Poel 2009), and where appropriate in the volume, we also address more traditional forms of experimentation, such as laboratory and field experiments as well as thought experiments and living labs.

In our conceptualization of the introduction of a new technology into society as a social experiment, we see that introduction as a learning process in which the consequences of it emerge only gradually. The learning may be more or less deliberate and may come at few or great social costs. Below we will distinguish between tacit and deliberate modes of experimentation in the introduction of new technology. The important point is that our conceptualization not only offers a framework to understand and analyze the process, but it also can help to judge morally the introduction of new technology into society and to design better ways for doing so.

We focus in this volume on the introduction of new technologies as well as experimentation as a way to perceive new developments and changing contexts. A main theme is how to approach the moral issues raised by new technology, but we look broader as we first need to understand the role of experimentation in the introduction of new technologies. The experimental perspective can help us discern what ethical issues (new) technologies raise in society. In addition, it may also be used to morally evaluate the experimental introduction of new technology as several contributions to the book do. Finally, it may draw attention to the fact that also the moral norms and values by which we judge new technology are in flux because they take shape in an experimental process.

By focusing on the moral issues that arise from new technologies and on processes of moral learning, our focus is broader than that in, for example, strategic niche management (Kemp, Schot and Hoogma 1998; Hoogma et al. 2002) and transition management approaches (Kemp, Loorbach and Rotmans 2007; Rotmans and Loorbach 2008). While experimentation and learning play a major role in those approaches, they aim at deliberately changing social and economic configurations, and focus on technical and institutional learning. We focus more sharply on normative and moral learning.

Below, we start by developing the idea further that a new technology's introduction into society is a form of social experimentation, and we point out how this idea plays a role in various contributions to this book. Next, we explore the forms of experimentation in sites beyond the laboratory, and thus we extend the traditional notion of experiment. We then briefly introduce the various contributions to the book. We end with a discussion on how we can move from tacit to more deliberate social experimentation with new technology.

The introduction of new technology as a social experiment

When new technologies are introduced into society, they amount to social experiments. This means that some of the risks and benefits, but also some of the ethical issues, which are raised by a new technology only become known after it is actually introduced into society. This also implies that efforts to anticipate the risks, benefits, and ethical issues raised before that technology is actually employed will only partly succeed (cf. Collingridge 1980). Due to ignorance and indeterminacy, surprises are likely to happen, and only after its implementation will we gradually learn about the impacts of a technology on society, the normative and moral issues raised by such processes, and the best way to embed it in society.

However, the notion of experiment not only refers to uncertainty and lack of control, but it also denotes a process in which uncertainty is reduced through deliberate and systematic learning. This process is aimed at increasing control over the intervention (i.e., the introduction of the new technology into society) and over the outcome of the intervention (i.e., risks, benefits, and other impacts). It must be recognized that sometimes the introduction of a new technology is described as a social experiment, even when this element of deliberate learning is absent and when experiments are not named as such. In these cases, we might speak of a tacit or *de facto* social experiment. Such tacit experimentation is to be distinguished from what might be called deliberate or proper social experiments, which are not only called that by name but in which there is a systematic and deliberate attempt to learn from the experiment.

In several chapters, authors explicitly discuss the idea that the introduction of a new technology into society amounts to a social experiment; for

example, Asveld and Stermerding focus on the bioeconomy, Pieters and Dechesne explore (cyber)security technologies, and Vincent and Jane investigate human enhancement technologies. In other cases, the terminology of the authors may be less explicit, but they describe cases consistent with the idea that the introduction of a new technology is a social experiment. Kokeler and Brandse describe the introduction of new information and communication technologies in social services in the Netherlands. They point out that there is deliberate experimentation in the ways new technologies are used with clients, but beyond that there is social experimentation with (conflicting) routines, norms, and values for providing social services for youths. This social experimentation is largely tacit, although it sometimes is deliberately organized in living labs. Umlauf, in his contribution, describes the introduction of new standardized tests for malaria intended to simplify diagnosis in the fragmented Ugandan health care infrastructure. He observed how the tests were not used as intended and thus resulted in an experiment with the new technology. The case also illuminates a tacit experiment with different social institutions involved in malaria diagnosis thus bringing together tacit social experimentation with a more deliberate experimentation. In Felt's contribution on the Fukushima nuclear accident, we also encounter a tacit social experiment. This experiment is not about the introduction of a new technology but rather about restoring normalcy after a technical accident. Felt argues that the experiment is about restoring (the image of) control and containment of nuclear technology.

What the samples in this edited volume suggest is that most social experimentation with new technology that takes place is tacit and *de facto* rather than deliberate and planned. There are two reasons for making tacit social experimentation more explicit and deliberate. The first reason is to improve the possibilities for learning, thus improving the introduction of a new technology into society by decreasing social costs in terms of risks and disadvantages. This possibility is explicitly discussed by Asveld and Stermerding who address two strategies for more deliberate experimentation and social and moral learning, namely scaling up and adaptability. They propose a number of recommendations to improve learning in the bioeconomy: one, that it should not focus on just one technology but rather on the whole chain of development and implementation; and two, that thought experiments should be performed earlier in the process of introducing new technology. Thought experiments are discussed in more detail in the chapters by Van de Poel, who includes a moral philosophical analysis, and Kupper, who argues that theatrical debates employ thought experiments to facilitate public discussion of the ethical issues raised by nanotechnology.

The second reason to increase deliberate (rather than tacit) experimentation is that the people who are subject to such experiments have a right to know so. In Van de Poel's contribution, he briefly discusses ethical issues brought about by different kinds of moral experimentation with new technology. However, in particular, the chapter by Vincent and Jane suggests that

turning tacit experimentation into explicit and deliberate experiments might not be enough to make it a responsible or an ethically acceptable mode of experimentation. They propose a methodology to make the introduction of new enhancement technologies into society more acceptable, a methodology that can be used for other technologies as well.

Experimentation beyond the laboratory: Different modes of experimentation

Studies on experimentation in the philosophy of science have, in most cases, focused on laboratory experiments (e.g., Popper 1963; Hacking 1983; Radder 2003). Although Popper discusses social experiments, he does so only in *The Open Society and its Enemies* (1945) where he does not consider in any detail the methodology of such experiments. Also in Science and Technology Studies (STS), the focus has traditionally been on laboratory experiments (e.g., Latour and Woolgar 1979; Collins 1985), although there is some scholarship on experiments in the real world (e.g., Krohn and Weingart 1987; Wynne 1988; Krohn and Weyer 1994).

This book focuses on sites other than the laboratory, in particular, on society as the locus of experimentation. Such experiments have been described as real-life experiments (Krohn and Weyer 1994), real-world experiments (Gross and Hoffmann-Riem 2005; Levidow and Carr 2007), societal experiments (Martin and Schinzinger 1983; Van de Poel 2009), social experiments (Krohn and Weingart 1987; Wynne 1988; Herbold 1995; Van de Poel 2011; Bösch 2013), and collective experimentation (Felt et al. 2007). These are experiments that are “often not called by name” (Felt et al. 2007, 68). The question can be asked whether these are real experiments or if the notion of experiment is used here metaphorically or critically. Similarly, one can wonder whether the learning in such social experiments is different from the trial-and-error learning that takes place in society more generally.

Some of these questions are taken up by Ansell and Bartenberger in their contribution to this volume. They argue for a broad notion of experiment. They understand experimentation as “taking an action with the intention of learning ‘what if.’” They distinguish between three types of experimentation: controlled experimentation, evolutionary experimentation, and generative experimentation. Ansell and Bartenberger thus allow for uncontrolled experiments. This is not uncontroversial, as control is often seen as a precondition for proper experimentation (e.g., Webster and Sell 2007; Gonzales 2010; Hansson 2015).

In conventional controlled experiments, control refers to the intervention taking place with control over the experimental conditions; it does not refer to control over the results. The outcome of an experiment is typically not known with certainty nor is it controlled; that is why it is worth performing. Control over the intervention and/or the experimental conditions is nevertheless considered essential to gain knowledge because control allows

for hypothesis testing; i.e., it allows us to create the conditions (or kinds of interventions) that theoretically would lead to a certain outcome and, through the experiment, we can test whether this outcome indeed occurs. Without control, we usually cannot create the conditions (or kinds of interventions) that meet the theoretical requirement. Nevertheless, as Morgan (2013) points out, sometimes control may occur without human intervention (e.g., in exceptional but naturally-occurring circumstances), or control may be achieved through statistical manipulation of observational data.

Even if we do not test a hypothesis in an experiment, control is often deemed necessary to find causal relations because, through experimental control, we can manipulate the independent variables and observe their effect on the dependent variable(s). By systematically varying the independent variables, we are better able to infer the effect of the independent variables on the dependent variables and thus make inferences about causal relations.

Kroes in his contribution discusses the control paradigm for scientific experiments and concludes that, if the introduction of new technology into society is conceived of as a social experiment, it cannot be a controlled experiment. Nevertheless, he recognizes the potential of uncontrolled social experiments. In this connection, he speaks of practical experiments, which are similar to what Ansell and Bartenberger in their contribution call “generative experimentation” and what Ansell (2012) has earlier called “design experiments.”

The notion of design experiments stems from aeronautics and artificial intelligence, where it is used to refer to the gradual process of gathering information about the implementation of an intervention or an innovation to adapt and improve it until it works well (Stoker and John 2009, 358). The notion has been taken up in educational design and more recently in policy design (cf. Ansell 2012). Also, forms of ecological reconstruction, such as the restoration of lakes (Gross 2010), can be conceived of as design experiments, where the aim is not general ecological knowledge but rather to learn from the actual reconstruction and its consequences to improve the reconstruction itself.

An example of a design experiment in technology is the introduction of the autopilot on Tesla cars. In 2016, a Tesla car had a fatal accident when the camera failed to recognize a white truck against a bright sky and the autopilot failed to brake (Boudette 2017). After the accident, Tesla declared that it “disables Autopilot by default and requires explicit acknowledgement that the system is new technology and [is] still in a public beta phase before it can be enabled” (Tesla Team 2016). Tesla thus explicitly recognizes the experimental nature of the technology and asks drivers for their informed consent before the system can be used. And,

[w]hen drivers activate Autopilot, the acknowledgment box explains, among other things, that Autopilot “is an assist feature that requires you to keep your hands on the steering wheel at all times”, and that

“you need to maintain control and responsibility for your vehicle” while using it.

(Tesla Team 2016)

The Tesla autopilot is not only a design experiment because Tesla explicitly recognizes the technology as experimental, but also because the experience of using it may lead to further improvements in the system. In the case of the fatal crash, one may not only think of technical improvements for the camera recognition but also of better ways for dealing with the behavior changes that an autopilot brings about. There is evidence that the driver in this case relied too much on the system. An investigation of the National Highway Traffic Safety Administration (NHTSA) found that the driver was not paying attention to the road (Boudette 2017). After the accident, Tesla updated the software of the system, and it now gives drivers more frequent warnings to keep their hands on the steering wheel. The system also shuts off after three warnings, and it can then only be restarted when the driver stops and restarts the car.

As the Tesla example shows, the learning that takes place in design experiments is neither fundamental scientific learning nor aimed at general knowledge of the technology. It is learning about the specific, namely a specific intervention (the introduction of a technology into) to improve it along the way. This type of learning requires neither the formulation of hypotheses to be tested nor controlled experimentation.

In addition to experiments in the lab, we can distinguish a category of what might be called “social design experiments”, and such social experiments may lack experimental control but can still be justified as experiments. Social and laboratory experiments may be seen as two extremes on one scale. In between, we find what have been called “field experiments” and “experimentation in living labs,” a notion which has become popular over the last few years. A living lab is a real-world setting in which experiments are performed and in which deliberate learning takes place on the basis of experimental observations.

In the scant philosophical and STS literature on field experiments, two interpretations of field experiments can be found. One is that field experiments are less controlled than laboratory ones or completely uncontrolled (e.g., Gross, Hoffmann-Riem, and Krohn 2005; Schwartz 2014), or two, that field experiments are controlled experiments, but the control is achieved in ways other than in the laboratory (e.g., Morgan 2013). An example of controlled field experiments are randomized controlled trials to test the efficacy (and possible side effects) of new drugs.

Field experiments can be either controlled or uncontrolled and both are possible if we distinguish between the site of experimentation (lab, field, or society at large) and the set-up of an experiment (e.g., controlled, evolutionary, or generative). Those who conceive of field experiments as uncontrolled typically compare them to design experiments while those who see them

as controlled assume the set-up of a controlled experiment (usually as controlled randomized trial). If we can indeed distinguish between the site and the set-up of an experiment, it would suggest that there can be, at least in theory, controlled social experiments. An example would be the experiments on social reform to evaluate public policy interventions in the US between the 1960s and 1980s, which were inspired by the work on social experiments by Donald Campbell (e.g., Campbell and Stanley 1966; Campbell and Russo 1999). These experiments were carried out as randomized controlled trials. Conversely, it would seem possible to have laboratory experiments without experimental control, for example, to observe new phenomena. As Hacking (1983, 154) indeed points out, in the laboratory “[o]ne can conduct an experiment simply out of curiosity to see what will happen.”

Experiments may not only take place in the laboratory, the field, or society at large, but also in imagined or imaginary worlds. One example is computer simulations; another, thought experiments. For both types of experiments, there has been discussion on whether they can be properly called experiments (see e.g., Sorensen 1992; Winsberg 2009, 2015; Brown and Fehige 2017). One reason to doubt that thought experiments and simulations are real experiments is that they, at least *prima facie*, do not produce new empirical data. A thought experiment may point out unexpected implications of a theory but in doing so does not reveal anything that was not yet implied in the theory. Similarly, computer simulations do not produce new empirical data; they produce new data, but these have been implicitly included in the computer algorithms.

In response to this, two arguments may be made for why thought experiments and (computer) simulations are nevertheless sometimes experiments. First, even in cases that do not produce new or fresh empirical data, they may point out certain implications of a theory or a simulated phenomenon that the experimenters did not expect or know, and as such, they may produce new insights and shed light on theories or phenomena. Second, as Van de Poel points out in his contribution about morally experimenting with new technology, moral thought experiments with new technology may lead to new moral experiences. Although thought experiments take place in imagined worlds, they may produce real and new moral experiences.

For moral experimentation, as Van de Poel argues, thought experiments have distinct advantages over social experiments. One advantage is that they produce no harm to the real world; another is that a series of thought experiments in which the parameters are varied systematically is relatively easy to do. The main disadvantage of thought experiments is that their external validity is limited, i.e., they may not be representative of people’s moral reactions if moral dilemmas occur in reality. In this sense, the relation between moral thought experiments and moral experimentation in society at large is somewhat similar as that between lab experiments and social experiments in terms of the functioning and social impact of new technologies. Lab experiments are less harmful and may be more easily repeatable than social experiments, but they may not

reliably predict the full spectrum of social impact of a new technology. It is for this reason that we call the introduction of a new technology in society a social experiment, regardless of how many laboratory experiments and thought experiments have taken place before it is introduced into society. This is not to say that lab and thought experiments are useless. On the contrary, they are likely to contribute to a better and less harmful introduction of a technology into society but usually not to the extent that the actual introduction is no longer experimental.

As there is a need for sites of experimentation between the lab and society at large—like field experiments and living labs—there is also a need for forms of moral experimentation between thought experiments and experiments in society at large. One such form is what Van de Poel calls “experiments in living,” a term coined by J.S. Mill in his *On Liberty* and which has also been used by, for example, Anderson (1991). Experiments in living are done by an individual to test out a morally-loaded vision of the good life. It might be argued that experiments in living are never fully individual as all people are part of a web of social relations and their well-being affects others. Nevertheless, experiments in living are usually less encompassing than social experiments, which, on the one hand, limits what can be morally learned from them, while on the other hand, limits the harm that may ensue from such experiments.

All in all, we can conclude that there is a wide variety of forms of experimentation beyond the laboratory. If we want to understand this variety, we should move beyond the simple dichotomy of controlled laboratory experiments versus uncontrolled social experiments. In the wide spectrum of intermediate forms, we should distinguish between the site of experimentation (virtual world, laboratory, field, and society), the set-up of the experiment (controlled, generative, and evolutionary), and the aims of an experiment (hypothesis testing, finding causal relations, improving an intervention along the way).

Overview of the contributions

The chapters in his book on experimentation, social experiments, and technology address conceptual aspects as well as the roles of moral and social learning in evaluating the introduction of technologies. Specific cases of social experiments are explored that shed light on a variety of consequences when new technologies are introduced into society. Taken together, they provide various perspectives of how we can benefit from approaching the introduction of technology as social experimentation.

Beginning with “Control in scientific and practical experiments,” Peter Kroes analyzes scientific and what he calls “practical” experiments by focusing on the role of control in their experimental paradigms. He argues that, because of their differing levels of control, the knowledge produced from these two types of experiments is fundamentally different. In conventional

scientific laboratory and controlled social scientific experiments, parameters are well defined, as is the experimental system that yields knowledge of the regularities of the natural world. Practical experimentation of sociotechnical systems cannot maintain the level of control of scientific experiments and thus does not illuminate regularities because they lack well-defined interventions and roles of the experimenter. Furthermore, unpredictable human behavior plays a role that cannot be controlled. While learning does take place in uncontrolled practical experiments, it differs in part because the results are not undeniable facts but rather subject to interpretation by a variety of stakeholders.

In Chapter 2, “The diversity of experimentation in the experimenting society,” Christopher Ansell and Martin Bartenberger address the diversity of three distinct logics of experimentation—controlled, evolutionary, and generative. They argue for a view of experimentation that expands the strict definition of control employed in scientific experimentation in an effort to find productive ways to experiment in society and learn valuable lessons. Evolutionary experimentation focuses on the systems level and thus applies less control. The goal is to increase variation in, for example, technical innovation. Generative experiments in social contexts aim to generate new design or problem-solving possibilities in, for example, policy development. Each of these logics produces different kinds of knowledge and they all have ethical consequences.

In “Moral experimentation with new technology,” Chapter 3, Ibo van de Poel examines moral experiments and moral learning. Arguing that we must see the introduction of new technologies as social experiments, he analyzes ways in which we can improve technological implementation. He draws from, and contributes to, moral philosophy. Van de Poel focuses on three modes of experiment: thought experiments, experiments in living, and social experiments. The three modes offer distinct possibilities for moral learning as well as pose different ethical consequences.

Frank Kupper’s Chapter 4, “The theatrical debate: experimenting with technologies on stage,” analyzes a specific experiment in creative ethical reflection. In the interactive theatrical debate performances of the project *Nano is Big* that took place in the Netherlands in 2010 and 2011, professional actors staged potential future situations of nanotechnology in society to an audience invited to share and explore their concerns and perspectives. Drawing on Dewey’s “dramatic rehearsal”, Kupper argues that this method of interactive theater acts as a rehearsal of the ways in which future technologies may emerge after introduction to society. He maintains that such theater projects function as simulation experiments—thought experiments—about new technologies that allow collective ethical reflection.

Lotte Asveld and Dirk Stemerding argue in Chapter 5, “Social learning in the bioeconomy: the Ecover case,” that the innovation process of sustainable technologies needs to incorporate an experimental approach that includes deliberate social and moral learning. They build on the case of a

new detergent marketed by Ecover as sustainably produced yet surprisingly criticized by those who saw the production method as unsustainable and contentious. The authors discuss how learning about the moral frameworks of the actors involved as well as about shared perspectives that may have emerged could have created a situation in which the conflict was avoided. More generally, moral learning that yields a shared understanding among all stakeholders can contribute to a more effective socially acceptable innovation trajectory for technologies.

In Chapter 6, Nicole A. Vincent and Emma A. Jane argue that the introduction of all new technologies are social experiments because of the uncertainties that their impact will have on individuals and the society in which we live. Their “Cognitive enhancement: a social experiment with technology” chapter reveals that the current debates on safety and ethical analyses of cognitive enhancement technologies frame the problem as one of medical safety, thereby neglecting the potential social consequences. They propose a methodology for the design and regulation of technologies to improve prediction, evaluation, and goal-setting and emphasize the need for critical self-reflection when embarking on the social experimentation.

In Chapter 7, “Living a real-world experiment: post-Fukushima imaginaries and spatial practices of ‘containing the nuclear,’” Ulrike Felt identifies ways in which people were subjected to uncertain technological, social, and scientific interventions in the experiment of Fukushima’s nuclear clean-up and containment. Probing shifting boundaries of the experimental space in this process, she describes the experimentation and learning that took place. Furthermore, she shows how redefining space was a political effort to re-instill confidence of the technical control of the nuclear for the Japanese population. Questioning the power of technical experts in evaluating acceptable risks, Felt suggests that we develop collective forms of social experimentation for new technologies.

René Umlauf’s Chapter 8, “‘Dormant parasites’: testing beyond the laboratory in Uganda’s malaria control program,” investigates the covert social experiment that took place with the implementation of the Rapid Diagnostic Test for malaria. This tool, designed to provide clear-cut results—simply positive or negative—was intended to be used by low-level health professionals in remote areas for easy and precise diagnosis without the need of a laboratory. In his ethnographic research, Umlauf observed how health workers interacted with patients and, questioning the validity of the new technology, reinterpreted the tests results. Moving the testing outside of the laboratory—the social experiment—transformed the practice of health workers as well as the context in which diagnosis takes place.

In Chapter 9, “Experimenting with ICT technologies in youth care: *Jeu-gdzorg* in the Netherlands,” Ben Kokkeler and Bertil Brandse study the introduction of information and communication technologies in the Dutch system of health and social services for minors. They explore various social experiments performed by administrators, for example, with new digital

client files, and by social workers, such as a creative project that encouraged client digital self-expression. Identifying the values held by different stakeholders in the organizations, the authors revealed ways in which, at times, those values conflicted. Moral dilemmas regarding trust and privacy, for example, emerged in the social experimentation thus, the authors suggest the sector incorporate efforts to develop responsible experimentation.

In the final Chapter 10, “Adversarial risks in social experiments with new technologies,” Wolter Pieters and Francien Dechesne accept the perspective that the introduction of new technologies are social experiments and note that many studies of technology implementation focus on unintentional consequences, in particular, on safety. The authors demonstrate the importance of including security and adversarial risks—especially prevalent in the cybertechnology domain—in analyses of new technologies. Illustrated with the distributed currency Bitcoin as an example, they demonstrate how this technology has been used by adversaries in ways unintended by the designers. They draw on actor-network theory to shed light on adversaries as human-technology networks. They advocate more serious consideration of potential adversarial uses of new cybertechnologies in the process of social experimentation.

From tacit to deliberate experimentation

The conceptualization of new technologies as social experiments offers an analytical and explanatory framework that increases our understanding of the introduction of new technologies into society. Furthermore, it enriches our analyses of and insight into the uncertainties that accompany technology implementation, and helps us conceptualize the learning processes in which these uncertainties gradually can be reduced. With relation to the ethics of technology, the framework enables us to recognize not only that ethical issues often become clear only after a technology has been introduced into society, but also why this is so. Moreover, our framework highlights that the moral frameworks, norms, and values with which we evaluate technologies continue to change in a moral learning process. One might think that the framework of new technologies as social experiments has, in particular, analytical and explanatory power for technologies with which there is relatively little operating experience. Examples include cybertechnology, biofuels, nanotechnology, synthetic biology, human enhancement, self-driving cars, and drones. However, our approach is also valuable for existing technologies, for example, those used in new contexts such as in the Rapid Diagnosis Test for malaria in remote health services or ICT technologies in social services. Also, in cases where the normal order of things is broken open, as in Fukushima, the framework clearly has an added analytical and explanatory value.

It should be noted that the conceptualization of new technologies as social experiments sheds light on the introduction of new technologies even when

the experimentation is tacit. Furthermore, by calling the introduction of a new technology a social experiment, characteristics may be revealed that are not seen by the actors themselves. In this sense, the conceptualization may also play an emancipatory role that can give actors a new perspective on their situation and help them find ways to improve the introduction of new technology.

The conceptualization of new technologies as social experiments offers a way to consider the acceptability of introducing a certain technology into society (cf. Van de Poel 2016) and may improve its introduction. Two main lessons follow from this edited volume. First is the need to transform tacit experimentation into deliberate experimentation for ethical and emancipatory reasons. Actors have the right to know that they are involved in an experimental process and this knowledge may provide new options for action (the emancipatory element). In addition, deliberate experimentation offers the opportunity for less socially harmful modes of experimentation in which better social and moral learning take place. Second, if we want to improve the introduction of new technology into society, we must think of that introduction not as one big social experiment, but rather as a staged process of experimentation. This experimentation, in fact, already starts before a technology is introduced, for example, in laboratory and thought experiments. Experimentation takes place not only before a technology is introduced in society and during the introduction phase, but also during intermediate stages, as seen in field experiments or living labs, and most well known in drug testing protocols. However, drug testing, as Vincent and Jane rightly point out in their contribution, limits testing to efficacy and harmful effects while neglecting potential social and moral problems. Drawing on the explorations in this book, we conclude that a staged model for moral experimentation with new technology would be extremely valuable. Ideally, this model would both recognize the introduction of new technology into society as a social and moral experiment, and create room for phases of limited and controlled experimentation before a technology is actually (fully) introduced into society. All in all, the framework examined in this volume points the way to improving our ability to evaluate social and moral consequences of new technologies.

Acknowledgments

This paper was written as part of the research program “New Technologies as Social Experiments,” which was supported by the Netherlands Organization for Scientific Research (NWO) under grant number 277-20-003.

References

- Anderson, Elizabeth S. 1991. “John Stuart Mill and Experiments in Living.” *Ethics* 102 (1):4–26. doi:10.2307/2381719.
- Ansell, Chris. 2012. “What Is a “Democratic Experiment”?” *Contemporary Pragmatism* 9 (2):159–80.

- Böschén, Stefan. 2013. "Modes of Constructing Evidence: Sustainable Development as Social Experimentation-The Cases of Chemical Regulations and Climate Change Politics." *Nature and Culture* 8 (1):74–96. doi:10.3167/nc.2013.080105.
- Boudette, Neal E. 2017. "Tesla's Self-Driving Tech Cleared in Crash." *New York Times*, January 20, 2017. www.nytimes.com/2017/01/19/business/tesla-model-s-autopilot-fatal-crash.html. Accessed March, 1 2017.
- Brown, James Robert, and Yiftach Fehige. 2017. Thought Experiments. In *The Stanford Encyclopedia of Philosophy (Summer 2017 Edition)*, edited by Edward N. Zalta. <https://plato.stanford.edu/archives/sum2017/entries/thought-experiment>
- Campbell, Donald T., and M. Jean Russo. 1999. *Social Experimentation, Sage Classics Series*. Thousand Oaks: Sage.
- Campbell, Donald T., and Julian C. Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research*. Chicago: McNally.
- Collingridge, D. 1980. *The Social Control of Technology*. London: Frances Pinter.
- Collins, H. M. 1985. *Changing Order: Replication and Induction in Scientific Practice*. London: Sage.
- Felt, Ulrike, Brian Wynne, Michel Callon, Maria Eduarda Gonçalves, Sheila Jasanoff, Maria Jepsen, Pierre-Benoît Joly, Zdenek Konopasek, Stefan May, Claudia Neubauer, Arie Rip, Karen Siune, Andy Stirling, and Mariachiarra Tallacchini. 2007. Taking European Knowledge Society Seriously. Report of the Expert Group on Science and Governance to the Science, Economy and Society Directorate, Directorate-General for Research, European Commission. Brussels: Directorate-General for Research, Science, Economy and Society.
- Gonzales, Wenceslao J. 2010. "Recent Approaches on Observation and Experimentation: A Philosophical-methodological Viewpoint." In *New Methodological Perspectives on Observation and Experimentation in Science*, edited by Wenceslao J. Gonzales, 9–48. La Coruna: Netbiblo.
- Gross, Matthias. 2010. *Ignorance and Surprise. Science, Society, and Ecological Design, Inside Technology*. Cambridge, MA: MIT Press.
- Gross, Matthias, and Holger Hoffmann-Riem. 2005. "Ecological Restoration as a Real-world Experiment: Designing Robust Implementation Strategies in an Urban Environment." *Public Understanding of Science* 14 (3):269–84. doi:10.1177/0963662505050791.
- Gross, Matthias, Holger Hoffmann-Riem, and Wolfgang Krohn. 2005. *Realexperimente: ökologische Gestaltungsprozesse in der Wissensgesellschaft, Science studies*. Bielefeld: Transcript.
- Hacking, Ian. 1983. *Representing and Intervening. Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.
- Hansson, Sven Ove. 2015. "Experiments before Science. What Science Learned from Technological Experiments." In *The Role of Technology in Science: Philosophical Perspectives*, edited by Sven Ove Hansson, 81–110. Dordrecht: Springer Netherlands.
- Herbold, Ralf. 1995. "Technologies as Social Experiments. The Construction and Implementation of a High-Tech Waste Disposal Site." In *Managing Technology in Society. The Approach of Constructive Technology Assessment*, edited by A. Rip, Th. Misa and J. Schot, 185–98. London and New York: Pinter.
- Hoogma, Remco, Rene Kemp, Johan Schot, and Bernhard Truffer. 2002. *Experimenting for Sustainable Transport. The Approach of Strategic Niche Management*. London: Spon.

- Kemp, René, Johan Schot, and Remco Hoogma. 1998. "Regime Shifts to Sustainability through Processes of Niche Formation: The Approach of Strategic Niche Management." *Technology Analysis & Strategic Management* 10 (2):175–95.
- Kemp, René, Derk Loorbach, and Jan Rotmans. 2007. "Transition Management as a Model for Managing Processes of Co-evolution towards Sustainable Development." *International Journal of Sustainable Development & World Ecology* 14 (1):78–91. doi:10.1080/13504500709469709.
- Krohn, Wolfgang, and Peter Weingart. 1987. "Commentary: Nuclear Power as a Social Experiment-European Political "Fall Out" from the Chernobyl Meltdown." *Science, Technology, & Human Values* 12 (2):52–58.
- Krohn, Wolfgang, and Johannes Weyer. 1994. "Society as a Laboratory. The Social Risks of Experimental Research." *Science and Public Policy* 21 (3):173–83.
- Latour, Bruno, and Steve Woolgar. 1979. *Laboratory Life: The Social Construction of Scientific Facts*. Beverly Hills: Sage.
- Levidow, Les, and Susan Carr. 2007. "GM Crops on Trial: Technological Development as a Real-world Experiment." *Futures* 39 (4):408–31. doi:10.1016/j.futures.2006.08.002.
- Martin, Mike W., and Roland Schinzinger. 1983. *Ethics in Engineering*. New York: McGraw-Hill.
- Mill, John Stuart. 1869. *On Liberty*. 4th edition, *Library of Economics and Liberty*.
- Morgan, Mary S. 2013. "Nature's Experiments and Natural Experiments in the Social Sciences." *Philosophy of the Social Sciences* 43 (3):341–57. doi:10.1177/0048393113489100.
- Popper, Karl R. 1945. *The Open Society and Its Enemies*. 2 vols. London: Routledge.
- Popper, Karl R. 1963. *Conjectures and Refutations; The Growth of Scientific Knowledge*. London: Routledge and Kegan Paul.
- Radder, Hans, ed. 2003. *The Philosophy of Scientific Experimentation*. Pittsburgh, PA: University of Pittsburgh Press.
- Rotmans, J., and D. Loorbach. 2008. "Transition Management: Reflexive Steering of Societal Complexity through Searching, Learning and Experimenting." In *Managing the Transition to Renewable Energy: Theory and Practice from Local, Regional and Macro Perspectives*, edited by Jeroen C. J. M. van den Bergh and Frank R. Bruinsma, 15–46. Cheltenham: Edward Elgar.
- Schwartz, Astrid. 2014. *Experiments in Practice*. Edited by Alfred Nordmann. Vol. 2, *History and Philosophy of Technoscience*. London: Pickering & Chatto.
- Sorensen, Roy A. 1992. *Thought Experiments*. New York: Oxford University Press.
- Stoker, Gerry, and Peter John. 2009. "Design Experiments: Engaging Policy Makers in the Search for Evidence about What Works." *Political Studies* 57 (2):356–73. doi:10.1111/j.1467-9248.2008.00756.x.
- Tesla Team. 2016. "A Tragic Loss." June 30, 2016. www.tesla.com/en_GB/blog/tragic-loss. Accessed March, 1 2017.
- Van de Poel, Ibo. 2009. "The Introduction of Nanotechnology as a Societal Experiment." In *Technoscience in Progress. Managing the Uncertainty of Nanotechnology*, edited by S. Arnaldi, A. Lorenzet and F. Russo, 129–42. Amsterdam: IOS Press.
- Van de Poel, Ibo. 2011. "Nuclear Energy as a Social Experiment." *Ethics, Policy & Environment* 14 (3):285–90. doi:10.1080/21550085.2011.605855.
- Van de Poel, Ibo. 2016. "An Ethical Framework for Evaluating Experimental Technology." *Science and Engineering Ethics* 22 (3):667–86. doi:10.1007/s11948-015-9724-3.

- Webster, Murray, and Jane Sell. 2007. "Why Do Experiments?" In *Laboratory Experiments in the Social Sciences*, edited by Murray Webster and Jane Sell, 5–23. Amsterdam: Elsevier.
- Winsberg, Eric. 2009. "A Tale of Two Methods." *Synthese* 169 (3):575–92. doi:10.1007/s11229-008-9437-0.
- Winsberg, Eric. 2015. Computer Simulations in Science. In *The Stanford Encyclopedia of Philosophy (Summer 2015 Edition)*, edited by Edward N. Zalta. <https://plato.stanford.edu/archives/sum2015/entries/simulations-science/>
- Wynne, B. 1988. "Unruly Technology: Practical Rules, Impractical Discourses and Public Understanding." *Social Studies of Science* 18 (1):147–67.

1 Control in scientific and practical experiments

Peter Kroes

Introduction

My aim in this chapter is to analyze and compare two types of experiments which I shall refer to as “scientific” and “practical” experiments. Scientific experiments are performed from a theoretical perspective; they have a purely epistemic aim and play a role in producing a specific kind of knowledge, namely knowledge about regularities in our world. Practical experiments are performed from a practical perspective; they serve practical interests and are concerned with how to act in the world in order to bring about a desired state of affairs. Whereas scientific experiments play a key role in understanding the world we live in, practical experiments play a key role in changing our world. In analyzing and comparing these two types of experiments, I will focus on the notion of control. This notion is crucial for the modern conception of scientific experiments. I will argue that the notion of control also plays an important but different role in practical experiments, and that the kind of knowledge produced in practical experiments is different from the kind of knowledge produced in scientific experiments.

I will employ a rather broad notion of the term ‘experiment’: experiments are not only performed with the aim of gathering a particular kind of knowledge about the world but also with the aim of changing the world. The first aim of experiments is generally acknowledged and traditionally associated with scientific experiments. However, many experiments performed in a technological context fall under my category of scientific experiments. Any technological experiment that is performed in order to learn regularities about how to act in or change the world is in my terminology a scientific experiment. They lead to what Hansson (2016) calls “action knowledge,” that is, knowledge of actions that if performed (adequately) in particular situations will lead to a desired result. Their aim is still knowledge of regularities, but of regularities of a special kind, namely concerning the effects of human actions. Practical experiments, by contrast, aim at bringing about a certain state of affairs in the world. The production of knowledge, be it of regularities or something else, comes subsidiary to realizing a practical result, that is, a desired state of affairs. Of course, these two aims do not exclude each

other and may be combined to give rise to a whole spectrum of experiments, from experiments with a purely epistemic aim to a practical aim and hybrid forms in between.

The reason why, in the context of this book, I opt for a broad interpretation of the notion of experiment is that when the introduction of new technologies in society is conceived of as a kind of experiment, we are dealing first and foremost with a situation in which we aim at changing the world to bring about a desired state of affairs. By calling such a situation an experiment, we are drawing attention to the fact that, apart from its practical aim, we also want to learn something from it. But exactly what is it that we want to learn? And what is it that we can learn? I will argue that for scientific and practical experiments, different issues come into focus when we analyze in more detail what it means to learn from experiments, especially in relation to controlling experiments.

At first sight, my distinction between scientific and practical experiments appears to run closely parallel to Hansson's (2015, 2016) distinction between epistemic and action-guiding experiments. Epistemic experiments aim at factual knowledge about the workings of the world, whereas an experiment is action-guiding according to Hansson if and only if the following two criteria are satisfied (2016, 617):

- 1 The outcome looked for should consist in the attainment of some desired goal of human action, and
- 2 the interventions studied should be potential candidates for being performed in a nonexperimental setting in order to achieve that goal.

The first condition of action-guiding experiments is about changing the world and is therefore very much in line with my notion of practical experiments. The second, however, shifts the focus from bringing about a particular state of affairs to *learning* how to bring about that state of affairs in order to be able to apply what is learned in nonexperimental settings. For Hansson, this implies that the aim of action-guiding experiments is to produce knowledge of regularities about how to act in order to achieve a goal; that knowledge may then be applied in nonexperimental settings. This is a rather strong requirement that ties experimental learning to learning of regularities and excludes, for instance, learning from experiments on the basis of analogies with nonexperimental settings. Apart from this focus on learning of regularities, there is yet another reason why action-guiding experiments are different from practical experiments. Hansson's definition of action-guiding experiments refers to the distinction between experimental and nonexperimental settings. Although he does not elaborate on this distinction, it appears to refer roughly to a distinction between artificial, controlled, laboratory, etc. situations and real-life situations. Starting from this distinction, practical experiments are always real-life experiments because they aim at realizing a desired state of affairs in the real world and therefore

have to be characterized as nonexperimental. Thus, with regard to practical experiments, the distinction between experimental and nonexperimental settings will have to take on a different meaning (see below).

Clearly, Hansson's distinction between epistemic and action-guiding experiments is a distinction that falls squarely within what I have called scientific experiments: the aim of action-guiding experiments is primarily the production of knowledge of regularities (albeit of a particular kind), just as is the aim of epistemic experiments. Hansson analyses action-guiding experiments from a theoretical perspective because he is interested in the production of scientific knowledge. For him, scientific knowledge includes both factual and action-guiding knowledge, and therefore both epistemic and action-guiding experiments are important when studying the role of experiments in science. He observes, and I fully agree with him, that nevertheless action-guiding experiments have been almost completely neglected in historical and philosophical studies of experiments in science.

The distinction between scientific and practical experiments is more wide-ranging than Hansson's distinction between epistemic and action-guiding experiments. Hansson's definitions of epistemic and action-guiding experiments tie experiments in general to the production of a particular kind of knowledge, namely knowledge about regularities in the world. He emphasizes this aspect in his definition of experiments in general (Hansson 2016, 616):

...by an experiment I will mean a procedure in which some object of study is subjected to interventions (manipulations) that aim at obtaining a predictable outcome or at least predictable aspects of the outcome. Predictability of the outcome, usually expressed as repeatability of the experiment, is an essential component of the definition. Experiments provide us with information about regularities, and without predictability or repeatability we do not have evidence of anything regular.

For action-guiding experiments these regularities take more or less the form of conditional imperatives: "If you want to achieve Y, do X." Niiniluoto (1993) refers to these conditionals as "technical norms;" knowledge of such norms constitutes action knowledge or means-ends knowledge. It is this rigid coupling of experiments in general to regularities, predictability, and repeatability that sits uncomfortably with the idea of the introduction of new technologies in society as experiments. These experiments in society are not about learning of regularities but rather about learning how to achieve a particular state of affairs by intervening in the world. Therefore, we need a broader conception of experiments but not one that is too broad: not every goal-directed intervention in the world qualifies as an experiment.

What is missing in habitual, nonexperimental, goal-directed interventions in the world to qualify as experiments is the element of learning. In particular, I will take a goal-directed intervention in the world to be an

experiment if, apart from the intention of bringing about its goal, there is also the subsidiary intention to use this intervention to learn how to reach that goal. This may involve not only learning about the means to achieve the goal but also about how to adjust the goal in view of the available means. The intervention must go hand in hand with a form of systematic inquiry. In the pragmatist spirit of Dewey, this form of inquiry may be characterized as involving an indeterminate situation in which we are “uncertain, unsettled, disturbed” (1938, 105) and in which “existential consequences are anticipated; when environing conditions are examined with reference to their potentialities; and when responsive activities are selected and ordered with reference to actualization of some of the potentialities, rather than others, in a final existential situation” (p. 107). What is interesting about this view is that it connects inquiry and learning to responsive activities (interventions) and existential aspects. This learning by experimentation is not primarily of an intellectualistic type that results in action knowledge of (abstract) regularities; it is learning about how to adapt our interventions in the light of the goals pursued or how to adapt our goals in the light of the available means for achieving them. This kind of learning does not necessarily lead to action knowledge that can easily be expressed as the regularities considered above. This is because the justification of action knowledge in the form of regularities presupposes a form of experimental control that in many practical experiments is not available. In particular, I will argue that this is not the case for experiments involving the introduction of new technologies in society. First, however, it will be necessary to have a closer look at the issue of control in scientific experiments.

Control and scientific experiments

The aims of the types of experiments discussed in this section are the same: to provide information or evidence for the production of knowledge of regularities. This may, in the words of Hansson, be factual or action knowledge and therefore both his epistemic and action-guiding experiments fall under this heading. I will analyze the role of control in these types of experiments in three different domains, namely the physical sciences, the social sciences, and the technological or engineering sciences.

I will start with a brief look at the various kinds of experiments performed in the physical sciences.¹ Depending on the locus of the experiment, we may distinguish between thought experiments which take place in the mind, computational/simulation experiments in computers, and experiments in the real world; the last may take place in a laboratory setting or in the wild (I will mainly focus on laboratory experiments below). The epistemic role of thought and computational experiments has been and still is contested, one of the issues being whether or not they lead to new knowledge about the world (see, for instance (Mach 1976 (1897); Kuhn 1977)). This is not the case for experiments performed in the real world; the epistemic relevance of the

outcomes of these experiments is not disputed. Another distinction, which cuts across this one, is between qualitative and quantitative experiments. Qualitative experiments show the existence of particular phenomena, such as the quantum-interference of electrons in the double slit experiment. Quantitative experiments focus on quantitative relations between physical quantities and generally make use of the principle of parameter variation (here we may think of Boyle's experiments with gases to show the relation between pressure and volume that bears his name). Another distinction is based on the relation between experiment and theory. Depending on whether an experiment is intended to explore phenomena without the guidance of theory or is intended to test a theory, experiments may be divided, only schematically of course, into exploratory and hypothesis testing.

There are at least two reasons for performing laboratory experiments in the physical sciences:

- 1 experiments enable the study of spontaneously-occurring physical objects and phenomena under conditions that do not occur spontaneously in the world; and
- 2 conditions may be created for the occurrence of physical objects and phenomena that do not occur spontaneously in the world; these objects and systems can therefore be studied only under experimental conditions.

Thus, the experimenter creates the appropriate conditions for studying physical phenomena and objects, and may even create those phenomena and systems themselves, by creating the appropriate conditions for their occurrence. These conditions are human-made and therefore artificial; that, however, does not imply that the physical objects and phenomena themselves are human creations and thus artificial (Kroes 2003).

What I am particularly interested in here is the extent to which the experimenter has control over the system on which the experiment is performed. In this respect, there are significant differences between thought, computer, and laboratory experiments. In thought experiments, the physical system under study and the conditions under which it is studied are created in the imagination, and the experimenter has, in principle, total control over the system and conditions; (s)he is even in a position to study physical systems under conditions that cannot be realized in the world (for instance, by assuming the validity of imaginary physical laws). However, there are restrictions on what kind of systems can be studied fruitfully in thought experiments. These restrictions find their origin in the reasoning powers of the experimenter; it is, for instance, no use to perform thought experiments on systems that are so complex that it is not possible to draw any interesting conclusions about their behavior.

More or less, the same applies to computer experiments (simulations). *Prima facie*, the experimenter appears to have almost unlimited freedom to

define the target system and its conditions. However, also in this case there are restrictions; the freedom of the experimenter is not unlimited due to constraints imposed by the computational device (computer). The computational power of the device limits the kinds of target systems that may be simulated and therefore also limits the control of the experimenter over the target system and its conditions. These limits on the experimenter's control find their origin in the technological limits (related to hardware and software) of the computational device.

In laboratory experiments, technological constraints determine the control of the experimenter over what kind of system may be studied under what kind of experimental conditions. It is not possible to study objects or systems that contradict the laws of nature, as in thought and computational experiments. In a limited case where the scientist has no control whatsoever over the object of study and the conditions under which it is studied (e.g., the occurrence of a supernova), the scientist is dependent on Nature to perform the experiment so as to make it possible to study the system or phenomenon (see Morgan's notion of Nature's experiment (Morgan 2013)). This does not mean that the scientist is condemned to the role of a totally passive observer and that no issues of control emerge. The observation of naturally-occurring phenomena may require the control of scientific instrumentation, for example, for measurements, during the observations.

Note that in these various kinds of experiments, the control of the experimenter stretches no further than control over the physical system studied and the conditions under which it is to be studied. The experimenter does not have control over the behavior of the system under those conditions; that is, the experimenter has no control over the outcome of the experiment.² If that would be the case, the performance of the experiment would lose much of its rationale: why perform an experiment if the outcome can be controlled and consequently constructed and predicted in advance?³ In that case, it would be difficult to explain how a scientist could learn anything about nature from experiments.⁴ Here, the theoretical perspective of these experiments shows itself: when it comes to gathering information or evidence about the world during an experiment, the experimenter is forced into the role of a passive observer, of someone who is a spectator watching nature perform its play.⁵ The role of the experiment is reduced, so to speak, to offer the scientist as a spectator the place from which (s)he can watch what (s)he is interested in.

I now move to experiments in the social sciences.⁶ Walker and Willer (2007, 25) distinguish between two "fundamentally different types of experiments" in the social sciences, namely empiricist and theory-driven experiments, each of which has its own method. This distinction runs more or less parallel to the distinction between exploratory and hypothesis-testing experiments above. Empiricist experiments aim at making generalizations from observations, whereas theory-driven experiments aim at testing theories. The examples of experiments they discuss fall into the category of laboratory experiments. They do not refer to thought experiments or computational

experiments; it appears that at least the latter also have come to play an important role in the social sciences. In the following, I will focus on laboratory experiments only.

What emerges from the discussion of laboratory experiments in the social sciences in part I of Webster and Sell (2007b) is that it is all about control. According to Webster and Sell (2007a, 8), “a study is an *experiment* only when a particular ordering occurs: when an investigator controls the level of independent variables *before* measuring the level of dependent variables.” Walker and Willer (2007, 25) state that a laboratory experiment is “an inquiry for which the investigator plans, builds, or otherwise controls the conditions under which phenomena are observed and measured...”. According to Thye (2007, 66), “The confluence of three features makes experimental research unique in scientific inquiry: random assignment, manipulation, and controlled measurement.” All of these three features are about control; the notion of a control group is based on randomization and the “first and most straightforward type of control is the *control group*” (Thye 2007, 79). Random assignment of persons in the experimental and control group is intended to eliminate (which is a specific form of control) the influence of spurious variables. This focus on controlling variables in laboratory experiments is of course closely related to the complexity of social phenomena given the number of potentially relevant variables for the phenomenon under study. To be able to draw reliable conclusions about the relation between the independent variables and the dependent variables, all other variables that may have an influence on the dependent variables must be kept constant.

The control of social phenomena in laboratory experiments has advantages and disadvantages. According to Webster and Sell (2007a, 11):

The greatest benefits of experiments reside in the fact that they are artificial. That is, experiments allow observation in a situation that has been designed and created by the investigators rather than one that occurs in nature. Artificiality means that a well-designed experiment can incorporate all the theoretically presumed causes of certain phenomena while eliminating or minimizing factors that have not been theoretically identified as causal.

Because the conditions are artificial and controlled, experiments may be replicated in order to test the internal validity of the outcomes. The disadvantage of artificiality is that it raises issues about generalizability and external validity: to what extent may the results from experiments be generalized to situations outside the laboratory settings? Internal and external validity are more or less in tension with each other—the more the conditions in the laboratory are artificial, that is, are different from the conditions in natural settings, the better it is for establishing internal validity, but the more problematic it may be for external validity (that is, for situations outside the laboratory) (Walker and Willer 2007, 51).

Finally, let us turn to experiments and their role in producing technological knowledge. To begin, I will make a schematic distinction between experiments *in* technology and experiments *with* technology. With experiments in technology, I mean experiments performed within the context of designing, developing, and producing a technology or technological artifacts. These are the experiments performed mainly by scientists, engineers, and technicians in R&D laboratories, on the shop floor, or elsewhere; they typically involve physical objects or processes that are supposed to perform technical functions. Experiments with technology are experiments in which technologies are implemented in real-life situations in order to achieve some practical goals during which the implementation is closely monitored for learning purposes. In the case of experiments with technology, it may be rather difficult to determine who is doing the experiment; such experiments may involve engineers, technicians, and users, but the ‘experimenter’, if any is clearly defined at all, is usually a company or a governmental institution. Admittedly, this distinction is schematic and it may be very problematic to apply in particular cases (for instance, when engineers test a new design among a user group for feedback). For my purposes, however, it makes sense to differentiate between these two cases because it helps to illustrate the differences in scientific and practical experiments and the kind of knowledge they produce. Here I will focus on experiments in technology; a discussion of experiments with technology will be postponed to our discussion of practical experiments.

Schematically, experiments in technology fall into two classes, namely scientific and practical experiments. Whereas scientific experiments in technology aim at the production of knowledge of technically relevant regularities, practical experiments aim primarily at the production of functioning (prototypes of) technical artifacts. Here we are mainly interested in scientific experiments. It should be noted that the nature of both kinds of experiments in technology and of the technological knowledge produced on their basis are philosophically a rather under-investigated topic.⁷ One of the best studies in the field is Vincenti’s (1990) *What engineers know and how they know it*. On the basis of detailed historical case studies, he distinguishes between various kinds of technological knowledge and the way they are produced. His analysis shows that, similar to the situation in physical experiments, controlled experiments play a major role in scientific experiments in technology. Given that much of science and technology have currently merged into technoscience, this conclusion is not very surprising. What appears to be specific for controlled experiments within a technological setting is that they are more often action-guiding experiments that give rise to action knowledge or knowledge of means-ends relations. This action knowledge is knowledge of regularities that have predictive force and are based on repeatable experiments. According to Hansson, its justification is based on the following procedure for experimental comparison (Hansson 2016, 626):

If you want to find out whether you can achieve Y by doing X, both do and refrain from doing X under similar circumstances, and find out whether the cases with X differ from those without X in that Y occurs more often or to a higher degree.

In order to make sure that X is the relevant causal factor for bringing about T, a control experiment is performed. So, just as control plays a crucial role in scientific experiments leading to factual knowledge, control plays a crucial role in technological experiments leading to action-guiding knowledge.

The control paradigm for scientific experiments

One common feature of scientific experiments in these various domains is the focus on control.⁸ In these domains, the experimenter strives for (total) control over the experiment. According to Schiaffonati and Verdicchio (2014, 363) an experiment can be seen as a *controlled experience*, namely “as a set of observations and actions, performed in a controlled context, to support a given hypothesis.” But what does control over the experiment mean? To answer this, we have to distinguish carefully between two different forms of control. One form concerns control over the kind of system on which experiments are to be performed. In laboratory settings, the kind of experimental system is usually designed and created by the experimenter; taking into account technological, social, ethical and other constraints, she decides on the basis of her research interests how to configure the system on which the experiments will be performed. The aim of the experiment is to study the behavior of this particular kind of experimental system which means that during an experiment, the system itself has to remain the same, that is, has to remain an instance of that kind of experimental system. So, once an experiment has been defined and actually started, the experimental system can no longer be changed. This means that the experimental system is no longer under control of the experimenter in the sense that any change in that kind of system implies either a termination of the original experiment and the start of a new kind of experiment, or simply a termination of the original experiment.⁹ However, this does not preclude that the experimenter may have full control over when to start and when to terminate a particular experiment.

The other form of control pertains generally to the conditions under which the experimental system is to be studied, and in particular to the interaction between the experimental system, once it is put in place, and its environment. This kind of control may take different forms. For instance, the experimenter may be interested in how the experimental system, starting from a given initial state, behaves in isolation from its environment. In that case, the experimenter brings the experimental system in the initial state through controlled interaction between the system and its environment. Once it is in this state, she closes off the system from all relevant

interaction with the environment.¹⁰ In experiments with independent and dependent variables, the experimenter sets the independent variable through controlled interaction with the environment of the experimental system. Finally, in experiments involving control groups, the spurious variables not under direct control are controlled by following statistical procedures in composing the experimental and control groups (this is an aspect of the set-up of the experimental system itself)¹¹ and by assuming that the interaction of the two groups with their environments during the experiment is the same *except for* the controlled intervention in or treatment of the experimental group. Under those conditions, it is possible to apply a statistical variant of Mill's method of difference and conclude that the observed (statistical) difference in the dependent variable between the experimental and control group is the effect of the controlled intervention or treatment.¹²

The basic assumption about control of experiments underlying the ideas above is that the experimenter may control the experiment either through intervention in the environment of the experimental system or through intervention in the experimental system itself. The notion of intervention has a clear meaning: the experimenter is not part of the system on which the experiment is performed nor of its environment. The experimenter operates from a center of control which is not part of (outside) the experimental system and its environment. From this vantage point, the experimenter controls the experimental system and its interaction with its environment, which means that she has control over setting the independent variables and over starting and terminating the experiment. I will refer to this idea as the *control paradigm for experiments*.

Of course, the idea of full control underlying this traditional control paradigm is based on an idealized view of experiments. In actual scientific practice, it may be difficult to realize full control of experiments in the two senses discussed above. Even in well-prepared physical experiments within the "secure" walls of laboratories, it may be difficult to eliminate (control) all interfering interactions. The ideal is to eliminate interferences as much as possible in order to "purify" the phenomenon under study. If interferences cannot be fully eliminated, then it is very often possible to take the effect of the occurring interferences into account in interpreting the results of the experiment.

One type of experiment in which the idea of control becomes problematic is the case of field experiments. According to Schwarz and Krohn (2011, 123), "[p]erhaps the most striking feature of field experiments is that they deal with objects 'outside,' in an uncontrolled environment." Field experiments, in agriculture for instance, are often performed precisely because one is interested to learn more about the behavior of the experimental object or phenomenon under uncontrolled conditions. This means that in field experiments, the experimenter tries to expose the experimental object or phenomenon to the conditions in the wild. This, paradoxically, presupposes a form

of control on the part of the experimenter, namely to maintain conditions that correspond as much as possible to the conditions in the wild. Whenever necessary or desirable, it may be possible, moreover, to register or measure and (partly) control the uncontrolled interactions in order to analyze their effect on the experimental system. The fact that field experiments are conducted in an uncontrolled environment does not per se imply that this environment is uncontrollable.

The most important feature of the kind of knowledge produced on the basis of these various controlled experiments is that they “provide us with information about regularities,” and with regularities come predictability and repeatability, since “without predictability or repeatability we do not have evidence of anything regular” (Hansson 2015, 4). This regularity is a feature of factual knowledge as well as action knowledge and is based on and guaranteed by the control paradigm. Because the experimenter is in control of the system studied and of the interventions in the system from the environment, it is possible to repeat the experiment and make predictions about what will happen. For action-guiding experiments, this means that under suitable conditions the regularities learned may be transferable to nonexperimental settings in order to make certain (or highly probable) that the result of a human intervention will lead to a desired outcome. It is precisely this aspect of gaining knowledge about regularities that is put into question by Schwarz and Krohn in their analysis of field experiments. Indeed, field experiments may raise questions about the extent to which the experimenter is able to satisfy the conditions of the control paradigm. If they are right that field experiments show features of “individuality, uniqueness, contingency, instability, and also potentially lack of safety,” then what is learned from field experiments may not be transferable and generalizable to other situations (2011, 123 and 130). Whether or not these features of field experiments indeed undercut the idea that they may lead to information of regularities will depend to a large extent upon the nature and context of the specific field experiment involved. Suffice it here to remark that many kinds of field experiments are considered to be scientific experiments and lead to knowledge of more or less transferable regularities.

To summarize, scientific experiments, including scientific experiments in technology, aim at knowledge of regularities, and to make this possible, these experiments must satisfy the conditions of the control paradigm. This implies that the notions of experimenter and of intervention in the experimental system are well-defined. In the next section, I will argue that the control paradigm may become problematic for experiments with technology. I will focus on specific experiments with technology, namely those in which the experimental systems are sociotechnical systems. I will consider these experiments with technology to be practical experiments, that is, experiments that aim primarily at changing the world but that also involve an element of learning.

Control and practical experiments

If we consider the introduction of new technologies in society as practical experiments, then these experiments may involve experimental systems of a special kind, namely sociotechnical systems.¹³ Examples of (complex) sociotechnical systems are infrastructural systems such as electric power supply systems or public transport systems. Their behavior is significantly affected by technical components, but the functioning of the systems as a whole depends as much on the technical components as on the social components (legal systems, billing systems, insurance systems, etc.) and the behavior of human actors. Sociotechnical systems are hybrid systems consisting of various elements, such as natural objects, technical artifacts, human actors, and social entities such as organizations and the rules and laws governing the behavior of human actors and social entities. Within sociotechnical systems, all of these elements have to be attuned to each other to guarantee a proper functioning of the system as a whole. This means that the introduction of new technologies in these systems with the practical aim of improving their operation and functioning will engender changes in their social subsystems. Therefore, the introduction of a new technology is not simply a practical experiment with technology but with a sociotechnical system.¹⁴

In the following I will argue that (1) the control paradigm, in particular, the notion of intervention, becomes problematic when dealing with practical experiments on sociotechnical systems, and (2) learning about control may take place during these experiments, but this is not necessarily learning of regularities as in the case of scientific experiments.

To start with the applicability of the control paradigm, elsewhere I (Kroes 2009) have argued that the traditional engineering design paradigm is no longer a suitable basic framework for the design and control of sociotechnical systems. This design paradigm is based on three pillars. First, it assumes that it is possible to separate clearly the object of design from its environment. Second, it addresses exclusively the design of hardware (the manual is more or less taken for granted). What is designed is a material technical object. Third, it assumes that the behavior of the systems designed can be fully controlled by controlling the behavior of its parts. Given that the technical artifact is made up of physical parts, this control amounts to the control of the behavior of these physical parts through a set of control parameters. Similar to the control paradigm for experiments, the traditional design paradigm is based on the assumption that the designer operates from a control center outside of the system to be designed.

For sociotechnical systems, the assumptions underlying the traditional design paradigm do not apply, and the same aspects that undermine the traditional engineering design paradigm for these systems also undermine the traditional control paradigm for experiments with sociotechnical systems. Firstly, there is the problem of where to draw the line between the experimental system and its environment. If the function of a system is taken to

be that which gives the system cohesion, then it is rather obvious that all elements relevant to the functioning of a system should be included. But how is the function of, for instance, an electric power supply system to be defined? Different actors may have different views on this and may therefore have different opinions on what constitutes part of the experimental system and what belongs to its environment. In whatever way the boundaries of the experimental system will be drawn, it is clear that, since we are dealing with sociotechnical systems, by definition human agents and social institutions will be integral parts of the experimental system. This means, secondly, that the nature of the experimental system to be controlled changes. Its inner environment no longer consists of material objects only. The control of these systems not only involves the control of technical elements but also of social elements. However, the behavior of human agents and social institutions cannot be controlled in the same way that the behavior of technological systems can be controlled. In so far human agents perform certain operator roles in sociotechnical systems, one may try to control their behavior explicitly by developing, for example, protocols that they are required to follow in their operator roles. Or one may try to control their behavior in more implicit ways by conditioning their working environment such that it elicits various forms of (desired) behavior. Nevertheless, there appears to be an essential limit to these forms of control of the behavior of human agents in sociotechnical systems, a limit that is related to the nature of human agency. The human agents who fulfill the operator roles in sociotechnical systems remain *autonomous* agents whose behavior, for *reasons of their own*, may deviate from prescribed protocols in an uncontrollable way. For instance, operators may decide to start a strike to support demands by their labor union, or deviate from a protocol because in a certain situation, it raises moral issues for the human agent/operator involved. Of course, the presence of this essential limit in controllability of sociotechnical systems may raise important moral issues when performing experiments on this kind of system.¹⁵ Finally, various actors from within the sociotechnical system, with their own interpretations of the function of the system and their role in realizing it, may try to change, control, or re-design parts of the system from within. As a result, the idea of controlling the experimental system from a control center outside of it becomes highly problematic.¹⁶

The main conclusion to be drawn is that the notions of control of and intervention in the system on which the experiment is performed will lose its standard meaning when we consider complex sociotechnical systems. Thus, the control paradigm that underlies scientific experiments is not applicable to experiments with sociotechnical systems. This is one of the features that makes practical experiments on sociotechnical systems different from scientific experiments.¹⁷

A second feature, related to the foregoing, that sets this kind of practical experiment apart from scientific experiments is the kind of learning that takes place during these experiments. The primary aim of a practical

experiment is to bring about a desirable outcome (e.g., to improve the functioning of sociotechnical systems), whereas according to Hansson (2015, 5), the outcome of epistemic experiments leading to factual knowledge “need not coincide with anything that a sensible person would wish to happen except as part of the experiment itself.” But any intervention to bring about a desired outcome that is to be considered a practical experiment must also have the objective of learning something about the intervention and/or the outcome. This learning, however, comes secondary to, and is intended to serve, bringing about the desired outcome; it is therefore mainly geared to managing problems, obstacles, and uncertainties that arise in attempts to realize this outcome. By contrast, learning in scientific experiments is geared to the production of knowledge of reliable regularities about the behavior of the experimental system and about interventions in that system. This different orientation in learning and the related fact that the control paradigm for experiments is not applicable appears to indicate that we are dealing here with a different kind of learning.

How can we characterize the kind of learning that takes place in practical experiments? This kind of learning is different from the learning that goes on in controlled scientific experiments and results in knowledge of regularities on the basis of which it is possible to predict the effects of planned interventions in sociotechnical systems and to redesign them in order to achieve the desired result. Because in experiments on sociotechnical systems there is no clearly defined center of control or of experimentation, it is difficult to give a clear-cut answer to the question “who learns from these experiments?” Different stakeholders are involved, each with their own interests and agendas, who act and react to the behavior of other stakeholders and learn lessons from these actions and reactions in the light of the goals they pursue. Metaphorically, one could say that the system as a whole learns or that we are dealing here with a form of distributive learning among all the stakeholders involved.

The question of what is learned, or what kind of knowledge is produced in these experiments if not primarily factual or action knowledge of regularities, appears even more difficult to answer. Of course, by closely monitoring the introduction of a new technology, one may learn about how the new technology is received and used, what social and institutional changes accompany its introduction, what moral issues arise, and how they are dealt with by various stakeholders, for example. This learning may result in an historical narrative, or a set of historical narratives considering the various stakeholders involved, each of which will tell its own story with its own moral. These narratives will contain detailed factual knowledge of what happened during the experiment, that is, knowledge of historic facts. But how do we get from knowledge of these historic facts to transferable knowledge that will be somehow applicable or useful in other experiments of more or less the same kind?¹⁸ According to Schwarz and Krohn, a similar problem presents itself with regard to learning from field experiments because of their

individual, unique, and contingent features. In order to draw transferable lessons in these cases, it will be necessary to distinguish between features that make an experiment unique and those that make it similar to other experiments. If that would be possible, then on the basis of observations of many experiments with sociotechnical systems, one could try to derive inductively generalizable knowledge that may be useful for action-guiding purposes.¹⁹

A clear conceptual framework for analyzing the learning that takes place during practical experiments—who learns what and how—is still missing. Various ways of analyzing or characterizing forms of learning or knowledge that have been proposed in the literature may be relevant for exploring this issue further, such as “learning by doing,” “tacit knowledge,” or “bricolage.”²⁰ The conclusion to be drawn here is that without a further explication of this essential learning element in the characterization of practical experiments, the notion of practical experiments itself remains rather problematic.

Discussion

The outcome of my analysis is that, if the introduction of a new technology in society is interpreted to be an experiment, then considering control over the experiment, we are dealing with a kind of experiment that is different from scientific experiments in the natural, social, and technological sciences. In closing, let me point briefly to a conception of experimentation that may be more appropriate to apply to the introduction of new technologies in society. Ansell (2012) refers to such experiments as “design experiments” (Ansell 2012, 16364):

Design experimentation starts with the presumption that the world is a messy place and that experiments will not be able to isolate the effects of single variables. In a design experiment, the experimenter presumes that the experiment will interact with the totality of the setting in which the experiment is conducted. The focus of a design experiment is not to definitely accept or reject a hypothesis, but rather to iteratively refine the intervention (design-redesign cycles). [...] Design experiments do not create a sharp distinction between researchers and subjects; instead, the practitioners often become experimenters. [...] In other words, design experiments do not fully control the conditions in which the experiment occurs, as laboratory experiments attempt to do.

Thus, in design experiments, there is no clear-cut distinction between the experimenter and the experimental system, and nobody controls the experimental system fully. Such a notion of experiment may be more fruitful when technological innovations are interpreted as social experiments. Ansell (2012, 172) remarks that “...this approach to experimentation loses the powerful mode of verification associated with controlled experiments (and for

this reason, some might argue that it is not experimental at all).” Indeed, because design experiments do not satisfy the conditions of the control paradigm, they do not qualify as scientific experiments. Apart from the loss of control, there is yet another reason to question whether design experiments are scientific experiments. The focus of a design experiment is to redesign the intervention on the fly. In the introduction of technological innovations in society, this means redesigning the technological innovation (and possibly redesigning its social context). That amounts to changing the kind of system on which the experiment is performed. However, by changing the experimental system during the experiment, it is no longer clear what kind of experiment is being performed, which means that it is no longer clear what regularities about what kind of system may be learned during the experiment. The foregoing, however, does not imply that design experiments are not experiments at all. They come close to our characterization of practical experiments, of experiments that aim at changing (redesigning parts of) the world and in which learning by doing or by trial and error takes place. The kind of knowledge gained from such experiments is not action knowledge of a well-defined experimental system, that is, knowledge of regular cause (intervention) and effect (change in experimental system) relations that makes it possible to control the experimental system and to move it in the desired direction by planned intervention with predictable outcomes. It is more learning about how to “iteratively refine the intervention” such that the design intervention has the desired effect.

Let me end with a general remark about the limits of control of the effects of human action in complex practical experiments, such as experiments with sociotechnical systems that involve many different actors. In her book *The human condition*, Arendt (1958) analyzes “the threefold frustration of action – the unpredictability of its outcome, the irreversibility of the process, and the anonymity of its authors” (p. 220). She argues that when there is a plurality of actors, the final outcome of human actions becomes unpredictable, irreversible, and cannot be attributed to a specific actor or set of actors. In other words, the outcome of human action appears to take on a life of its own which is not under control of any one actor or even of all the actors together. According to Arendt, the meaning of action:

reveals itself fully only to the storyteller, that is, to the backward glance of the historian, who indeed always knows better what it was all about than the participants. (...) Even though stories are the inevitable results of action, it is not the actor but the storyteller who perceives and “makes” the story.

(p. 192)

If the full meaning of action comprises what we may learn from its outcome, then we are back to what we already observed with regard to learning and practical experiments, namely storytelling. Whether there is just one story

to be told, as Arendt suggests, or many, remains a topic for further research. The final conclusion is not that we cannot learn from practical experiments with new technologies in society, but rather that whatever kind of learning is involved in such experiments, we should not mistake it for the kind of learning that takes place in scientific experiments, where learning is all about regularities and control.

Notes

- 1 The following discussion of the role of control in natural and social sciences is taken, with some minor modifications, from Kroes (2016).
- 2 Note that this may not be true for thought experiments. Think of Newton's famous bucket experiment; Newton simply assumed what would be the outcome of his thought experiment, namely, that the surface of a bucket of water rotating in absolute space would be curved. Mach questioned whether that would really be the case. This feature of thought experiments is closely connected to their contested epistemic status mentioned above.
- 3 This relates directly to the remark above that the physical phenomena themselves, although created in the experiment, are themselves natural phenomena and not artificial.
- 4 The idea that there is no control over the outcome of an experiment has been critiqued from social constructivist quarters. For an overview of the debate about the epistemic status of experimental results, see Franklin and Perovic (2015). They defend the idea that there is an epistemology of experiment and that, therefore, scientists may learn from experiments. Of course, given the complexity of many modern experiments, it may often be very difficult to establish what exactly may be learned from them, but that does not undermine the fact that in principle and de facto it is often possible.
- 5 One of the meanings of the Greek verb *θεωρεῖν* is being a spectator. Following up on the previous note, in actual practice it may be difficult, due to the complexity of many experiments, to ascertain what in the things observed belongs to the stage setting (the experimental set-up) and what to the play itself. But the underlying idea is that it is always in principle possible to distinguish between these two aspects (I leave aside here issues about the interpretation of quantum mechanical measurement).
- 6 The following is mainly based on the discussion of experiments in the social sciences in part I of *Laboratory Experiments in the Social Sciences* edited by Webster and Sell (2007b).
- 7 I agree with Nordmann (2016) that too often "technology is viewed through the lens of science." As a consequence, philosophical analyses of learning and knowledge production in technology focus mainly on scientific experiments, that is, experiments producing knowledge of regularities, with an almost total neglect of the knowledge production in practical experiments in which, for instance, technical artifacts are made; knowledge production in technology is not confined to experiments in the engineering (or technological) sciences, but also takes place in engineering itself, that is, in the design, development, making, repairing, etc. of technical artifacts.
- 8 The following discussion of the control paradigm for experiments is taken, with some modifications, from Kroes (2016).
- 9 Of course, this raises intricate problems about identity criteria for experimental systems and for experiments; in the present context, the precise nature of these identity criteria, if there are any, is not of interest; what matters here is

the contrast between control over the experimental system and control over the conditions under which experimental systems are studied.

- 10 Of course, any experiment involves making observations of (measurements on) the experimental system so it cannot be totally isolated from its environment.
- 11 “Control over the interaction” here does not necessarily mean that the experimenter controls or knows who is part of the experimental group and who of the control group; in order to avoid experimenter’s bias, it may be necessary to apply double-blind procedures.
- 12 “If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or a necessary part of the cause, of the phenomenon.” John Stuart Mill, *A System of Logic*, Vol. 1. 1843. p. 455.
- 13 The following discussion of experiments with sociotechnical systems is taken from Kroes (2016).
- 14 For more information about sociotechnical systems, see Vermaas et al. 2011 and Franssen and Kroes, 2009.
- 15 Note that these limits to the control of the behavior of agents performing operator roles in sociotechnical systems do not make it impossible to perform controlled experiments in the social sciences (as discussed in Section 2); the aim of controlled experiments in the social sciences is precisely to study the autonomous behavior of agents under controlled circumstances (just as the aim of physical experiments is to study the “autonomous” behavior of physical systems under controlled circumstances). Nevertheless, an essential limit in controllability, due to the fact that human beings are part of the experimental system, may manifest itself in experiments in the social sciences. The idea that the experimenter is not part of the environment of the experimental system may be difficult to realize in certain experiments in the social sciences, because subjects participating in an experiment may know that they do so and know that an experimenter is observing their behavior (i.e., is part of the environment), and this knowledge may influence their behavior in an uncontrollable way.
- 16 For yet another reason (the possible occurrence of emergent phenomena) that threatens the applicability of the traditional control paradigm for experiments to sociotechnical systems, see Kroes (2016).
- 17 Another indication that we are dealing here with a special kind of experiment is that experiments on sociotechnical systems do not fit into Morgan’s table of experimental forms; see Morgan (2013, 342).
- 18 This is the reason why Hansson introduces the second criterion in his definition of action-guiding experiments (see above).
- 19 But I agree with Hansson’s (2016, 618) rather strong claim: “For action-guiding purposes, an experiment is always epistemically preferable [...] to a nonexperimental observation.”
- 20 For “learning by doing,” popular in educational quarters, see for instance (Schank, Berman, and Macpherson 1999); the *locus classicus* for tacit knowledge is (Polanyi 1958); Lévi-Strauss discusses the notion of bricolage in chapter 1: “The science of the concrete” in his *The savage mind* (1966).

References

- Ansell, Chris 2012. “What Is a “Democratic Experiment”?” *Contemporary Pragmatism* 9 (2): 159–80.
- Arendt, Hannah 1958. *The Human Condition*. Chicago, IL: The University of Chicago Press.

- Dewey, John 1938. *Logic: The Theory of Inquiry*. New York: Henry Hotl and Company.
- Franklin, Allan, and Slobodan Perovic. 2015. "Experiments in Physics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2015 edition. <http://plato.stanford.edu/archives/sum2015/entries/physics-experiment/>
- Franssen, Maarten, and Peter Kroes. 2009. "Sociotechnical Systems." In *A Companion to the Philosophy of Technology*, edited by Jan Kyrre Berg Olsen, Stig Andur Pedersen, and Vincent F. Hendricks, 223–26. Chichester, UK; Malden, MA: Wiley-Blackwell.
- Hansson, Sven Ove 2015. "Experiments before Science. What Science Learned from Technological Experiments." In *The Role of Technology in Science: Philosophical Perspectives*, edited by Sven Ove Hansson, 81–110. Springer.
- Hansson, Sven Ove 2016. "Experiments: Why and How?" *Science and Engineering Ethics* 22 (3): 613–32.
- Kroes, Peter 2003. "Physics, Experiments and the Concept of Nature." In *The Philosophy of Scientific Experimentation*, edited by Hans Radder, 68–86. Pittsburgh, PA: University of Pittsburgh Press.
- Kroes, Peter 2009. "Foundational Issues of Engineering Design." In *Philosophy of Technology and Engineering Sciences*, edited by Anthonie Meijers, 513–41. Oxford: Elsevier.
- Kroes, Peter 2016. "Experiments on Sociotechnical Systems: The Problem of Control." *Science and Engineering Ethics* 22: 633–45.
- Kuhn, Thomas S. 1977. "A Function for Thought Experiments." In *The Essential Tension*. Chicago, IL: University of Chicago Press.
- Levi-Strauss, Claude 1966. *The Savage Mind*. Chicago, IL: University of Chicago Press.
- Mach, Ernst 1976 (1897). "On Thought Experiments." In *Knowledge and Error*, T. J. McCormack and P. Foulkes, trans. 134–47. Dordrecht: Reidel.
- Morgan, Mary S. 2013. "Nature's Experiments and Natural Experiments in the Social Sciences." *Philosophy of the Social Sciences* 43 (3): 341–57.
- Niiniluoto, Ilkka 1993. "The Aim and Structure of Applied Research." *Erkenntnis* 38 (1): 1–21.
- Nordmann, Alfred 2016. "Changing Perspectives: The Technological Turn in the Philosophies of Science and Technology." In *Philosophy of Technology after the Empirical Turn*, edited by Maarten Franssen, Pieter E. Vermaas, Peter Kroes, and Anthonie Meijers, 107–125. Springer.
- Polanyi, Michael 1958. *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago, IL: University of Chicago Press.
- Schank, R. C., T. R. Berman, and K. A. Macpherson. 1999. "Learning by Doing." In *Instructional-design Theories and Models: A New Paradigm of Instructional Theory*, vol. 2, edited by Charles M. Reigeluth, 161–81. New York: Routledge.
- Schiaffonati, Viola, and Mario Verdicchio. 2014. "Computing and Experiments." *Philosophy & Technology* 27 (3): 359–76.
- Schwarz, Astrid, and Wolfgang Krohn. 2011. "Experimenting with the Concept of Experiment." In *Science Transformed? Debating Claims of an Epochal Break*, edited by Alfred Nordmann, Hans Radder, and Gregor Schiemann, 119–34. Pittsburgh, PA: University of Pittsburgh Press.

- Thye, Shane R. 2007. "Logical and Philosophical Foundations of Experimental Research in the Social Sciences." In *Laboratory Experiments in the Social Sciences*, edited by Murray Webster Jr. and Jane Sell, 57–86. Amsterdam: Elsevier.
- Vermaas, Pieter E., Peter Kroes, Ibo van de Poel, Maarten Franssen, and Wybo Houkes. 2011. *A Philosophy of Technology: from Technical Artefacts to Sociotechnical Systems*. Morgan & Claypool.
- Vincenti, Walter G. 1990. *What Engineers Know and how they Know it*. Baltimore, MD: John Hopkins University Press.
- Walker, Henry A., and David Willer. 2007. "Experiments and the Science of Sociology." In *Laboratory Experiments in the Social Sciences*, edited by Murray Webster Jr. and Jane Sell, 25–55. Amsterdam: Elsevier.
- Webster Jr., Murray, and Jane Sell. 2007a. "Why Do Experiments?" In *Laboratory Experiments in the Social Sciences*, edited by Murray Webster Jr. and Jane Sell, 5–23. Amsterdam: Elsevier.
- Webster Jr., Murray, and Jane Sell, ed. 2007b. *Laboratory Experiments in the Social Sciences*. Amsterdam: Elsevier.

2 The diversity of experimentation in the experimenting society

Christopher Ansell and Martin Bartenberger

While the concept of experimentation has never been strictly limited to the laboratory, it is increasingly called upon to serve a role as a strategy of social innovation and change. A much wider agenda for experimentation is now being imagined, one where experimentation operates in the “real world” (Gross and Krohn 2005) to address “society’s grand challenges” (Hoffmann 2011; Ferraro, Etzion and Gehman 2015). Often drawing inspiration from ideas developed earlier by John Dewey (Ansell 2012) or Donald Campbell (1969), this emerging agenda takes many forms, including policy experiments, experimentalist governance, adaptive management, design experiments, urban laboratories, pilot projects, prototyping, and niche experiments to encourage the development of sustainable technologies. Along with this variety comes different conceptions and understandings of what it means to experiment.

Our basic view is that this diversity of experimentalism is a strength because it enlists experimentation for different but often complementary purposes. As the agenda for experimentalism widens, however, we risk talking past each other. “Experiment” is clearly an elastic term (Karvonen and van Heur 2014; Steffen 2014). It is often used casually and with great rhetorical effect, but it also carries much intellectual baggage and has specific connotations for different groups. In this chapter, our focus is primarily upon situations where experimentation is being called upon to serve as a strategy for improving policy, technology, institutions, or governance. Terms like “the experimenting society” (Campbell 1998), “democratic experimentalism” (Dorf and Sabel 1998), “real-world experiments,” (Krohn and Weyer 1994; Gross and Hoffmann-Reim 2005; Gross 2010), “public experiments” (Collins 1988), “social experimentation” (Gross and Krohn 2005; van de Poel 2015), “bounded socio-technical experiments” (Brown et al. 2003), and “governance experiments” (Bos and Brown 2012; Bulkeley et al. 2012; Doorn 2015) all point to a broad agenda for conducting experiments in, for, and with the public or society.

Perhaps Donald Campbell’s idea of “the experimenting society” comes closest to expressing the scope of our interest in this chapter. The experimenting society is:

...a society in which policy-relevant knowledge is created, critically assessed, and communicated in real-life or natural (not laboratory)

settings, with the aim of discovering, through policy experimentation, new forms of public action which signify a gain in the problem-solving capacities of society.

(Dunn 1998, ix–x)

While we would not necessarily exclude the laboratory and would want to interpret “policy experimentation” broadly to include institutions, technologies, and governance, the phrase “new forms of public action which signify a gain in the problem-solving capacities of society” captures rather well what we mean by an expanding experimentalist agenda. However, we would argue that this agenda also requires a more expansive, though still bounded, conception of experimentation. Our contention is that the experimenting society is best served by having a conception of experimentation broad enough to appreciate different logics of experimentation but specific enough to distinguish the practice of experimentation from innovation or social change.

It is possible to take a restrictive position on what counts as an experiment. For example, in an essay on the history of social experiments, Brown observes that “[d]emonstrations and pilot projects are neither experimental nor quasi-experimental since they do not employ randomization, control groups, replication of treatment, or systematic measurement” (Brown 1997, 8). As philosophical Pragmatists (see Ansell (2012) and Ansell and Bartenberger (2016) for details), we regard this narrow association of the concept of experiment with control as limiting the potential for the productive use of experimentation in human affairs. It is also possible to take a much more expansive view of experiment, using the term to refer to societal actions that are novel, innovative, or previously untested. This is broadly implied in ideas such as “society as experiment” or “the city as laboratory” (Gross and Krohn 2005). As Pragmatists, we approve of this wider ambit but also worry that associating the concept of experimentation too broadly with novelty robs it of any distinctive analytical value. In the broadest use of the term, nearly any social change, policy innovation, or new technology might qualify as an experiment.

Building on our prior work (Ansell 2012; Ansell and Bartenberger 2016), we take a meta-position on the relationship between a restrictive and an expansive view of experiment, describing three distinct logics of experimentation—controlled, evolutionary, and generative. Controlled experimentation focuses on isolating causal factors and as a result equates experimentation with techniques of control (e.g., randomization, control groups, etc.). It is the understanding of experimentation aligned with the restrictive view of experiment. By contrast, evolutionary and generative logics of experimentation are less concerned with control and more concerned with, respectively, maximizing rates of innovation and creating new ideas. These views are more compatible with the expansive view of experimentation.

Thus, our goal in this chapter is to foster an appreciation for the diversity of experimentation while making it clear how different logics are distinctively experimental. We begin by examining what makes each of these three logics, despite their differences, experimental.

The common core of experimentation

While our goal is to describe the diversity of experimentation, we begin by pointing to what we think is common to most ideas of experiment. Donald Schön offers a useful and succinct starting point: “In the most generic sense, to experiment is to act in order to see what action leads to. The most fundamental experimental question is, ‘What if?’” (1983, 145). This definition actually has several important components. First, it suggests that “to experiment is to act,” implying that experimentation is a form of learning-by-doing. Second, it suggests that “an experiment is to act in order to see,” implying an intention to learn from the action. And third, it suggests that the key experimental question is “What if?”, which implies either uncertainty or a critical reflexivity about current beliefs.

Building on Schön’s definition, we argue that the *intention to learn* is at the core of what it means to experiment. Learning is itself an elastic concept and there is a vast literature on the topic. Here we utilize a definition offered by Jack Mezirow, who developed the theory of transformative learning: “Learning may be defined as the process of making a new or revised interpretation of the meaning of an experience, which guides subsequent understanding, appreciation, and action” (1990, 1). We further elaborate that “making a new or revised interpretation of the meaning of an experience” implies drawing an inference from action. Thus, learning is drawing an inference from action that guides subsequent understanding, appreciation, and action.

As Mezirow notes, however, not all learning is intentional; it can also occur non-intentionally via processes like socialization or operant conditioning. Even when learning is intentional, it may only occur post hoc after an action is taken. Experimentation implies a level of self-conscious reflexivity about the goal of learning from an action that is prior to the action itself. It is prospective rather than retrospective. Even intentionality, however, is insufficient to define experimentation. For example, students doing a problem set to learn mathematical concepts may be learning intentionally but not experimentally. These students may be simply following a procedure described in a textbook. Therefore, experimentation also implies some degree of uncertainty about, or critical scrutiny of, the action in question. This uncertainty is conveyed by the question “what if?”. Drawing these ideas together, we argue that the common feature of experimentation is *taking an action with the intention of learning “what if?”*.

While this definition is broad enough to encompass a variety of specific meanings of experiment, it is delimited enough to distinguish experiments from other types of action. Most policy interventions, for example, produce some degree of experiential learning. However, this does not make them experimental because this learning may be inadvertent, unintentional, or post hoc. Sometimes we call a program or policy experimental merely because it is being tried out for the first time or is novel in some way. But a novel

intervention may simply be the best or the only thing we can think of to do in a difficult situation. Again, there is no intention to learn. Finally, most contemporary policy interventions incorporate an element of evaluation. However, this does not make them experimental because the goal of learning is typically a secondary one. Policy or program interventions may thus vary in terms of how much they are designed for learning. Interventions become more experimental to the extent that the intention to learn is central to their purpose.

The “intention to learn” therefore also depends on the objectives of the designers, participants, or observers of the intervention. Designers of an intervention may have the intention to learn from it, while participants may be simply interested in producing some specific outcome. In some cases, designers and participants may not be strongly focused on learning, but observers of the intervention might be. Thus, the same intervention may be experimental for some but not for others. The point is that the “intention to learn” is a relatively simple criterion for identifying what is distinctive about experimentation and how it might differ from learning or social change understood more broadly. At the same time, it helps us appreciate that the boundary around what is or is not an experiment is not necessarily clear cut. The degree to which an intervention is experimental will vary with the intentions of those involved.

Building on recent scholarship on laboratories, Karvonen and van Heur extend the conception of experimentation so that it can apply more broadly to cities but are also mindful of the need to distinguish it from social change more generally. Thus, they argue that “...it is helpful to understand experimentation as (1) involving a specific set-up of instruments and people that (2) aims for the controlled inducement of changes and (3) the measurement of these changes” (2014, 383). These are more restrictive criteria than the “intention to learn,” though we can appreciate that the intention to learn may entail making specific preparations (e.g., specific set-up, controlled inducement, and measurement). To canvass the wider diversity of experimentation, however, it is best to leave these specific preparations more open-ended.

The criterion of “intention to learn” is, in fact, arguably too restrictive. One potential challenge to this criterion comes from Krohn and Weingart’s (1987) description of the Chernobyl nuclear accident as an “implicit experiment.” They argue that risky technologies such as nuclear power are sold to the public as tested and safe but this disguises the deep uncertainty that they entail. Given the uncertainty that surrounds this technology, each new nuclear power plant is like an experimental trial. For the public, however, Chernobyl revealed the experimental character of the technology *retrospectively*. The public had no intention to learn from the technology prospectively because it had not appreciated the risk and uncertainty surrounding the technology. For Krohn and Weingart, Chernobyl was an experiment because it was an action taken under conditions of high uncertainty (ignorance) that taught us something. It was an *implicit* experiment, however, because the

learning was retrospective. Krohn and Weingart's argument can be interpreted as saying that technologies like nuclear power are being introduced *as if* we were comfortable learning about them as we go.¹ By our definition, "natural experiments" are also "as if" experiments—not because the experimenter did not design the experimental controls, but because learning is retrospective.

Some scholars describe action as experimental when actors are engaged in "bricolage" or "intelligent tinkering." In such cases, the actors involved may not have a well-developed "intention to learn." For example, Dorstewitz describes the social process of converting a closed colliery in Zollverein German into an historical site as an "iterative experimental exploration of the place and its potentials" (2013, 437). He notes, however, that the experimental quality of the project only became clear in retrospect. The participants did not begin the process by saying to themselves "we are initiating a major experiment in the use of old industrial space." Rather, they confronted an indeterminate situation that called for inquiry and action. Each step in their process of response created new conditions and questions that responded to the outcome of the prior step. New purposes emerged over time. The result of this exploration was "...a tendency to move from improvised and provisional experiments to larger, more systematic experimental structures" (2013, 437).

In this example, there is no question that the participants engaged in inquiry about what actions to take or that they learned from these actions and incorporated the lessons into subsequent activity. However, it is much less clear—at least in the account provided—that actions were taken with an intention to learn or with learning as one of the chief objectives. They did fashion novel strategies through situated inquiry under conditions of uncertainty, and they did learn from their interventions. For us, this example sits at the boundary between experimental and nonexperimental. To judge whether it is experimental, we would have to know more about the participants' intentions as they made decisions about how to proceed.

The "intention to learn" does not, however, only imply deductive theory-testing experimentation and can therefore encompass a wider range of experimental practice. Cabin, for example, describes a pattern of "intelligent tinkering" in ecological restoration that builds on an adaptive, trial-and-error approach. One restoration ecologist who Cabin interviews nicely sums up the approach: "We really just learn by doing. We try plausible things, watch what happens, and adjust our practices accordingly" (Cabin 2011, 186). This approach is very similar to the process Dorstewitz describes in the Zollverein colliery case. The difference is that the intention to learn is clear, and based on this criterion "intelligent tinkering" is experimental.

In the early 1980s, Hacking (1982, 1983) led a charge among philosophers and historians of science to appreciate more fully the richness of experimental practice. A key point of his work was to argue that experiments are not necessarily limited to testing theory. Building on Hacking's lead, Steinle

(1997) distinguished between “exploratory” and “theory-oriented” experimentation. The former is more inductive in nature and oriented toward building rather than testing theory.² As a result, it is flexible and open-ended in its use of instrumentation and oriented to the context of discovery rather than to the context of justification.

In summary, we argue that the common feature of all experimentation is *taking an action with the intention of learning* “*what if?*” Such a definition is meant to be broad so that it can encompass a variety of types of experimentation. Yet this definition also limits the scope of the concept. The most important limit is the *intention to learn*, which implies a prospective and self-conscious reflexivity about the goal of drawing an inference from an action that will guide subsequent understanding, appreciation, and action. Action may be more or less experimental depending on how central the intention to learn is to the goals of those who are conducting or observing it. Retrospective learning from action may qualify as “as if” experimentation but does not fully meet the criteria established by our definition. Building on this basic understanding of experimentation, we now turn to describing three different logics of experimentation.

Controlled experimentation

The first logic of experimentation comes closest to the conventional view of how experiments work in the laboratory and in science more generally. Controlled experimentation is characterized by its search for valid inferences about cause and effect. The main tools for this search are a tightly controlled environment and the isolation of relevant factors. Factors are conceptualized as variables prior to the experiment, which is carefully designed to isolate the effects of certain factors and to avoid confounding the effects of other factors. As Schön notes, this kind of experimentation “...succeeds when it effects an intended discrimination among competing hypotheses” (1983, 146). Controlled experiments clearly have an “intention to learn.”

Controlled experimentation has deep historical roots that can be traced back at least to the mid-eighteenth century. The use of “control” and “treatment” groups to achieve control was first developed, followed later by the strategy of randomization to reduce bias and to control for unknown factors. These ideas came together in the mid-twentieth century to create the randomized controlled trial, which randomly assigns subjects to control and treatment groups (Dehue 2001; Manzi 2012). Controlled experimentation demands strong separation between the attitudes of the researcher and the conduct of the experiment because the goals and expectations of the researcher can themselves become confounding factors. Careful experimental design is therefore required to minimize experimenter bias (Kroes 2015).

The results from controlled experiments are typically understood to be less biased than observational data, i.e., where the available data has solely to be observed and analyzed by the researcher (Cox and Reid 2000; Gerber

and Green 2008; Greenstone and Gayer 2009; Manzi 2012). The chief difference between experimental and observational data lies in the greater control over variables achieved by experimental settings. Randomized controlled trials (RCTs) are therefore regarded as the “gold standard” of controlled experimentation (Cartwright 2007, 2011; Ettelt, Mays and Allen 2015) because their design minimizes selection and allocation bias. As Cartwright notes (2007), RCT’s often have a deductive character because if the treatment produces a particular outcome, a hypothesis can be *deduced* to be true.

One of the best known large-scale controlled policy experiments was the Moving to Opportunity (MTO) experiment conducted in five cities in the United States between 1994 and 1998. MTO was designed to determine whether a housing voucher program would have positive social and economic effects. Nearly 5000 low-income families living in public housing projects in high-poverty neighborhoods were recruited for the experiment and then randomly assigned to one of three treatment groups: a group receiving a housing voucher that could only be used to move to housing in low poverty areas combined with counseling and housing assistance services; a second group that received a housing voucher that could be used anywhere; and a third “control” group that did not receive a housing voucher but remained eligible for traditional public housing. The program then conducted “interim” (4–7 years after enrollment) and “long-term” (10–15 years after enrollment) evaluations of treatment effects. Among the key long-term findings were that the vouchers did enable participants in the two voucher programs to live in lower poverty and safer neighborhoods and did produce positive physical (adults only) and mental health benefits (adults and female youth only). However, the vouchers did not affect the economic self-sufficiency of participants (Moving to Opportunity Fair Housing Demonstration Program website, <http://www.nber.org/mtopublic/> Accessed May 8, 2016).

One common issue that arises in the design of controlled experiments is “external validity.” Laboratory experiments are able to control better the factors associated with interventions, but the conditions required for control may be unrepresentative of real world conditions. “Field experiments” (like the MTO experiment) are seen as sacrificing some degree of control for more representative conditions and hence greater “external validity” (Campbell 1969).³ In fact, there is a range of variation between the “artificial” conditions of the lab and the “real world” conditions of the field. For economic experiments, Harrison and List (2004) distinguish three dimensions along which experiments may vary:

- 1 Standard versus non-standard subject pools (the standard subject pool is university undergraduates);
- 2 Abstract versus realistic framing (abstract framing describes a set of goods, incentives, etc. in abstract terms; realistic framing uses actual goods or incentives);
- 3 Imposed rules versus rules set naturally by the context.

In this framework, a “conventional lab experiment” is one in which there is a standard subject pool, an abstract framing of the experiment, and imposed rules. A “natural field experiment” is one in which the subjects “naturally undertake these tasks and where the subjects do not know they are in an experiment” (Harrison and List 2004, 1014).

Evolutionary experimentation

The second logic of experimentation builds on basic evolutionary principles of variation, selection, and retention. We label this logic *evolutionary experimentation* because these principles are often inspired by the processes of variation, differential reproduction, and heredity.⁴ The “blind” selection processes associated with natural evolution, however, are incompatible with the intentionality inherent in our understanding of experimentation (Campbell 1974). Nevertheless, the concept of natural evolution is often extended by analogy to social evolution, as suggested in the following statement:

Adaptation in ecosystems typically occurs by the action of selection on diversity...The same principle is often true in human societies; adoption of a greater diversity of policies or management approaches may make it more likely that a successful solution (whether adaptation or mitigation) will emerge.

(Cumming et al. 2013, 1144)

As this statement suggests, variation, either over time or at the population level, is central to the idea of evolutionary experimentation. However, as the evolutionary economics literature on strategic niches clarifies, directed as opposed to blind variation is the key mechanism and learning is the source of this directed variation (Schot and Geels 2007).

Isolating causation is not the chief aim of evolutionary experimentation. Instead, its chief value arises from producing new traits through the conduct of many trials. By producing variation through many different trials and by examining which of them are successful (according to certain specified criteria), it is more likely to generate novel ideas and innovations, and also quickly sift out trials that do not work. Although evolutionary experimentation is certainly less widely known or appreciated than controlled experimentation, these ideas have been taken up in evolutionary economics and in related work on technological innovation and change (Dosi and Nelson 1994). The literatures on transition management and strategic niche management explicitly adopt these ideas about evolutionary experimentation.⁵

One of the most ambitious examples of evolutionary experimentation has been the Dutch Energy Transition Project (ETP). In order to promote a societal transition from conventional energy use to a more sustainable energy model, the Dutch Ministry of Economic Affairs adopted an evolutionary strategy of promoting a range of “transition experiments” across multiple

energy sectors. These experiments were intended to encourage system innovation and aimed to encourage learning about different technological “transition pathways.” Forty-eight transition experiments were initiated in the first round of development (2004–2007; Kemp 2010). Kemp describes the program as a form of “guided evolution” that encourages “[a] portfolio of options...generated in a bottom-up, forward-looking manner in which special attention is given to system innovation” (Kemp 2010, 291, 309).

Applied to social problem-solving, evolutionary experimentation suggests that the rate of individual experimentation needs to be increased and the number of units experimenting expanded. In general, this logic puts more faith in successful innovation arising probabilistically from a large number of trials than from a more teleological conception of rational design. By contrast with controlled experimentation, evolutionary experimentation is more *inductive* than deductive, relying on trial-and-error learning. This evolutionary experimentation embraces the power of large numbers (Menand 2001) but, whereas controlled experiments make use of the law of large numbers via randomized sampling to eliminate bias and confounding influences, evolutionary experimentation uses large numbers to increase the frequency of innovation. In other words, controlled experimentation requires sample sizes large enough to guarantee sufficient degrees of freedom to estimate statistical differences between treatment and control groups. By contrast, the evolutionary logic requires a high degree of freedom in order to generate sufficient variation.

As noted, the literature on Strategic Niche Management explicitly adopts a variation-selection-retention framework for promoting sustainability experiments (Raven et al. 2008). The goal is to “...improve the functioning of the variation selection process by increasing the variety of technology options upon which the selection process operates” (Hoogma et al. 2002; loc 4821). Strategic niches can *shield*, *nurture*, and *empower* variations, protecting them from selection pressures, as Verhees et al. (2013) have illustrated for the Dutch photovoltaics sector. A distinctive feature of an evolutionary view of experimentation is that it shifts the focus from individual experiments to systems or ecologies of experimentation.

The key to the evolutionary logic is to generate variation (Nair and Roy 2009). One strategy for doing this is to promote “parallel experimentation,” which Ellerman (2004) argues can generate a portfolio of “best practices.” Another strategy is to rapidly run many experiments in sequence, a strategy called “rapid experimentation” (Thomke 2003). Simulation modeling is sometimes used in this way. In a discussion of resource management, for example, Jager and Mosler argue that agent-based modeling creates the possibility for “...conduct[ing] thousands of experiments in a very short time, allowing for exploring the effects of many different combinations of factors” (2007, 99). Given its faith in the power of large numbers, evolutionary experimentation is unique in its toleration and even embrace of failure. As Thomke argues, it is important to “fail early and often” (2003, 13).

The literatures on “democratic experimentalism” (Dorf and Sabel 1998; Sabel and Zeitlin 2010, 2012) and “laboratory federalism” (Polsby 1984; Oates 1999; Kerber and Eckhard 2007; Saam and Kerber 2013) both contain elements of an evolutionary logic of experimentation. The democratic experimentalism literature, for example, begins with the idea that many different local units (which could be firms, local governments, state governments, etc.) experiment in parallel to achieve broad framework goals. These parallel experiments are then monitored and information is pooled and peer-reviewed. Best practices are identified and then fed back to the units to inform subsequent experiments.

Generative experimentation

In addition to controlled and evolutionary experimentation, a third distinct logic of experimentation is *generative experimentation*. It can be thought of as a process of generating and iteratively refining a solution concept (an idea, innovation, design, policy, program, etc.) based on continuous feedback and with the goal of addressing a particular problem.⁶ This type of experimentation embodies what Donald Schön called a “move-testing experiment,” which takes action to achieve a purpose and evaluates whether that purpose has been achieved (Schön 1983). It is often associated with design and, in this context, generative experiments are often referred to as design experiments or prototyping. Although such experiments typically relax control over experimental conditions, they still have an intention to learn from their interventions.

Like controlled experiments, generative experiments concentrate on a single experiment at a time, but discerning causal mechanisms is not their primary goal. Their aim is to stimulate production and analysis of information about an intervention in order “...to help re-specify and re-calibrate it until it works” (Stoker and John 2009, 358). In other words, generative experiments focus on the process of elaborating new and innovative solutions to existing problems. Generative experiments often adopt a “probe and learn” strategy in which a prototype is introduced and then successively refined based on feedback (Lynn, Morone, and Paulson 1996). Many strategies of adaptive problem-solving—from intelligent tinkering (Cabin 2011) to problem-driven iterative adaptation (Andrews, Pritchett, and Woolcock 2013)—reflect the logic of generative experimentation, as does exploratory experimentation (Steinle 1997; Burian 2013; Schiaffonati 2015).⁷ The inferential logic is more *abductive* (creatively developing new hypotheses from incomplete information) than deductive or inductive (Ansell and Bartenberger 2016).

Since control is not the *raison d’être* of generative experiments, they are more likely to be conducted as “real-world” (Gross and Hoffmann-Riem 2005; Gross and Krohn 2005; Gross 2010) or “wild” experiments (Lorimer and Driessen 2014) than as laboratory experiments. Krohn (2007) has drawn

an analogy between laboratory versus real-world experimentation and nomothetic (seeking general laws) versus idiographic knowledge (seeking historical specificity). He argues that for real-world experimentation, “[l]earning focuses on the local conditions to keep an installation going, not on finding the parameters of a general solution” (2007, 143). Generative experiments are often more idiographic than nomothetic.

A major urban planning intervention in Vienna, Austria, offers a good example of generative experimentation (Bartenberger and Sześciło 2015). In 2010, Vienna’s city government announced plans to redesign its largest shopping street, *Mariahilferstrasse*, which is also a major urban thoroughfare. City planners presented citizens with three design options and opened up a dialogue process with citizens and stakeholders. Based on the results of this dialogue, the city then proposed a final design. However, unlike many urban planning projects, this was not the end of the process. The city then conducted a “test” of the design between August 2013 and February 2014, which led to several adaptations to the design. A neighborhood referendum was then held on the design and a majority of residents voted in favor of implementing it.

Pilot projects often exhibit the logic of generative experimentation. In an analysis of water governance pilot projects in the Netherlands, for example, Vreugendhil et al. (2010) distinguish pilot projects from laboratory (controlled) experiments. In contrast with laboratory experiments, the interaction of pilot projects with contextual factors is only controlled to a limited extent. They identify one prominent type of pilot—an *explorative pilot*—that is used “to test and refine innovations in their context and gain experience” (2010, 11). This type of pilot comes closest to the generative logic.

Though both evolutionary and generative experimentation typically occur in “real world” contexts, the logic of evolutionary experimentation always refers to a population or ecology of experiments while generative experimentation refers to the iterative refinement of a single experiment. Evolutionary experimentation hopes to discover a successful trial from a population of many less successful, or even failed, trials. Hence, it regards failure as a necessary condition for success and thus places great value on increasing the number of trials. Generative experimentation does not think in population or ecosystem terms. Instead, it focuses on designing and redesigning a single solution concept until it is successful. Both logics may require multiple interventions. While the evolutionary logic is about increasing the *chance* of success through many independent trials, the generative logic is about *designing* a successful solution through accumulated knowledge and experience. “Rapid prototyping” combines both logics.

In the previous section, we described the Strategic Niche Management literature as a logic of evolutionary experimentation. However, this literature and the related literature on transition management typically describe individual experiments in generative terms. Transition or niche experiments

typically take place in a real-life societal context as opposed to a laboratory (Van den Bosch 2010, 61; see also Kroes 2015). They are characterized by a mode of action described as “learning-by-doing” or “probe and learn” (Hoogma et al. 2002, loc 4744) and they have an “emergent project design” realized through an “iterative, cyclical process” (de Wildt-Liesveld, Joske and Barbara 2015, 156).⁸ A closely-related literature on “bounded socio-technical experiments” characterizes them as “...learning by doing, doing by learning, trying out new strategies and new technological solutions, and continuous course correction...” (Brown et al. 2003, 292; Brown and Vergragt 2008; see also Ceschin 2014, 3).

Challenges to the experimenting society

The idea of the “experimenting society” expresses a hope that broadening the use of experimentation will improve society’s capacity for problem-solving. But as Donald Campbell was well aware, this wider agenda for experimentation encounters many challenges. A number of these challenges have been well aired before (Dunn 1998). In this final section, our goal is to begin an exploration of how the different logics of experimentation we have described might relate to such challenges. To some degree, all three types face similar ethical, political, and administrative challenges, but our focus will be on the particular strengths and weaknesses that arise from different logics of experimentation. Controlled experimentation has received by far the greatest scholarly attention, so it naturally becomes our main case. However, we try to balance this disproportionate emphasis on controlled experimentation by explicitly considering the parallel challenges confronted by evolutionary and generative experimentation.

Perhaps the most obvious challenges to the experimenting society are ethical in nature (Weber 2011; van de Poel 2015; Doorn, Spruit and Robaey 2016). The literature on this topic is extensive and long-standing, and we make no attempt to do it justice here. We merely try to call attention to some of the ways that these ethical issues may arise from the different logics of experimentation. Controlled experimentation, for instance, raises several kinds of ethical issues that arise specifically from the imperatives of control, particularly randomization and assignment to control or treatment groups. Because random assignment is blind to differences in status, identity, or resources, random assignment may be understood to be an ethical strength. However, the ethics become more problematic when treatments are unequal or perceived as unequal (Lilford and Jackson 1995). In the context of controlled experimentation in American education, for example, Cook notes the resistance of schools to random assignment because a randomized experiment implies that schools must, among other things, “...surrender choice over the treatment they will receive...” (2003, 132). However, Oakley et al. (2003) find that randomization is generally an acceptable practice for educational experiments in the United Kingdom.

In a review of criminology experiments, Weisburd asks: “Is it ethical to assign criminal justice sanctions or treatments on the basis of research rather than legal criteria?” (2000, 184).⁹ In practice, he argues that this is less of a problem (e.g., it does not produce resistance) if the control group receives the status quo treatment and the treatment group receives a modified, but preferential, treatment. For example, he ascribes the lack of objections to the California Reduced Prison Experiment to the fact that the experiment was designed to release some prisoners earlier than they would otherwise be released. His point is that ethical issues associated with controlled experimentation can, in part, be addressed through careful experimental design.¹⁰

Evolutionary and generative experimentation also face ethical challenges that may arise from their respective logics. As described earlier, both types place less emphasis on control. Van de Poel (2011) argues that because social experiments (which may be either evolutionary or generative, depending on perspective) operate under less controlled conditions, their consequences may be more likely to have uncontrolled and potentially irreversible harms. As described by Irvine and Kaplan (2001), conducting “small experiments” may be one valuable strategy for dealing with such harms.

Van de Poel (2011) also notes that because social experiments take place in society, they may be less likely to be recognized as experiments by those potentially affected by them. He proposes criteria for responsible social experimentation, some of which are concerned with responsible management of potentially negative consequences and others which are concerned with achieving consent from involved or affected parties. Devon and van de Poel (2004) also explicitly consider the ethical challenges of iterative design-redesign cycles (a key feature of generative experimentation). Design experiments, they note, may suffer from a diffusion of responsibility because different designers are often involved at different stages of the design process. A social ethics of design, they argue, requires heightened reflexivity and learning about experimental responsibility and consequences.

Beyond ethical issues, the experimenting society also raises practical challenges stemming from the political use of experimentation. A number of authors have noted these problems for controlled experimentation, which reflects the great authority controlled experiments enjoy in policy debates due to their claim to provide definitive scientific knowledge. Brodtkin and Kaufman observe that controlled experimentation in social policy seems ideally suited to speak truth to power, but in practice “...it seems to have distorted debates over welfare reform, providing tidbits of data that were used more often as weapons to advance existing positions than as evidence to shape them.” (2000, 521–22).¹¹ Controlled social policy experiments have not been neutral arbiters of policy design, they argue, but are often used instead as “shadow institutions” to surreptitiously protect and advance particular designs. Ettelt et al. (2015) arrive at similar conclusions about RCT experiments in the United Kingdom, arguing that they were used to demonstrate rather than evaluate the effectiveness of a particular policy.

While controlled experiments raise concerns about the politicization that can stem from their claim to neutral authority, generative and evolutionary experimentation may raise other kinds of political challenges. Both generative and evolutionary experimentation, for example, potentially create value by which specific interests may wish to profit. In our example of evolutionary experimentation—the Dutch energy transition project—Kern and Smith (2008) found that there was a tendency for experiments to be captured by the existing energy producers.

Another political challenge is a tension between the goals of learning and demonstrating success. Real world experiments face pressures to demonstrate success. In the case of controlled experimentation, Ettelt argues that this attitude of searching for success violates the assumption of *ex ante* uncertainty that should characterize RCT experimentation. Although searching for what works does not violate the logic of generative experimentation, there is still a difference between searching for what works (which makes the “intention to learn” central) and using demonstration “for the purpose of winning the policy argument” (Ettelt et al. 2015, 301). Pilot and demonstration projects often face tensions between the goals of evaluating an idea, exploring its limits, and demonstrating its success (Sanderson 2002; Markusson, Ishii and Stevens 2011; van der Heijden 2014, 2015; Nair and Howlett 2015b).

Some political challenges to a broader experimentalist agenda may affect all three types of experimentation. For example, Cook (2003) and Brodtkin and Kaufman (2000) found that education and social policy experiments, respectively, create timing problems because the results from these controlled experiments are delayed until after the political agenda has moved on. Vreugdenhil et al. (2010) note that pilot projects—which are typically generative experiments—also face similar timing problems.

The efficacy and efficiency of experimentation as a mode of social problem-solving is another practical challenge for the experimenting society. In part, this issue may require tailoring the type of experimentation to specific purposes. Cabin (2011), for example, describes giving up on controlled experimentation as a strategy of ecological restoration because it could not really cope with the diversity, context-specificity, and scale inherent in restoration projects. This led him to formulate his idea of “intelligent tinkering” described above as a form of generative experimentation. Similarly, Kroes (2015) challenges the efficacy of controlled experimentation for socio-technical systems because socio-technical experiments are subject to both the autonomy of societal actors and the possibility of “emergent” phenomenon.

While controlled experimentation raises specific challenges, so too do evolutionary and generative experimentation. Pilot projects, demonstration projects, and policy-design experiments often fail or are disappointing in their results (Van der Heijden 2014). This may be more of a problem for generative rather than evolutionary experimentation because the latter expects

higher rates of failure. Nevertheless, it is difficult to diffuse successful pilots, presenting a challenge to the evolutionary logic as well (Vreugdenhil et al. 2010; Nair and Howlett 2015a). This problem may arise because the results of context-specific experimentation are difficult to generalize. The Strategic Niche Management literature has been particularly attuned to diffusion challenges. It observes that even successful experiments are unlikely to diffuse because of challenges encountered when they must go to scale. As a result, it is common for sustainable technology experiments to become “one-off” innovations (Raven et al. 2016).

One important issue that arises for an experimentalist agenda is who has influence over the design and conduct of experiments. Böschen (2013) suggests that the more that experimentation engages the “real-world,” the more important it becomes to mobilize agreement from broader publics, and Karvonen and van Heur (2013) argue that public experiments must always justify themselves to the public. Van de Poel (2011, 2015) makes a similar argument for social experiments, and Doorn (2015) provides some useful insights into how to structure stakeholder input into governance experiments. The intensity of public or stakeholder concern will often tend to reflect the potential scale and severity of experimental effects. No one is particularly concerned about the ethics of “thought experiments,” but concern is heightened when experiments may have wider consequences. As a result, there have been recent calls for new modes of “collective experimentation” that change the relationship between scientists and citizens (Felt and Wynne 2007). Experiments are troubling to stakeholders when they become subjects of experiments that they perceive to not serve or represent their interests. Cook (2003), for instance, notes that local schools in the United States resist controlled experimentation because they regard such experiments as representing the interests or objectives of the federal and state governments rather than their own.

Controlled experimentation may face particular challenges in structuring public or stakeholder involvement because effective control may require sharper separation between the design and conduct of experiments. However, it is still possible to take public and stakeholder input into account in these circumstances. Oakley et al. (2003) found that educational experiments in the United Kingdom worked better using randomized assignment when strong partnerships were created between researchers and service providers. Still, the relaxed control and in situ experimentation associated with evolutionary and generative experimentation may open new possibilities for collaboration and participant input. Less strict design requirements may create more openness to the input of different stakeholders, though the collaborative process is often quite challenging even in these cases (van der Heijden 2014). It may be possible, however, that some forms of experimentation enhance collaboration (Criado, Rodríguez-Giralt and Mencaroni 2015).

To conclude this section, it is clear that there are many practical challenges to the experimenting society, though we do not regard them as fatal to the goal of widening an experimentalist agenda. Experiments in, for, and with society are neither inherently good, nor inherently bad. They surely pose threats and dangers but may also promise valuable innovations, refined problem-solving strategies, and new knowledge. Although some challenges—like timing—appear common to all three logics of experimentation, others may be more specific to the logic of experimentation. Some ethical and political challenges arise specifically from the logic of control (e.g., ethical concerns related to assignment to control and treatment); others are more specific to the less-controlled nature of evolutionary or generative experimentation (e.g., concerns about irreversibility of uncontrolled processes). The wider lesson we draw from this discussion is that all three logics require careful reflection on their design, operation, and use.

Conclusion

To broaden the agenda of social experimentation, we must grapple with the elasticity of the concept of experimentation. While different meanings of experiment may perhaps happily coexist in a polysemic fashion, this elasticity may also lead to confusion, frustration, and contradiction. To avoid talking past each other, it is useful to enlist some constraints on our use of the concept while acknowledging the real diversity of experimental logics. Toward this end, this chapter defines the common criterion of experimentation as *taking an action with the intention of learning “what if?”* This criterion is broad and can encompass a wide range of different experimental strategies, but it also sets some useful limits that distinguish experimentation from social change in general.

The chapter distinguishes three different logics of experimentation: controlled, evolutionary, and generative. Controlled experimentation is our classic model of scientific experiment. Careful controls are established to discern and isolate causal effects. By contrast, generative experimentation is less concerned with control and aims instead to generate new solutions concepts through iterative refinement. The two logics are in tension, though not necessarily diametrically opposed. Controlled experiments often sacrifice laboratory controls for the greater external validity of field experiments, and a sequence of controlled experiments may be run to refine knowledge over time. Generative experimentation sacrifices control with the aim of creatively generating new designs or problem-solving strategies, often in a real-world context, but without necessarily relinquishing all attempts at control. Indeed, prototypes are often developed in controlled laboratory settings.

Evolutionary experimentation differs from controlled or generative experimentation by taking a population or ecological perspective on experiments

and by stressing the importance of variation and the value of failure. However, at the level of individual experiments, this logic is compatible with either controlled or generative experimentation. Since the logic of evolutionary experimentation draws lessons inductively at the systems level, it is perhaps less likely to be concerned about rigorous control at the unit level. But there is no necessary reason that evolutionary and controlled experimentation cannot be combined.

We argued in the introduction that the diversity of experimentation can be a strength for the experimenting society. As developed in earlier sections, the reason is that each logic has a different goal. Controlled experimentation is primarily designed to isolate causation while evolutionary experimentation aims to increase systemic innovation and generative experimentation seeks to create and refine new solution concepts. If the experimenting society is "...a society in which policy-relevant knowledge is created...with the aim of discovering...new forms of public action which signify a gain in the problem-solving capacities of society" (Dunn 1998, ix–x), we can see that all three logics are needed.

Notes

- 1 Building on the tradition of Krohn, Weingart and Gross (2010) have recently described how experiments can be conceived of as generating learning through surprise. They write that "...an experiment expedites and assists a surprise by acknowledging and (ideally) documenting what has not been known before so that the stakes are more transparent for the actors involved." (2010, 79). This perspective emphasizes the association of experiments with uncertainty and ignorance, but also appreciates that experiments are an occasion for learning.
- 2 Steinle doesn't use the term inductive, but he does describe exploratory experimentation as having "...the goal of finding empirical rules and systems of those rules" (1997, S71). For a discussion of the inductive nature of exploratory experimentation, see Burian (2013). For a recent discussion of the use of exploratory experiments in computer science, see Schiaffonati (2015).
- 3 The contrast between laboratory versus field experiments is similar to the contrast in biology and chemistry between *in vitro* (in test tube) versus *in vivo* (on live animals) experiments. Laboratory and *in vitro* experiments achieve great control and are often very efficient. However, results are criticized for being unrepresentative and reductionist. On *in vitro* and *in vivo* experimentation, see Lipinski and Hopkins (2004).
- 4 The variation of certain traits in a given population leads to differences in reproduction (either due to nonrandom environmental influences or sexual selection). Because these traits are inheritable, the composition and character of the population changes over time, leading to the differential reproductive success of certain groups of the population that possess the beneficial traits (Mayr 2001, chap. 6).
- 5 Scholars of health systems have also begun to imagine how complex health systems might adopt an evolutionary strategy of retaining small successful changes and discarding less successful ones (Martiniuk, Abimbola and Zwarenstein 2015).
- 6 Hacking (1982) notes that scientists use experiments to create phenomena and not merely to test theories.

- 7 Note, however, that Schiaffonati (2015) argues that exploratory experimentation is a form of controlled experimentation, but the control is a posteriori rather than a priori.
- 8 Pisano (1996) distinguishes experiments oriented to “learning before doing” and “learning by doing.” He also argues that experiments can be distinguished in terms of how representative they are of their final “target” uses.
- 9 Randomized experimentation in public policy seems to go through periods of what Farrington, referring to the field of criminology, calls “feasts or famine” (Farrington 2003, 224).
- 10 Weisburd notes that where citizens perceive unequal treatment, they may produce political demands that work against experimentation. As a result, lower-profile experiments may be easier to implement. He suggests, however, that practical challenges may be more of a barrier to controlled experimentation than ethical challenges. Similar conclusions are drawn by Cook (2003) about educational experiments.
- 11 For a much more positive account of this “golden era” of controlled policy experimentation, see Oakley (1998).

References

- Andrews, Matt, Lant Pritchett, and Michael Woolcock. 2013. “Escaping Capability Traps Through Problem Driven Iterative Adaptation (PDIA).” *World Development* 51:234–44.
- Ansell, Christopher. 2012. “What Is a ‘Democratic Experiment’?” *Contemporary Pragmatism* 9 (2):158–79.
- Ansell, Christopher, and Martin Bartenberger. 2016. “Varieties of Experimentalism.” *Ecological Economics* 130:64–73.
- Bartenberger, Martin, and Dawid Sześciło. 2015. “The Benefits and Risks of Experimental Co-Production: The Case of Urban Redesign in Vienna.” *Public Administration* 94 (2):509–25.
- Bos, J. J., and Rebekah R. Brown. 2012. “Governance Experimentation and Factors of Success in Socio-Technical Transitions in the Urban Water Sector.” *Technological Forecasting and Social Change* 79 (7):1340–353.
- Böschen, Stefan. 2013. “Modes of Constructing Evidence: Sustainable Development as Social Experimentation—The Cases of Chemical Regulations and Climate Change Politics.” *Nature and Culture* 8 (1):74–96.
- Brodtkin, Evelyn Z., and Alexander Kaufman. 2000. “Policy Experiments and Poverty Politics.” *Social Service Review* 74 (4):507–32.
- Brown, Halina S., and Philip J. Vergragt. 2008. “Bounded Socio-Technical Experiments as Agents of Systemic Change: The Case of a Zero-Energy Residential Building.” *Technological Forecasting and Social Change* 75 (1):107–30.
- Brown, Halina S., Philip Vergragt, Ken Green Philip, and Luca Berchicci. 2003. “Learning for Sustainability Transition Through Bounded Socio-Technical Experiments in Personal Mobility.” *Technology Analysis & Strategic Management* 15 (3):291–315.
- Brown, Robert. 1997. “The Delayed Birth of Social Experiments.” *History of the Human Sciences* 10 (2):1–21.
- Bulkeley, Harriett, Matthew J. Hoffmann, Stacy VanDeveer, and Victoria Milledge. 2012. “Transnational Governance Experiments.” *Global Environmental Governance Reconsidered*, edited by Frank Biermann and Philipp Pattberg, 149–71. Cambridge, MA: MIT Press.

- Burian, Richard M. 2013. "Exploratory Experimentation." In *Encyclopedia of Systems Biology*, edited by Werner Dubitzky, Olaf Wolkenhauer, Kwang Hyun Cho and Hiroki Yokota, 720–23. New York: Springer.
- Cabin, R. J. 2011. *Intelligent Tinkering: Bridging the Gap between Science and Practice*. Washington, D.C.: Island Press.
- Campbell, Donald T. 1969. "Reforms as Experiments." *American Psychologist* 24 (4):409–29.
- Campbell, Donald T. 1974. "Unjustified Variation and Selective Retention in Scientific Discovery." In *Studies in the Philosophy of Biology*, edited by Francisco J. Ayala and Theodosius Dobzhansky, 139–61. New York: Springer.
- Campbell, Donald T. 1998. "The Experimenting Society." In *The Experimenting Society: Essays in Honour of Donald T. Campbell*, edited by William N. Dunn, 35–68. New Brunswick, NJ: Transaction Publishers.
- Cartwright, Nancy. 2007. "Are RCTs the Gold Standard?" *BioSocieties* 2 (1):11–20.
- Cartwright, Nancy. 2011. "A Philosopher's View of the Long Road from RCTs to Effectiveness." *The Lancet* 377:1400–401.
- Ceschin, F. 2014. "How the Design of Socio-technical Experiments can Enable Radical Changes for Sustainability," *International Journal of Design*, 8(3): 1–21.
- Collins, Harry M. 1988. "Public Experiments and Displays of Virtuosity: The Core-Set Revisited." *Social Studies of Science* 18 (4):725–48.
- Cook, Thomas D. 2003. "Why Have Educational Evaluators Chosen not to do Randomized Experiments?" *The Annals of the American Academy of Political and Social Science* 589 (1):114–49.
- Cox, David R., and Nancy Reid. 2000. *The Theory of the Design of Experiments*. London: Chapman & Hall.
- Criado, Tomàs S., Israel Rodríguez-Giralt, and Arriana Mencaroni. 2015. "Care in the (Critical) Making. Open Prototyping, or the Radicalisation of Independent-Living Politics." *ALTER-European Journal of Disability Research/Revue Européenne de Recherche sur le Handicap* 10 (1):24–39.
- Cumming, Graeme S., Per Olsson, F.S. Chapin III, and C.S. Holling. 2013. "Resilience, Experimentation, and Scale Mismatches in Social-Ecological Landscapes." *Landscape Ecology* 28 (6):1139–50.
- Dehue, Trudy. 2001. "Establishing the Experimenting Society: The Historical Origin of Social Experimentation According to the Randomized Controlled Design." *The American Journal of Psychology* 114 (2):283.
- Devon, Richard, and Ibo van de Poel. 2004. "Design Ethics: The Social Ethics Paradigm." *International Journal of Engineering Education* 20 (3):461–69.
- Doorn, Neelke. 2015. "Governance Experiments in Water Management: From Interests to Building Blocks." *Science and Engineering Ethics* 22 (3):1–20.
- Doorn, Neelke, Shannon Spruit, and Zoe Robaey. 2016. "Editors' Overview: Experiments, Ethics, and New Technologies." (editorial) *Science and Engineering Ethics* 22:607–11.
- Dorf, Michael C., and Charles Sabel. 1998. "A Constitution of Democratic Experimentalism." *Columbia Law Review* 98:267–473.
- Dosi, Giovanni, and Richard R. Nelson. 1994. "An Introduction to Evolutionary Theories in Economics." *Journal of Evolutionary Economics* 4 (3):153–72.
- de Wildt-Liesveld, Renée, Joske F. G. Bunders, and Barbara J. Regeer. 2015. "Governance Strategies to Enhance the Adaptive Capacity of Niche Experiments." *Environmental Innovation and Societal Transitions* 16:154–72.sDunn,

- William N. (ed.). 1998. *The Experimenting Society: Essays in Honor of Donald T. Campbell*. New Brunswick, NJ: Transaction Publishers.
- Ellerman, David P. 2004. "Parallel Experimentation and the Problem of Variation." *Knowledge, Technology & Policy* 16 (4):77–90.
- Ettelt, Stefanie, Nicholas Mays, and Pauline Allen. 2015. "Policy Experiments: Investigating Effectiveness or Confirming Direction?" *Evaluation* 21 (3):292–307.
- Farrington, David P. 2003. "A Short History of Randomized Experiments in Criminology: A Meager Feast." *Evaluation Review* 27 (3):218–27.
- Felt, Ulrike, and Brian Wynne. 2007. *Taking European Knowledge Society Seriously*. Luxembourg: DG for Research.
- Ferraro, Fabrizio, Dror Etzion, and Joel Gehman. 2015. "Tackling Grand Challenges Pragmatically: Robust Action Revisited." *Organization Studies* 36 (3):363–90.
- Gerber, Andrew, and Donald Green. 2008. "Field Experiments and Natural Experiments." In *The Oxford Handbook of Political Methodology*, edited by Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier, 357–81. Oxford: Oxford University Press.
- Greenstone, Michael, and Ted Gayer. 2009. "Quasi-Experimental and Experimental Approaches to Environmental Economics." *Journal of Environmental Economics and Management* 57 (1):21–44.
- Gross, Matthias. (2010). *Ignorance and Surprise: Science, Society, and Ecological Design*. Cambridge, MA: MIT Press.
- Gross, Matthias, and Holger Hoffmann-Riem. 2005. "Ecological Restoration as a Real-World Experiment: Designing Robust Implementation Strategies in an Urban Environment." *Public Understanding of Science* 14 (3):269–84.
- Gross, Matthias, and Wolfgang Krohn. 2005. "Society as Experiment: Sociological Foundations for a Self-Experimental Society." *History of the Human Sciences* 18 (2):63–86.
- Hacking, Ian. 1982. "Experimentation and Scientific Realism." *Philosophical Topics* 13 (1):71–87.
- Hacking, Ian. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.
- Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42 (4):1009–55.
- Hoffmann, Matthew J. 2011. *Climate Governance at the Crossroads: Experimenting with a Global Response after Kyoto*. Oxford: Oxford University Press.
- Hoogma, Remco, Rene Kemp, Johan Schot, and Bernhard Truffer. 2002. *Experimenting for Sustainable Transport: The Approach of Strategic Niche Management*. London: Taylor & Francis.
- Irvine, Katherine N., and Stephen Kaplan. 2001. "Coping with Change: The Small Experiment as a Strategic Approach to Environmental Sustainability." *Environmental Management* 28 (6):713–25.
- Karvonen, A. and van Heur, B. (2014). "Urban laboratories: Experiments in reworking cities," *International Journal of Urban and Regional Research*, 38(2): 379–392.
- Kemp, René. 2010. "The Dutch Energy Transition Approach." *International Economics and Economic Policy* 7:291–316.
- Kerber, W., & Eckardt, M. 2007. "Policy Learning in Europe: The Open Method of Co-ordination and Laboratory Federalism," *Journal of European Public Policy*, 14(2): 227–247.

- Kern, Florian, and Adrian Smith. 2008. "Restructuring Energy Systems for Sustainability? Energy Transition Policy in the Netherlands." *Energy Policy* 36 (11):4093–103.
- Kroes, Peter. 2015. "Experiments on Socio-Technical Systems: The Problem of Control." *Science and Engineering Ethics* 22:633–45.
- Krohn, W. 2007. "Nature, Technology, and the Acknowledgment of Waste," *Nature and Culture*, 2(2): 139–160.
- Krohn, Wolfgang, and Johannes Weyer. 1994. "Society as a Laboratory: The Social Risks of Experimental Research." *Science and Public Policy* 21 (3):173–83.
- Krohn, Wolfgang, and Peter Weingart. 1987. "Commentary: Nuclear Power as a Social Experiment—European Political "Fall Out" from the Chernobyl Meltdown." *Science, Technology & Human Values* 12 (2):52–58.
- Lilford, Richard J., and Jennifer Jackson. 1995. "Equipoise and the Ethics of Randomization." *Journal of the Royal Society of Medicine* 88 (10):552–59.
- Lipinski, Christopher, and Andrew Hopkins. 2004. "Navigating Chemical Space for Biology and Medicine." *Nature* 432 (7019):855–61.
- Lorimer, Jamie, and Clemens Driessen. 2014. "Wild Experiments at the Oostvaardersplassen: Rethinking Environmentalism in the Anthropocene." *Transactions of the Institute of British Geographers* 39 (2):169–181.
- Lynn, Gary S., Joseph G. Morone, and Albert S. Paulson. 1996. "Marketing and Discontinuous Innovation: The Probe and Learn process." *California Management Review* 38 (3):8–37.
- Manzi, Jim. 2012. *Uncontrolled: The Surprising Payoff of Trial-and-error for Business, Politics, and Society*. New York: Basic Books.
- Markusson, Nils, Atushi Ishii, and Jennie C. Stephens. 2011. "The Social and Political Complexities of Learning in Carbon Capture and Storage Demonstration Projects." *Global Environmental Change* 21 (2):293–302.
- Martiniuk, Alexandra L., Seye Abimbola, and Merrick Zwarenstein. 2015. "Evaluation as Evolution: A Darwinian Proposal for Health Policy and Systems Research." *Health Research Policy and Systems* 13 (1):1–5.
- Mayr, Ernst. 2001. *What Evolution Is*. New York: Basic Books.
- Menand, Louis. 2001. *The Metaphysical Club*. New York: Farrar, Straus, and Giroux.
- Nair, Sreeja, and Dimple Roy. 2009. "Promoting Variation." In *Creating Adaptive Policies: A Guide for Policymaking in an Uncertain World*, edited by Darren Swanson and Bhadwal Suruchi, 95–105. Thousand Oaks: SAGE Publications.
- Nair, Sreeja, and Michael Howlett. 2015a. "Scaling up of Policy Experiments and Pilots: A Qualitative Comparative Analysis and Lessons for the Water Sector." *Water Resources Management* 29 (14):4945–961.
- Nair, Sreeja, and Michael Howlett. 2015b. "Meaning and Power in the Design and Development of Policy Experiments." *Futures* 76:67–74.
- Oakley, Ann. 1998. "Public Policy Experimentation: Lessons from America." *Policy Studies* 19 (2):93–114.
- Oakley, Ann, Vicki Strange, Tami Toroyan, Meg Wiggins, Ian Roberts, and Judith Stephenson. (2003). "Using Random Allocation to Evaluate Social Interventions: Three Recent UK Examples." *The Annals of the American Academy of Political*

- and *Social Science* 589 (1):170–89. Oates, Wallace E. 1999. “An Essay on Fiscal Federalism.” *Journal of Economic Literature* 37 (3):1120–149.
- Pisano, Gary P. 1996. “Learning-before-doing in the Development of New Process Technology.” *Research Policy* 25 (7):1097–119.
- Polsby, Nelson W. 1984. *Political Innovation in America: The Politics of Policy Initiation*. New Haven, CT: Yale University Press.
- Raven, Rob, Eva Heiskanen, Raimo Lovio, Mike Hodson, and Bettina Brohmann. 2008. “The Contribution of Local Experiments and Negotiation Processes to Field-level Learning in Emerging (Niche) Technologies. Meta-analysis of 27 New Energy Projects in Europe.” *Bulletin of Science, Technology & Society* 28 (6):464–77.
- Raven, Rob, Florian Kern, Bram Verhees, and Adrian Smith. 2016. “Niche Construction and Empowerment Through Socio-Political Work. A Meta-Analysis of Six Low-Carbon Technology Cases.” *Environmental Innovation and Societal Transitions* 18:164–80.
- Saam, N. J., & Kerber, W. 2013. “Policy Innovation, Decentralised Experimentation, and Laboratory Federalism,” *Journal of Artificial Societies and Social Simulation*, 16, 1: 1–15.
- Sabel, Charles F., and Jonathan Zeitlin. 2010. *Experimentalist Governance in the European Union: Towards a New Architecture*. Oxford: Oxford University Press.
- Sabel, Charles F., and Jonathan Zeitlin. 2012. “Experimentalist Governance.” In *Oxford Handbook of Governance*, edited by David Levi-Faur, 169–83. Oxford: Oxford University Press.
- Sanderson, Ian. 2002. “Evaluation, Policy Learning and Evidence-Based Policy Making.” *Public Administration* 80 (1):1–22.
- Schiaffonati, Viola. 2015. “Stretching the Traditional Notion of Experiment in Computing: Explorative Experiments.” *Science and Engineering Ethics* 22:1–19.
- Schön, Donald A. 1983. *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
- Schot, Johan, and Frank W. Geels. 2007. “Niches in Evolutionary Theories of Technical Change.” *Journal of Evolutionary Economics* 17 (5):605–22.
- Steffen, Dagmar. 2014. “New Experimentalism in Design Research: Characteristics and Interferences of Experiments in Science, the Arts and in Design Research.” *Artifact* 3 (2):1–16.
- Steinle, Friedrich. 1997. “Entering New Fields: Exploratory Uses of Experimentation.” *Philosophy of Science* 64, S65–S74.
- Stoker, Gerry, and Peter John. 2009. “Design Experiments: Engaging Policy Makers in the Search for Evidence About What Works.” *Political Studies* 57:356–73.
- Thomke, Stefan H. 2003. *Experimentation Matters: Unlocking the Potential of New Technologies for Innovation*. Cambridge, MA: Harvard Business Press.
- van den Bosch, Suzanne J. M. 2010. “Transition Experiments: Exploring Societal Changes Towards Sustainability.” PhD diss., Erasmus University Rotterdam.
- Van der Heijden, Jeroen. 2014. “Experimentation in Policy Design: Insights from the Building Sector.” *Policy Sciences* 47 (3):249–66.
- Van de Poel, Ibo. 2011. “Nuclear Energy as a Social Experiment.” *Ethics, Policy & Environment* 14 (3):285–90.
- Van de Poel, Ibo. 2015. “An Ethical Framework for Evaluating Experimental Technology.” *Science and Engineering Ethics* 22:667–86.

- Verhees, B., Raven, R., Veraart, F., Smith, A., & Kern, F. 2013. The Development of Solar PV in The Netherlands: A Case of Survival in Unfriendly Contexts," *Renewable and Sustainable Energy Reviews*, 19: 275–289.
- Vreugdenhil, Heleen, Jill Slinger, Wil Thissen, and Philippe Ker Rault. 2010. "Pilot Projects in Water Management." *Ecology and Society* 15 (3):1–26.
- Weber, Eric T. 2011. "What Experimentalism Means in Ethics." *Journal of Speculative Philosophy* 25:98–115.
- Weisburd, David. 2000. "Randomized Experiments in Criminal Justice Policy: Prospects and Problems." *Crime & Delinquency* 46 (2):181–93.

3 Moral experimentation with new technology

Ibo van de Poel

Introduction

By conceiving of new technologies as social experiments, attention is drawn to issues such as how to learn from such experiments while minimizing harm to society, and under what conditions we consider these experiments socially and morally acceptable. In terms of learning, I have argued in an earlier publication that social experimentation with new technology in society may result in three different types of learning, i.e., learning about the impacts of a technology in society (impact learning), learning about the institutions that are needed to properly embed technology in society (institutional learning), and learning about normative and moral issues (moral learning) (Van de Poel 2017).

In this chapter, I focus on moral learning during the experimental introduction of new technologies into society. This learning sometimes concerns the discovery of new moral issues that are triggered by the introduction of new technology. An example is that the small-scale experimental introduction of Google Glass revealed not only the expected privacy issues in terms of what information should be shared with whom, but also moral issues that relate to (perceived) physical intrusion in the private sphere and social norms of interaction (Kudina and Verbeek submitted). These issues might have contributed to the reasons why Google has, at least for the time being, decided not to introduce Google Glass to the consumer market. Moral learning may also take place with respect to moral frameworks, and norms and values that we use to judge a technology morally. An example is the notion of meaningful human control that was developed in response to moral and responsibility issues brought about by the use of military drones and other automated weapon systems (UNIDIR 2014; Horowitz and Scharre 2015). This notion is useful to evaluate technological development and can help guide morally the development of new technology.

In this connection, I will speak of moral experiments in this chapter. By using this term, I do not wish to suggest that the moral experiment of introducing a new technology into society can be neatly separated from what may be called the institutional experiment or the experiment in which we learn about impacts. In reality, these experiments will be intertwined. The notion of

moral experiment, then, refers to an analytical category in which the focus is on experimental moral learning. Using the terminology of moral experiments allows me to connect to ideas on experimentation in moral philosophy.

One way in which moral experimentation takes place is when a technology is actually introduced into society. However, this is not the only mode of moral experimentation with new technology. It may also take place before a technology is introduced into society, for example, through thought experiments. Moral thought experiments typically involve the thinking out of possible scenarios that anticipate what will happen and how people will morally react when a technology is actually introduced into society. Although such thought experiments may not be able to anticipate all moral issues that arise, they do facilitate forms of moral learning before the introduction of a new technology.

My aim in this chapter is to describe three different modes of moral experimentation with new technology—thought experiments, experiments in living, and social experiments. I explore their distinct possibilities and limitations for moral learning and shed light on the different kinds of ethical issues that they each raise. The chapter is structured as follows: I start with a very general characterization of moral experiments that is appropriate for the main aim of this chapter. Next, I discuss in detail three possible modes of moral experimentation with new technology, i.e., thought experiments, experiments in living, and social experiments. In the next section, I systematically look at what types of moral learning these three types of experiments offer. After that, I briefly discuss ethical issues that each of the three modes of moral experimentation may raise, and I end with an outlook on the role of moral experimentation in the introduction of new technology into society.

What are moral experiments?

Moral experiments involve an *intervention* in the real, or an imagined, world that results in new *experiences* (broadly conceived), which are deliberately and systematically used for moral *learning* or for attempts to morally learn from these experiences. I include interventions in imagined (or imaginary) worlds because I want to include the category of thought experiments that is important in moral philosophy. I take the notion of experiences broadly; it includes not only observations but also thoughts and moral judgments about cases or situations. In my definition, deliberate and systematic attempts to learn morally are crucial. However, that does not necessarily mean that such learning is the sole aim of doing the experiment.

I take moral learning to include, but not be limited to, the learning about the following issues:

- 1 New moral issues raised by the introduction of a new technology into society;
- 2 Ways to specify or apply existing normative standards (rules, principles, norms, values) to judge the introduction of a new technology into society or the moral issues raised by it;

- 3 The discovery of the relevance of new normative standards to judge the introduction of a new technology and moral issues raised by it.

In sum, I am interested in moral learning with the aim to improve both moral judgements of the introduction of a new technology into society and the process of that introduction.

My characterization of moral experiments is deliberately broad. It is, for example, much broader than experiments that aim at hypothesis testing. It also does not suppose that experiments are controlled because the category of moral experiments I am interested in here is neither aimed at causal knowledge nor at hypothesis testing. It is not aimed at causal knowledge because none of the topics I consider for which learning may occur involves causal mechanisms. Still, moral experiments in principle could involve hypothesis testing; not of scientific theories but of moral theories. The way experiments have been used in moral philosophy (see e.g. Alfano and Loeb 2016) often seems to include or focus on hypothesis testing. Such testing indeed seems appropriate if we focus on learning about existing moral theories, insights, conceptions of the good, etc. However, my goal is not to illuminate moral theories but rather moral learning in relation to the introduction of a technology into society.

Nevertheless, in further developing the idea of doing moral experiments with new technology, I will take inspiration from the use of moral experiments in moral philosophy. Recently, experimental moral philosophy has been articulated as an area of moral philosophy. However, the idea of experimentation in moral philosophy is not new. For example, John Dewey stressed the importance of experimentation in moral philosophy. Appiah (2008) contends that experimentation, and more generally, empirical investigations, have long been a proper part of moral philosophy but were excluded from moral philosophy in the twentieth century, especially by analytical philosophers attempting to distinguish their discipline from psychology. He points out, for example, that the subtitle of Hume's *A Treatise of Human Nature* (1739) is *Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*.

One important role of moral experiments in moral philosophy is to test moral hypotheses. Dewey, for example, argues that moral principles should be seen as hypotheses that are tested in experiments:

Principles exist as hypothesis with which to experiment. Human history is long. There is a long record of past experimentation in conduct, and there are cumulative verifications which give many principles a well-earned prestige. Lightly to disregard of them is the height of foolishness. But social situations alter, and it is also foolish not to observe how old principles actually work under new conditions and not to modify them so that they will be more effectual instruments in judging new cases.

(Dewey 1922, 239)

Weber (2011, 100) even holds that true experimentalism in ethics always involves hypothesis testing (and making adjustments before one tries again).¹ Nevertheless, there are also moral experiments that do not aim primarily at hypothesis testing—namely, thought experiments. Although thought experiments may play a role in hypothesis testing, they may also have a number of other uses in moral theorizing. For example, they may be used to make a theoretical point or to provide difficult cases that a moral theory should be able to elucidate.

Three types of moral experiments

I explore three types of moral experiments: thought experiments, experiments in living, and social experiments. The first two already play a role in moral philosophy and I will discuss in what way they can also be employed as moral experiments with new technology. The third, social experiments, has been discussed in the literature mainly in the context of policymaking or state formation. I argue that they may also be used as a model for moral experiments with new technology.

These three types each offer distinct possibilities for moral learning, and thus all three play a role in morally experimenting with new technology. As will become clear, social experiments may offer the most ample opportunities for moral learning with respect to new technologies, but because this learning takes place in society, learning may be gained at considerable social costs; therefore, they may raise serious ethical concerns. Thought experiments and experiments in living in general accrue lower social costs and are ethically less questionable but what can be learned may be less encompassing than in the case of social experiments. Additionally, the external validity of such experiments may be lower than social experiments.

Thought experiments

Thought experiments are quite popular in moral philosophy. Famous examples are the trolley problem and experience machine. In the trolley problem, there is a runaway trolley that is approaching five people tied to the railway track and the trolley is about to kill them. However, there is a switch in the track that allows you to send the trolley to another track where it will kill only one person. You are standing some distance away from the track next to a lever with which you can turn the switch (or not). The question is what you should do. Over time, the trolley problem, which was introduced by Philippa Foot (1967),² has developed in many variations and has been taken up outside of philosophy, in particular in psychology and neuroscience.

The experience machine is a thought experiment in which you have the choice to enter a machine that gives you all the pleasurable experiences you would like to have. However, entering the machine would mean that you would have to give up your normal life and would be tied to the machine the

rest of your life. Robert Nozick (1974, 42–45) devised this thought experiment in an attempt to show that happiness was more than just experiences of pleasure (as hedonism holds). He believed that people would choose not to enter the experience machine because they want real experiences that make them happy rather than the mere illusion or feeling of happiness.³

The experience machine is a typical example of how thought experiments are used in philosophy. A hypothetical scenario is presented, and the reply to the scenario is usually taken for granted or assumed to show a general or theoretical point.⁴ For example, when we wonder whether torture can ever be acceptable, philosophers (and others) have come up with (far-fetched) scenarios in which torture seems necessary to avoid a disaster to show that, in some situations, torture is justifiable (cf. Shue 1978; Allhoff 2005; Davis 2005).

Typically, such traditional thoughts experiments in philosophy are doubly fictional. They are fictional in the sense that the scenario presented is not real but invented to make a point. They are also fictional in the sense that the response is taken for granted or assumed in contrast to, for example, a survey done to test the response to the scenario. Both types of fictionality may be problematic.

Fictional scenarios are problematic when they describe things that are physically or logically impossible, as it is questionable whether anything follows from our reaction to an impossible scenario (Davis 2012). Even if they are not strictly impossible, it may be problematic to draw conclusions from far-fetched or unlikely scenarios. Of course, unlikely but possible scenarios may perhaps falsify general philosophical theories, but in other cases it may be doubtful what general conclusions can be drawn from possible but unlikely scenarios. For example, should we allow torture by law if it turns out that there are some unlikely scenarios in which we would want to allow torture? One might well argue that despite these exceptional cases, there are still good reasons to forbid torture by law.

Traditional thought experiments are often fictional because the response to them is taken as given or assumed. This is not to say that the response cannot be contested but rather that it is not empirically tested. Sorensen (1992, 250–52) has characterized thought experiments as those that are not executed and therefore do not produce “fresh information.” Given this characterization, one might wonder whether unexecuted thought experiments fit my definition of moral experiments because they do not result in new experiences or new observations.⁵

Still, there is also a category of moral thought experiments that result in new experiences, such as those invoked by Dewey when he characterizes moral deliberation as dramatic rehearsal:

Deliberation is actually an imaginative rehearsal of various courses of conduct. We give way, in our mind, to some impulse; we try, in our mind, some plan. Following its career through various steps, we find

ourselves in imagination in the presence of the consequences that would follow: and as we then like and approve, or dislike and disapprove, these consequences, we find the original impulse or plan good or bad.

(Dewey and Tufts 1908, 323)

It should be noted that dramatic rehearsal, although it operates through what has been called moral imagination, produces real experiences (“we like and approve or dislike and disapprove consequences”), not just imagined or fictional experiences, just like reading a novel or watching a movie leads to real experiences even if the novel or movie are fictional.

Such experiences or reactions to thought experiments may also be systematically gathered. For example, the trolley problem has set off a large range of empirical experiments in which the reactions of people to it (and its many variations) are investigated rather than assumed. Nevertheless, even in such cases, thought experiments may give rise to questions about their external validity.⁶ The issue is not simply whether the group investigated is representative of the larger population, but also whether people’s reaction to fictional or imagined scenarios is representative for how they would react if these scenarios actually occurred. Another issue, as Michael Davis (2012) has pointed out, is that thought experiments usually invoke uncommon situations and people’s reactions to uncommon situations may be a less reliable test of their moral commitments than their reactions to more common situations. Another problem is that people’s reactions to fictional scenarios are subject to framing effects, i.e., to the way the scenario is presented rather than the content of the scenario. Indeed, for trolley experiments it has been argued that their external validity is quite low; it is questionable whether people’s reactions to a typical trolley problem say anything about how they would react in real-life situations or if they reveal their underlying moral commitments or judgements (Bauman et al. 2014).

Although the fictional character of thought experiments has drawbacks, it has advantages as well (cf. Dewey and Tufts 1908, 323–24; Sorensen 1992, 197–202). One is that, unlike real experiments, thought experiments do not have real consequences, so no harm ensues if a thought experiment is done.⁷ Another advantage is that we can easily perform a series of thought experiments, for example, by varying certain parameters of the situation, while it might require too much time and effort, or be more difficult or even impossible, to execute a series of real experiments. Thought experiments also allow us to investigate people’s moral reactions to scenarios that cannot, or at least cannot yet, be realized but which may nevertheless arise in the future, for example, when a new technology is introduced into society.

Given these advantages, it is not surprising to see that thought experiments have also become popular in the sociology and philosophy of technology to explore the ethical issues raised by new technologies. For example, in the Value Sensitive Design (VSD) approach, which aims at integrating moral values early in a technical design process, so-called “value scenarios” are a

technique to investigate the possible value implications and moral issues raised by new technologies (Davis and Nathan 2015). Swierstra, Stemerding, and Boenink (2009) and Boenink, Swierstra, and Stemerding (2010) propose techno-moral scenarios or vignettes (see also Lucivero 2016). These are “narratives evoking alternative future worlds” (Swierstra, Stemerding, and Boenink 2009, 120) that, in particular, address the interaction between technology and morality. That is to say, they do not only address morally relevant consequences of new technologies but also how the introduction of new technology may lead to moral change, i.e., changes in the moral standards. Like value scenarios, the aim of these scenarios is not to predict what will happen but rather to enable moral deliberation about new technologies and their consequences, including moral change. Another form that moral thought experiments with new technology may take are the theatrical debates described in the contribution by Koppers to this volume (Chapter 4). He explicitly refers to Dewey’s notion of dramatic rehearsal to understand interactive theatrical debates that employ scenarios and are enacted as collective thought experiments.

These thought experiments about new technology are distinct from traditional (philosophical) thought experiments in a number of ways. First, they do not aim at hypothesis testing or making a theoretical point. Rather, they aim at forms of moral learning that more directly relate to the introduction of a new technology. Second, they result in new moral experiences (or judgments), albeit in response to fictional scenarios and as such are appropriately called experiments according to my definition. Third, they are often not just aimed at new individual moral judgments about possible future technology scenarios, but also at more collective forms of deliberation.

Thought experiments clearly have a place in considering the future moral consequences of new technologies. However, despite their advantages, they also share disadvantages with traditional moral thought experiments. First, they are based on hypothetical scenarios that may be far-fetched or unlikely. Second, even if they are tested on real people, e.g., through surveys or deliberative methods, it remains unclear if participants’ reactions to these scenarios would be the same when the technologies were actually introduced into society. The real test of new technologies therefore requires more than a thought experiment.

Experiments in living

I now turn to a second category of moral experiments: “experiments in living,” a notion that was coined by John Stuart Mill in his *On liberty* (1869). An experiment in living is one that an individual performs by living his or her life in a certain way; what is tested out in such an experiment is a certain mode of living or a certain conception of the good life, and by living that life, the individual will learn whether it brings the goods it promises. According to Mill, such experiments in living are needed because mankind

is imperfect and “the worth of different modes of life should be proved practically” (Mill 1869, III.1). According to Mill, the government should allow a diversity of experiments in living rather than enforce its own vision of the good life:

Government operations tend to be everywhere alike. With individuals and voluntary associations, on the contrary, there are varied experiments, and endless diversity of experience. What the State can usefully do, is to make itself a central depository, and active circulator and diffuser, of the experience resulting from many trials. Its business is to enable each experimentalist to benefit from the experiments of others, instead of tolerating no experiments but its own

(Mill 1869, V.19)

Elisabeth Anderson (1991) has reconstructed Mill’s own life as an experiment in living. Mill’s father raised him according to the principles of the utilitarian philosophy of Jeremy Bentham. He believed that a happy life results from rational calculation and requires taming the sentiments as potentially dangerous dispositions. However, in 1826, Mill fell into a severe depression. He only recovered when he started reading poetry, which was discouraged by Benthamite principles. According to Anderson, Bentham’s theory failed Mill in his own experiment in living. First, it could not explain the onset of the depression. Second, it did not offer a successful remedy. Only when Mill started reading poetry—after having tried various strategies suggested by traditional utilitarianism to overcome the depression—did he begin to recover. In response to this experience, Mill developed his own version of utilitarianism that distinguishes higher from lower pleasures and maintains that rather than aiming to be happy, happiness arises as the result of useful endeavors.

A few things are to be noted about Mill’s own experiment in living. First, it delivered him an experience that he could probably never have had in a thought experiment. Mill was convinced that living according to Bentham’s theory would make him happy, and had he done a thought experiment, it would probably only have confirmed his conviction. Second, partly due to contingent circumstances, Mill’s experiment in living almost perfectly met the conditions of Bentham’s theory; it was, for example, free of contradictory influences such as religion (Anderson 1991, 15). Therefore, it amounted to a hypothesis-testing experiment under more or less controlled circumstances, and it turned into a refutation of Bentham’s theory (at least for Mill).

However, the broader notion of experiment in living that Mill uses in *On liberty* does not assume hypothesis testing. Although people’s experiments in living presumably are based on a notion of how to live a good life, which may be touted a hypothesis, in most cases this would not amount to the testing of hypotheses that are based on well-formulated moral theory, as in the case of Mill’s own experiment in living. The degree to which Mills’

actual life met the conditions of Bentham's theory also was exceptional; most experiments of living typically occur in less controlled circumstances. Moreover, a person's experience would, in most cases, not just be a confirmation or refutation of an underlying notion of a good life, but would rather result in smaller or larger adaptations in how that person lives his or her life rather than the paradigmatic change (also in moral theory) that it triggered in Mill's case.

This broader notion of experiment in living is useful for morally experimenting with new technology. There are quite a few examples in which the use of new technology amounts to an experiment in living. Think for example of people who experiment with drugs that improve cognitive or physical performances. As Vincent and Jane point out in Chapter 6 in this volume, a large variety of people are already experimenting on themselves with putative cognitive enhancement technologies, expecting benefits such as "superior memory, focus, reflexes, calmness, clarity of thought, problem-solving ability, mental stamina, and ability to function well with little sleep (this volume, 125)."

Consider the use of such devices as smart phones that influence how people live their lives and organize their contacts, or social media such as Facebook and Twitter. Some of these technologies indeed promise to give people a better or happier life with satisfying social interaction. Often the experimentation with such technologies, and the learning from it, remains implicit. However, sometimes it becomes more explicit, as in the case of Mat Honan who experimented for a few months on himself with Google Glass. Google Glass is a wearable computer in the shape of eyeglasses. It functions like a smart phone but can be operated hands-free as it reacts to voice commands.

In April 2013, Google made Glass available for a limited number of people (so-called Glass Explorers) willing to pay 1500 dollars to test it. Honan describes his experiences in an article in *Wired* in December 2013. He sums up what he learned from experimenting with Google Glass:

Even in less intimate situations, Glass is socially awkward. Again and again, I made people very uncomfortable. That made me very uncomfortable.... People get angry at Glass. They get angry at you for wearing Glass. They talk about you openly.... My Glass experiences have left me a little wary of wearables because I'm never sure where they're welcome. I'm not wearing my \$1,500 face computer on public transit where there's a good chance it might be yanked from my face. I won't wear it out to dinner, because it seems as rude as holding a phone in my hand during a meal. I won't wear it to a bar. I won't wear it to a movie. I can't wear it to the playground or my kid's school because sometimes it scares children. It is pretty great when you are on the road—as long as you are not around other people, or do not care when they think you're a knob.

(Honan 2013)

Honan's experience may be described as an experiment in living. Although wearing Google Glass for almost a year is perhaps not as comprehensive as the experiment in living that J.S. Mill did, there are interesting parallels. Like Mill, Honan was testing a way of living and he was expecting the best of it. Like Mill, his experience was different in important ways from what he expected and in ways that a thought experiment could never have revealed to him. What is also interesting is that in Honan's case, the learning extended beyond what was tried out in his experiment in living:

Glass kind of made me hate my phone—or any phone. It made me realize how much they have captured our attention. Phones separate us from our lives in all sorts of ways. Here we are together, looking at little screens, interacting (at best) with people who aren't here.... Glass helped me appreciate what a monster I have become, tethered to the thing in my pocket. I'm too absent.

(Honan 2013)

Nevertheless, Honan still believes that wearables are the future and that we better prepare to live with them.

Technological experiments in living, thus, are already taking place at quite a scale. However, in many cases they are currently not done systematically and they do not amount to deliberate attempts for moral learning. They could, however, be turned into more deliberate and systematic attempts at moral learning, for example, by pooling experiences and deliberating on them, not unlike what Mill suggests that the State should do with respect to individuals' experiments in living.

Experiments in living may be called moral experiments because they try out morally loaded visions of the good life in an individual's life. They may lead to new experiences and new moral insights, and also to the awareness of new moral issues. By their nature, they are individualistic enterprises as they test out moral implications for the individual, but this is not to deny that they may have broader moral implications in society. Take the use of enhancement drugs that is discussed in the chapter by Vincent and Jane (Chapter 6). Increased possibilities for human enhancement are likely to have individual moral ramifications. Individuals may now be confronted with choices they did not have before (i.e., to enhance themselves or not), and they may not perceive having such choices as moral progress; in addition, the use of enhancers may have a variety of effects on human health and well-being. Aside from these more individual effects, there are also social moral issues; for example, the use of enhancers may raise justice issues (as some people can perhaps afford them while others cannot), or it may create undesirable social pressure on individuals, which is also a social rather than just an individual moral issue. Social moral issues may already be faced in individual experiments in living. Since individuals are always embedded in a web of social relations, experiments in living are never completely individualistic;

they always have social consequences. However, testing whether socially-encompassing moral issues indeed occur, and morally experimenting with them, requires more than just an experiment in living. If we want to empirically verify, rather than just speculate, whether human enhancement raises problems of social justice, and if we want to experimentally find out how we should morally respond to these issues, what is needed is more than an experiment in living. An experiment in living will not be enough to recognize, and morally experiment with, the social justice issues. Rather, learning requires a social form of experimentation that likely involves people who have not voluntarily chosen to be part of the social experiment.

Social experiments

I now will consider a third mode of moral experimentation which I will call social experiments. These experiments are different from thought experiments in that they take place in society rather than being fictional. They are different from experiments in living that take place in society but merely concern the life of an individual. Social experiments are social in nature, concern a broader range of societal moral issues, and involve a broader range of people.

Classical examples of such social experiments are those with policy, regulation, law, or the formation of states. According to John Dewey, all such processes are essentially experimental in nature. For example, in his *Logic* he writes that “every measure of policy put into operation is, logically, and should be actually, of the nature of an experiment” (Dewey 1938, 508–509). Similarly, in *The public and its problems*, Dewey states that the “formation of states must be an experimental process” (Dewey 1927, 32).

It is interesting that John Stuart Mill advocated experiments in living but believed that the formation of governments and in fact most of moral philosophy could not be experimental. He writes:

There is a property common to almost all the moral sciences, and by which they are distinguished from many of the physical; this is, that it is seldom in our power to make experiments in them. In chemistry and natural philosophy, we can not only observe what happens under all the combinations of circumstances which nature brings together, but we may also try an indefinite number of new combinations. This we can seldom do in ethical, and scarcely ever in political science. We cannot try forms of government and systems of national policy on a diminutive scale in our laboratories, shaping our experiments as we think they may most conduce to the advancement of knowledge.

(Mill 1874, V.51)

Mill argues that controlled experimentation is not possible in moral science, and hence we cannot develop moral knowledge through experimentation.

Moral and, in particular, political science must therefore be a deductive endeavor. While Mill was no doubt right that there are limits to controlled experimentation in political and moral science, it does not follow that such sciences are necessarily deductive. In fact, new experiences gained, sometimes painfully, in for example new forms of government or policies, have led to new political and moral insights.

In as far as experiments with state formation or government policies cannot be controlled experiments, it indeed follows that these are not moral experiments that can generate *general* moral knowledge. However, they can still be moral *design* experiments, and as we will see below, that is indeed how they have been understood by Dewey and other authors.

It should also be noted that Mill himself did not rule out the possibility of uncontrolled experiments in morality as he advocated experiments in living, which are largely uncontrolled moral experiments. Nevertheless, as Mill suggests, there is a difference between experiments in living and experiments with forms of government. While many experiments in living can be done simultaneously (by different people), it is more difficult, if not impossible, to experiment with different forms of government in parallel. This may be seen as an argument against the possibility of what Ansell and Bartenberger in Chapter 2 call “evolutionary experimentation” but not against what they call “generative experimentation” in which, for example, a form of government or policy is continuously adapted in the process of experimentation and learning.

It is indeed this kind of generative experimentation that John Dewey has in mind when he talks about the application of the experimental method in political and moral investigations. He recognizes that such experiments cannot be controlled but argues that it is better to call new policies, or in our case new technologies in society, experiments because it offers additional possibilities for learning which we would otherwise be lacking (Dewey 1938, 509). Moreover, he stresses the importance of the experimental *method* of inquiry, in which “policies and proposals for social action” are “treated as working hypotheses” and are “subject to constant and well-equipped observation of the consequences they entail when acted upon, and subject to ready and flexible revision in the light of observed consequences” (Dewey 1927, 202).

The learning that takes place in such experimentation is not about general causal mechanisms but rather about the actual policy that is introduced into society and, in the process, learning how to revise and improve that policy in light of the developing consequences. This mode of learning is similar to what Karl Popper has called “piecemeal social experiments,” which involve “repeated experiments and continuous readjustments” (Popper 1945, 144).

The idea of generative experiments in policy making, law, regulation, or state formation can be extended to the introduction of new technology into society. In fact, authors like Winner (1980), Lessig (1999), and Sclove (1995) have argued that technologies regulate human behavior not unlike social structures, laws, or regulations. As we morally judge laws, regulations, and

social structures, we may similarly judge technologies morally for how they regulate and influence human behavior. In the case of laws and policies, these evaluations have an experimental component because we only discover the full consequences when they are actually implemented. Similarly, the moral evaluation of new technologies also has an experimental component because some consequences only become clear when they are actually implemented.

This is not to deny that there are differences between laws, regulations, and policies on the one hand and technologies on the other hand. Technologies have a material component whereas laws and policies in most cases lack one. With laws and policies, it is usually clear how they intend to regulate or influence human behavior, while in the case of technology, the influence on human behavior may be implicit or unintended. These factors may make it more complicated to morally evaluate new technologies, and increase rather than reduce the experimental component in evaluation of new technologies because exactly what technologies actually do will require ample empirical evidence.

In Chapter 9, Kokkeler and Brandse describe several social experiments with ICT technologies in health and social services for youths, some of which comprise moral experimentation. One important new use of digital technology in these services is for client dossiers and portfolios. From the viewpoint of the formal organization, control and accountability are important values to consider when using such technologies. However, for the clients, the making of their own portfolios aids them to express themselves and they find the values of self-expression, autonomy, and trust critical. This not only conflicted with the values of the organization (accountability and control) but also with the professional norm that professional workers should have access to the portfolios of their young clients to provide the best care possible. In their chapter, one of the examples they describe is a living lab in which clients could experiment with their own portfolios without interference by their professional coaches. The young clients expressed a strong preference to decide for themselves with whom to share their portfolios and they even threatened to withdraw from the experiment if this right was not granted. This created tension with the professional code of the health professionals. Nevertheless, the health professionals eventually accepted an external gatekeeper to coach their clients about what to share and with whom. The gatekeeper was accepted because the workers experienced intensified and enriched interactions with their clients who made digital portfolios rather than diminished communication, as they had expected.

This case described by Kokkeler and Brandse amounts to a moral experiment for several reasons. First, it led to the discovery, or at least a more explicit articulation, of the moral value of client autonomy and the related norm that minors should be able to decide with whom to share their portfolios. Second, it led to a moral change as this norm gradually became generally acknowledged despite the conflict with the organizational values of

accountability and control, and previous values and norms of professional health workers. Third, this moral change was accepted by professional caregivers because they *experienced* that granting more autonomy to clients improved their communication.

Learning in moral experiments

Above I distinguished several kinds of issues that might be learned in moral experiments with new technology, in particular: (1) learning about new moral issues raised by a technology, (2) learning about how to specify existing normative standards to judge these moral issues, and (3) learning about new normative standards that are required to adequately deal with the new moral issues.⁸ Below, I specify what type of learning can be expected in thought experiments, experiments in living, and social experiments with new technology.

First, the kind of learning occurring will, in part, depend on how the experiment is carried out, i.e., its set-up. I follow the three types of experimentation suggested by Ansell and Bartenberger in Chapter 2: controlled, evolutionary, and generative. In addition, I would like to suggest a fourth: explorative experimentation. I propose explorative moral experiments as those that aim to explore moral issues in connection with the anticipated introduction of a new technology into society.

With each of the methodological set-ups of moral experiments, a specific epistemological or design aim corresponds:

- Controlled experiments aim at testing hypotheses or finding causal relation;
- Explorative experiments aim at the exploration of moral issues raised by a new technology;
- Generative experiments aim at improving the process of the introduction of a new technology into society through iterative moral learning;
- Evolutionary experiments aim at selecting experimentally the best way to fulfill a task—in our case, selecting the best technology (from a moral point of view) to provide a certain function, and the best way to introduce, regulate, or use a technology.

The three types of experiments discussed in the previous section can be carried out in most of these set-ups. The only impossible combination is carrying out a thought experiment as a generative or evolutionary experiment because these set-ups require the actual introduction of a technology into (a part of) society while thought experiments take place in an imaginary world. Although the exact possibilities for moral learning in experiments will depend on their exact methodological set-up, the following general points can be made about what kind of moral learning is to be expected for the three types of experiments discussed above.

Thought experiments with new technology are explorative as they help to explore new ethical issues raised by a technology. They may also help to establish the extent to which existing normative standards are useful or appropriate to judge these ethical issues. It is doubtful whether they can also lead to learning about new specifications of existing standards or to new normative standards. As we have seen, thought experiments have a low external validity. It is therefore doubtful if they can contribute to how we would or should judge new ethical issues. This low external validity also suggests that explorative thought experiments cannot fully predict the moral issues raised by a new technology.

The advantage of experiments in living over thought experiments is that they are based on real experiences and therefore result in forms of moral learning not possible in thought experiments. In particular, they help to evaluate different ways of living made possible or stimulated by a new technology. It should be stressed that such evaluations are not simply based on the new experiences that living with the technology brings, but require normative reflection and judgement (cf. Dewey 1929, 259). As Anderson stresses with respect to Mill's experiment in living:

The crucial test for a conception of the good is that it provide a perspective of self-understanding which is both personally compelling (has normative force for the agent) and capable of explaining and resolving her predicament—the reasons for crisis and for recovery from it.

(Anderson 1991, 24)

This is not a matter of simple observation but requires judgement. Although thought experiments in principle could also lead to new moral judgements, they are not empirically grounded in the way that experiments in living are (cf. Anderson 1991, 24).

Experiments in living can result in learning at the individual level (as generative experimentation) as well as at more collective levels (as evolutionary experimentation). However, the learning that can take place is limited. They may result in learning about individual moral issues but are less appropriate for learning about social moral issues, such as justice. Although experiments in living may have consequences for people other than the individual doing the experiment, they are too limited for moral experiments with socially encompassing issues.

Social experiments result in moral learning during the implementation and use of a new technology. In this way, they contribute to the discovery of new ethical issues raised by a technology, and to the learning about, and the adoption of, new specifications of current moral standards. Generative social experiments can also lead to the articulation of new moral values and to moral change, as we saw in the case of health and social services above. As in the case of experiments in living, part of the learning stems from new experiences that generative social experiments bring and which may change

existing moral evaluations. Social experiments may also contribute to the development of new moral standards or frameworks that are needed to adequately deal with the moral issues brought about by a new technology.

The ethics of moral experimentation

Like any kind of experiment, moral experiments with new technology sometimes raise ethical issues. However, the kind of ethical issues that are raised by thought experiments, experiments in living, and social experiments are distinctly different.

Thought experiments do not have direct consequences in the real world. Of course, as a result of thought experiments with new technologies, certain policies may be pursued that have consequences in the real world. It is also conceivable that thought experiments have psychological consequences for their participants or may even do psychological harm to them, but they would normally not result in direct physical harm. So while thought experiments are not entirely without ethical consequences, their consequences are relatively innocent, and in general, doing thought experiments would raise limited ethical concern. The main ethical concern regards the actions taken on the basis of the thought experiment's outcome rather than carrying out the experiment.

Experiments in living are clearly less innocent as they have direct consequences in the real world. Mill went through a serious depression as a consequence of his experiment in living. However, in most cases there are primarily consequences only for the individual carrying out the experiment. Moreover, they would typically be done voluntarily. Experiments in living would then usually meet the condition of informed consent, an important consideration in the moral acceptability of experiments with human subjects (e.g., Beauchamp and Childress 2013). Informed consent basically means that an experiment is morally acceptable when the participants have voluntarily consented after being fully informed about the possible negative consequences of the experiment. Although it has been argued that informed consent is neither a necessary nor sufficient condition for responsible experimentation with human subjects (Emanuel, Wendler, and Grady 2000), experiments in living raise limited ethical concern in as far as they are individual and voluntary. One may nevertheless object that in real life, experiments in living are never completely individual since everyone is part of a web of social relations. If an individual chooses to test a certain notion of the good life, it is likely to have social consequences for at least his or her family and friends, who might not have consented to the experiment.

Social experiments obviously raise more ethical concerns than thought experiments and experiments in living. They may result in significant harm and irreversible consequences, and they may not meet the informed consent condition. Some may argue, therefore, that we should not do social experiments at all. However, the introduction of new technology into society is usually an experimental process anyway, whether we organize it as a social

experiment (which is aimed at deliberate and systematic learning) or not. It is often better to turn it into a deliberate social experiment to insure deliberate and systematic learning. The ethical issue is therefore often best phrased as: under what circumstances is it acceptable to experimentally introduce a new technology into society? In an earlier publication, I have proposed an ethical framework to judge whether and when the experimental introduction of a new technology is acceptable (Van de Poel 2016). If we decide to experimentally introduce a new technology, the next question is how to best organize such social and moral experimentation. Important issues with respect to more responsible social and moral experimentation are also discussed in several chapters in this edited volume, in particular the chapters by Vincent and Jane (Chapter 6), and by Asveld and Stemerding (Chapter 5).

Outlook

Explicit and deliberate moral experimentation with new technology is, in practice, still the exception rather than the rule. Thought experiments are sometimes carried out for technologies such as nanotechnology, synthetic biology, and robotics in an attempt to anticipate the possible ethical issues that might arise. Experiments in living are done by some individuals with technologies such as mobile phones, wearables like Google Glass, and enhancement drugs, but they are usually not conceived of as experiments, nor do they usually result in deliberate and systematic learning. Deliberate social experiments are even rarer. Most moral experimentation with new technology is tacit and *de facto*.

One plausible reason why moral experiments are often not called by that name is the fear that doing so might increase public resistance against new technologies. In fact, some protesters against biotechnology have framed the introduction of this technology as an unacceptable social experiment (BBC 2008). However, denying the experimental character of new technology may in the long run erode public trust as uncertainties abound and unexpected risks and disadvantages will emerge.

Whether moral experiments with new technology are called by name or not, the introduction of some new technologies raises ethical and social issues that need to be addressed. Recognizing the experimental character of these technologies will address these ethical and social issues better, and will help to achieve better learning processes that contribute to better technologies in a better society. What is needed is more than making tacit experimentation with new technology explicit and deliberate. Once the introduction of new technology into society is recognized and deliberately organized as an experiment, the question becomes what is the best and most responsible way to organize such experiments.

One of the important lessons from this chapter is that the three modes of moral experimentation (thought experiments, experiments in living, and social experiments) I have discussed offer distinctly different possibilities

for moral learning, and that they do so at different social costs. In particular, they offer different trade-offs between the possibilities for moral learning and the social costs they incur (and consequently the amount of ethical concern they raise). Moral experiments in society at large, or social experiments, may offer the most ample opportunities for moral learning, but they also do so against the highest social costs and therefore raise the most ethical concerns. Thought experiments have considerably lower social costs and therefore have a place in moral experimentation with new technology, even if their external validity may be low. Experiments in living that take place in the real world have a larger external validity and therefore offer better possibilities for learning; social costs are higher but not as high as in social experiments. However, they are often not appropriate to test the full spectrum of social and moral consequences of a new technology.

Proper moral experimentation with new technology then requires a combination of different kinds of experimentation (thought experiments, experiments in living, and social experiments) in different set-ups (explorative, controlled, generative, and evolutionary experimentation) at different sites (virtual worlds, the laboratory, field, and society). In particular, it would require intelligent combinations of different ways of experimenting deliberately sequenced in specific ways (e.g., going from thought experiments, to experiments in living in living labs, to generative social experimentation, for example) to increase the chances of systematic social and moral learning while minimizing harm to individuals, society, and nature.

Acknowledgements

This chapter was written as part of the research program “New Technologies as Social Experiments,” which was supported by the Netherlands Organization for Scientific Research (NWO) under grant number 277-20-003. An earlier version of this chapter was presented at the International Conference on Experimenting with New Technologies in Society, which was held August 20–22, 2015, in Delft, The Netherlands. I would like to thank the conference participants and, in particular, Donna Mehos and Lotte Asveld for comments on earlier versions.

Notes

- 1 As explained above, I do not take hypothesis testing to be a prerequisite to speak about an experiment. Nevertheless, I agree with Weber that a deliberate and systematic attempt to learn from the experiment should be present to properly speak of an experiment. Part of the disagreement may have to do with how the notion of “hypothesis” is understood. I assume that hypothesis testing requires a well-established scientific or moral theory from which a hypothesis is derived. Weber, and also Dewey, seems to employ a much more liberal notion of hypothesis in which an idea or plan that is tried out in an experiment counts as a hypothesis, even if it is not based on a well-established theory.

- 2 Foot (1967) introduced the thought experiment to show the difference in moral weight between negative and positive duties.
- 3 Nozick gives three reasons for not plugging into the experience machine: (1) we want to do things and not just have the experience of doing these things, (2) we want to be a certain kind of person (and not just to have certain experiences) and (3) plugging into the machine limits us to man-made reality, which means we cannot make contact with any deeper reality.
- 4 According to Sorensen (1992, 165), most thought experiments may be represented as stylized paradoxes that aim at showing inconsistencies. The experience machine and the ticking-time bomb, a thought experiment that has been used to argue that torture is permissible (Allhoff 2005), clearly fit this pattern.
- 5 Sorensen nevertheless believes that thought experiments are a limiting case of experiments. He defines an experiment as “a procedure for answering or raising a question about the relationship between variables by varying one (or more) of them and tracking any response by the other or others” (186). He argues that thought experiments also fit this definition.
- 6 External validity is a measure for the extent to which the results of the experiment can be generalized to other situations, in particular real-world situations.
- 7 That is to say, thought experiments at least do not result in physical harm. Still, they may result in other kinds of harm. For example, if people are asked whether they would be inclined to engage in a certain fictional but immoral scenario, this may give such a scenario nevertheless a kind of legitimacy that may not be deserved, and it may even lower the bar for immoral behavior. It is also conceivable that people are harmed if they have to choose between two immoral options in a fictional scenario (like in the trolley problem). See also the discussion in Sorensen (1992, 243–246).
- 8 I also distinguished learning about general moral theories or general moral insights, but this is less important for moral experiments *with new technology*.

References

- Alfano, Mark, and Don Loeb. 2016. Experimental Moral Philosophy. In *The Stanford Encyclopedia of Philosophy (Spring 2016 Edition)*, edited by Edward N. Zalta. <http://plato.stanford.edu/archives/spr2016/entries/experimental-moral/>
- Allhoff, Fritz. 2005. “A Defense of Torture. Separation of Cases, Ticking Time-bombs, and Moral Justification.” *International Journal of Applied Philosophy* 19 (2):243–64.
- Anderson, Elizabeth S. 1991. “John Stuart Mill and Experiments in Living.” *Ethics* 102 (1):4–26. doi: 10.2307/2381719.
- Appiah, Anthony. 2008. *Experiments in Ethics, The Mary Flexner Lectures*. Cambridge, MA: Harvard University Press.
- Bauman, Christopher W., A. Peter McGraw, Daniel M. Bartels, and Caleb Warren. 2014. “Revisiting External Validity: Concerns about Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology.” *Social and Personality Psychology Compass* 8 (9):536–54.
- BBC. 2008. “Charles in GM ‘Disaster’ Warning.” BBC. Accessed December 12, 2014. http://news.bbc.co.uk/2/hi/uk_news/7557644.stm.
- Beauchamp, Tom L., and James F. Childress. 2013. *Principles of Biomedical Ethics*. 7th ed. New York: Oxford University Press.
- Boenink, Marianne, Tsjalling Swierstra, and Dirk Stemerding. 2010. “Anticipating the Interaction between Technology and Morality: A Scenario Study of Experimenting with Humans in Bionanotechnology.” *Studies in Ethics, Law, and Technology* 4 (2):1–38.

- Davis, Janet, and Lisa P. Nathan. 2015. "Value Sensitive Design: Applications, Adaptations and Critiques." In *Handbook of Ethics and Values in Technological Design*, edited by Jeroen van den Hoven, Pieter E. Vermaas, and Ibo Van de Poel, 11–40. Dordrecht: Springer.
- Davis, Michael. 2005. "The Moral Justifiability of Torture and Other Cruel, Inhuman, or Degrading Treatment." *International Journal of Applied Philosophy* 19 (2):161–78.
- Davis, Michael. 2012. "Imaginary Cases in Ethics." *International Journal of Applied Philosophy* 26 (1):1–17.
- Dewey, John. 1922. *Human Nature and Conduct; an Introduction to Social Psychology*. New York: Holt.
- Dewey, John. 1927. *The Public and its Problems*. New York: Holt.
- Dewey, John. 1929. *The Quest for Certainty*. New York: Minton, Balch & company.
- Dewey, John. 1938. *Logic, the Theory of Inquiry*. New York: Holt.
- Dewey, John, and James Hayden Tufts. 1908. *Ethics, American Science Series*. New York: Holt.
- Emanuel, Ezekiel J., David Wendler, and Christine Grady. 2000. "What Makes Clinical Research Ethical?" *The Journal of the American Medical Association* 283 (20):2701–711.
- Foot, Philippa. 1967. "The Problem of Abortion and the Doctrine of Double Effect." *Oxford Review* 5:5–15.
- Honan, Mat. 2013. "I, Glasshole: My Year with Google Glass." *Wired*, December 12, 2013. <https://www.wired.com/2013/12/glasshole/>.
- Horowitz, Michael, and Paul Scharre. 2015. *Meaningful Human Control in Weapon Systems: A Primer*. Center for a New American Security.
- Kudina, Olya, and Peter-Paul Verbeek. submitted. "Ethics from Within: Google Glass and the Mediated Value of Privacy." Submitted to *Science, Technology, & Human Values*.
- Lessig, Lawrence. 1999. *Code and Other Laws of Cyberspace*. New York: Basic Books.
- Lucivero, Federica. 2016. *Ethical Assessments of Emerging Technologies Appraising the Moral Plausibility of Technological Visions*. Vol. 15, *The International Library of Ethics, Law and Technology*. Dordrecht: Springer.
- Mill, John Stuart. 1869. *On Liberty*. 4th ed., *Library of Economics and Liberty*.
- Mill, John Stuart. 1874. *Essays on Some Unsettled Questions of Political Economy, Library of Economics and Liberty*.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic books.
- Popper, Karl R. 1945. *The Open Society and its Enemies*. 2 vols. London: Routledge.
- Sclove, Richard E. 1995. *Democracy and Technology*. New York: The Guilford Press.
- Shue, Henry. 1978. "Torture." *Philosophy & Public Affairs* 7 (2):124–43.
- Sorensen, Roy A. 1992. *Thought Experiments*. New York: Oxford University Press.
- Swierstra, Tsjalling, Dirk Stemmerding, and Marianne Boenink. 2009. "Exploring Techno-Moral Change. The Case of the Obesity Pill." In *Evaluating New Technologies*, edited by P. Sollicie and M. Düwell, 119–38. Dordrecht: Springer.
- UNIDIR. 2014. *The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward*. UNIDIR (United Nations Institute for Disarmament Research).

- Van de Poel, Ibo. 2016. "An Ethical Framework for Evaluating Experimental Technology." *Science and Engineering Ethics* 22 (3):667–686.
- Van de Poel, Ibo. 2017. "Society as a Laboratory to Experiment with New Technologies." In *Embedding New Technologies into Society: A Regulatory, Ethical and Societal Perspective*, edited by Diana M. Bowman, Elen Stokes, and Arie Rip, 61–87. Singapore: Pan Stanford Publishing.
- Weber, Eric Thomas. 2011. "What Experimentalism Means in Ethics." *Journal of Speculative Philosophy* 25 (1):98–115.
- Winner, Langdon. 1980. "Do Artifacts Have Politics?" *Daedalus* 109:121–36.

4 The theatrical debate

Experimenting with technologies on stage

Frank Kupper

Introduction

The theatrical debate is an interactive theater format that brings citizens together for a joint exploration of their ideas and concerns. Staging the societal future of emerging technologies in vivid, open-ended situations that invite multiple interpretations, the theatrical debate offers a reflexive methodology to engage in experimental and creative ethical reflection. Such an experimental approach to ethical reflection is needed to grasp the dynamic complexity of the introduction of emerging technologies in society. The experimental perspective elaborated on in this volume conceives the introduction of technology in society as a social experiment. As Ibo van de Poel (Chapter 3) argues, such an experiment involves learning about the social impacts of technology, about the embedding of technology in society, and how to evaluate different impacts normatively. The theatrical debate methodology may be conceived of as a simulation experiment that enables its participants to explore these questions in a virtual learning environment even before the introduction of technology in society. In a collective process of inquiry, the participants of the theatrical debate anticipate the future social embedding of emerging technologies, reflect on underlying values, needs, and expectations, and discern desirable and acceptable directions for change. As I will argue at the end of this chapter, what is learned from the theatrical debate as a simulation experiment may help the management of the experimental introduction of technologies in the future.

We developed the theatrical debate production *Nano is Big* specifically for *Nanopodium*, the Dutch national dialogue on the societal and ethical aspects of nanotechnology in 2010–2011, and later modified it to fit debates about other emerging technologies such as functional foods, neurotechnologies, and synthetic biology. In this chapter, I will reflect on the *Nano is Big* experience to develop an understanding of the theatrical debate methodology as part of an experimental approach to ethics. As I will argue, the pragmatism of the American philosopher John Dewey provides a helpful framework to appreciate the contribution that the theatrical debate methodology can make. Dewey strongly emphasized the importance of creative-intelligent

inquiry and democratic deliberation to produce new understandings and evaluations in a changing world. A key element of Dewey's approach to inquiry and deliberation is the notion of "dramatic rehearsal": the exploration of an indeterminate situation through the imagination of alternative courses of action and their consequences. Such dramatic rehearsals may be performed through thought experiments and narratives. As I will show in this chapter, the theatrical debate employs the idea of dramatic rehearsal in an almost literal sense: the use of interactive theater as a dramatic rehearsal of how the introduction of emerging technologies may unfold.

Emerging technologies and ethical assessment

"Emerging technologies" is a catchphrase used frequently throughout the literature to depict the rise of a set of technologies that is currently being developed—or will develop in the near future—and is expected to alter our lives and environments in substantial ways. Examples of emerging technologies are nanotechnologies, biotechnologies, and information technologies. The possible convergence of these technologies towards integrated applications is considered even more transformative. In contrast to incremental innovation processes, emerging technologies are much more unpredictable, disruptive, and complex. Nanotechnology is one of the iconic examples of emerging technologies. Nanotechnology involves the manufacture, operation, and control of materials at the level of atoms and molecules, resulting in either the development of new materials, enhancement of existing properties of materials, or the application of new properties to existing materials. As is common for emerging technologies, nanotechnology is praised for its "revolutionary" potential to generate numerous benefits for domains as diverse as product development, environmental conservation, medicine, and information technology (Roco and Bainbridge 2001). At the same time, it gives rise to numerous concerns about potential health risks and environmental hazards. In addition, nanotechnology raises wider social and ethical issues regarding unintended long-term consequences, social and financial risks, issues of governance and control, and fundamental issues about life and human identity (Pidgeon and Rogers-Hayden 2007).

The question is how to deal with these impacts ethically. What is required to grasp the dynamic context of emerging technologies? When and where should ethical reflection take place? Who should be involved? Since the 1970s, practices of technology assessment (TA) emerged, especially in the United States and Europe, predominantly positioned as an early warning system to prevent negative impacts of technologies on society. Since the 1990s, TA practices made a participatory-deliberative turn, involving a broad spectrum of actors such as scientists, technology developers, policy-makers, representatives of societal organizations, and laypeople in a proliferation of participation experiments (Palm and Hansson 2006). This change coincided with a more general call for public influence and

participation in decision-making (Irwin 2001; Wilsdon, Wynne, and Stilgoe 2005). Attention focused on public engagement at early stages of the research and innovation process, when normative choices about their trajectories were still considered open to debate and public views less solidified (Wilsdon and Willis 2004; Felt et al. 2014). Especially nanotechnology has been welcomed by different authors as an excellent opportunity to move engagement “upstream” (see for example Macnaghten, Kearnes, and Wynne 2005), generating the employment of a broad variety of public engagement methods (see for an overview Delgado, Kjølberg, and Wickson 2010). The participatory-deliberative turn has increasingly made the ethical assessment of emerging technologies a matter of public deliberation. This certainly opened up science and technology for public scrutiny. However, it has not been easy to engage citizens and stakeholders meaningfully in technology assessment. An important question that continues to resurface is how to create platforms and spaces for public deliberation and shape the moral conversation about emerging technologies in public.

Three interrelated issues complicate this question. First, as a basic observation, sociologists and philosophers have convincingly shown that technology and society are deeply intertwined and coevolve while mutually shaping each other (see for example Rip and Kemp 1998; Grin and Grunwald 2000). Technologies shape societies, including the ways in which we perceive, understand, and evaluate the world around us (Verbeek 2009). At the same time, technologies are themselves shaped socially by the actions, decisions, and negotiations of the actors involved (Bijker, Hughes, and Pinch 1987; Jasanoff 2003). The coevolution of technology and society results in an open-ended process subject to contingency and change. Second, there is the related challenge of moral change. According to Swierstra, Stemerding, and Boenink (2009), this aspect of the dynamic coevolution of technology and society has not yet received much attention. Swierstra and Rip (2007) show that moral routines are usually taken as self-evident truths. We only become aware of them when conflicts between these routines and new possibilities and dilemmas arise. These “morals,” in the words of Swierstra and Rip, then become a matter of ethics again, i.e., they have to be reopened for critical inquiry and deliberation. Next to a rethinking of moral routines, techno-moral change might also affect the nature of ethical questions that are asked. Instead of the predominant focus on negative impacts on health, safety, and environment, emerging technologies might turn the attention of ethics towards the broader question of whether these technologies contribute to the flourishing of human and nonhuman nature (see Swierstra 2013).

Third, the appraisal of emerging technologies in the public sphere has extended the range of actors involved, and consequently the range of issues and concerns that could, or should, be addressed. As Wynne (2001) explained, differences in values and beliefs among various social actors produce different frames of the social controversy surrounding a particular technological development. Frames are particular ways of making sense of a

complex reality and guiding our actions. They consist of structures of values and beliefs about a certain situation or object that provide the constructs and patterns that make the situation or object meaningful to ourselves and others (Schön and Rein 1994). Although in human communication, frames are continuously produced and reproduced “unfinished constructions,” they are simultaneously embedded in underlying values, assumptions, and worldviews (Kupper and de Cock Buning 2011). The pluralism of democratic societies challenges both the practice of ethical technology assessment as well as the practice of political decision-making. Because of their different ways of framing, social actors differ in the aspects they feel should be publicly discussed. The issue of pluralism not only relates to the question of whether or not a particular ethical concern is legitimate, but also extends to the questions how should ethical decisions be made and what is an appropriate level and role of public engagement. Moreover, insensitivity to pluralism seems to misrepresent our experience of the diversity and richness of our relationships with the world (Smith 2003).

Dewey’s pragmatism: experimental and pluralistic

From the 1980s onwards, there has been a revival of philosophical pragmatism in different fields such as political theory, philosophy, and ethics (Hickman 1998; Caspary 2000). In different fields such as animal ethics, environmental ethics, and technology ethics, the pragmatist alternative has been welcomed. Especially Dewey’s work has inspired many scholars reflecting on the dynamic relationship between humans, technology, and society (see Keulartz et al. 2004; Shelley-Egan 2011; Krabbenborg 2013a). Also, the theatrical debate methodology articulated in this chapter roughly follows the reconstruction of ethics proposed by Dewey as it is interpreted by some of his contemporary followers in the field of technology ethics. But why pragmatism? A first merit of pragmatism is the recognition of the coevolution between science, technology, and society. As Hickman (2001) noted, one of the important insights of pragmatism is that our culture is basically a technological culture, characterized by the entanglement of ethical, social, and technological aspects. Also in Dewey’s time, the late-nineteenth and early-twentieth century, societies were facing the development of fairly disruptive technologies that changed life tremendously. The relationships between humanity and technology constituted an important aspect of Dewey’s philosophy. He considered the technologies with and by which we live to be an integral part of our “existential matrix of inquiry,” which he understood as the locus of our interactions—biological and cultural—with the environment (McGee 2003). Like language, technologies shape the way we look at others and ourselves, and provide the context of meaning we inhabit.

The introduction of emerging technologies such as nanotechnology in society is characterized by complexity and changing circumstances, and thus requires an ethical approach that is flexible and context-sensitive. This is

an important strength of Dewey's pragmatism. It is flexible. It can adapt to changing circumstances and practices because it is not tied to absolute truths that are divorced from people's everyday lived experience (McKenna and Light 2004). Dewey's philosophical pragmatism is known for its anti-foundational character (Dewey 1920; Keulartz et al. 2004). In general, Dewey's philosophy explores how humans can live a meaningful life in an everyday world (Krabbenborg 2013a). According to Dewey, this world is characterized by continuity, contingency, and change. The world we share is not a given. This implies that all our convictions inescapably have a provisional nature and should remain susceptible to critical appraisal. When we encounter new and unexpected situations, reflective inquiry is needed to develop an adequate response. Existing interpretive frameworks may not be suitable to grasp such a new reality. Dewey uses the notion of "indeterminacy" to signify these situations in which is not clear what happens and what is at stake. It is in such situations that people need to come together and collectively reflect on their experience to articulate the situation and construct new or adapted interpretive frameworks. Different authors pointed at the similarity of the context of emerging technologies in society. Krabbenborg (2013a), for example, argued for spaces of assembly and interaction that would allow for such a joint process of reflective inquiry in the context of nanotechnology.

Another strength of Dewey's pragmatism is that it is able to celebrate value pluralism because it accepts the historical context of moral experience, and it embraces an open, experimental approach to ethical claims (Minteer, Corley and Manning 2004). In Dewey's conception, human beings are situated in specific contexts and histories and relate to the world in different, unique ways depending on our experience. The world we share is therefore different for different individuals, groups, and societies across the globe and across time. We never experience the world fully, and we never see it, or think about it, the same. This implies that we need each other's perspectives to articulate the complexity of an indeterminate situation.

Dewey's reconstruction of ethics

Dewey's approach to ethics offers a conceptual framework to seek the development of moral guidelines for technology development without disregarding its dynamic character and the pluralism of our societies. Recognizing the dynamic and indeterminate character of human life, Dewey criticized the ethical theories and institutions of his time because of their quest for certainty and absolute moral truths. Rational foundationalism, he argued, would only obstruct creative-intelligent inquiry and the development of solutions to new and unexpected situations that arise. Dewey aimed at the reconstruction of ethics into a more open, experimental, and creative method for developing adequate responses in an ever-changing world (see for example Dewey 1932). A first characteristic of Dewey's approach is a *shift to contextualism*. Due to the complexity of indeterminate situations, any meaningful inquiry should occur

within the unique context of a situation. For example, different, conflicting values may be at stake that are difficult to reconcile. Individuals confronted with such a situation often find themselves in a persistent struggle with the complexity of moral judgment. Dewey criticized traditional approaches to the dilemmas of moral experience for not recognizing this complexity, as well as the novel demands and circumstances of indeterminate situations (Minteer, Corley, and Manning 2004). Dewey felt that the preoccupation with irrefutable theories and universal moral standards had distracted moral philosophers from moral experience as it was actually felt and lived. He was convinced it is always a “felt” question or confusion that makes us aware of a potentially morally problematic aspect of an indeterminate situation (Pappas 1998).

The shift to contextualism implies a strong focus on the process of moral inquiry. Dewey developed a model of reflective inquiry aimed at cultivating our human capacity for the reconstruction of habits and routines in the resolution of our problems. Following Krabbenborg (2013a) and Steen (2013), I distinguish three distinctive phases. In the first phase, an indeterminate situation is transformed into a problematic situation. The inquiry process starts with personal subjective experiences that question the situation at hand. Feeling is considered an important source of experience throughout each phase. Expressing (and sharing) feelings and concerns is critical to explore and define the problem. Engaged in a process of inquiry, one asks questions such as, “What do I find problematic about this situation?” and “What are my (and other people’s) experiences?” In the second phase of inquiry, the problem is perceived from multiple perspectives and possible solutions are conceived. This phase combines perception (one’s capacity to obtain an understanding of the problematic situation) and conception (the capacity to envision alternative scenarios) and typically involves dramatic rehearsal—the creative imagination of alternative courses of action using both thoughts and feelings. The third phase of inquiry focuses on trying out and evaluating solutions. This phase reevaluates the relationships between problem definitions, proposed solutions, and testing the solutions in the contexts of real world problems.

Finally, the way moral experience is “felt” and “lived” is always constructed in relation to other members of a moral community. The undeniably social context of morality, therefore, not only requires individual reflective inquiry but also *public deliberation*. Dewey’s philosophical project demonstrates a strong faith in the ability of human experience to produce from within the justification of values and beliefs. Moral deliberation in this view ultimately rests on the potential of individuals to engage collectively in moral inquiry and deliberation.

Inquiry, imagination, and dramatic rehearsal

Although Dewey appreciated the scientific method of empirical inquiry as an example for other forms of inquiry, his model of reflective inquiry is

intuitive and creative (Caspary 1991). Throughout the process of inquiry, imagination plays a crucial role. Fesmire (2003) distinguished two roles of imagination in Dewey's conception of moral inquiry and deliberation: (1) imagination enables us to respond directly and empathetically to other people's feelings and thoughts, not by projecting our own values and intentions, but by really taking the perspective of others; (2) imagination is a way to escape the inertia of habitual patterns of thought and to start seeing alternatives, creatively exploring a situation's possibilities. Imaginative thinking for Dewey is to perceive what is before us in light of what could be. Central to Dewey's account of moral imagination is the abovementioned notion of dramatic rehearsal. Deliberation implies trying out various alternative courses of action in imagination to foresee their outcomes (Dewey 1922). Caspary (1991) has thoroughly elucidated the role of dramatic rehearsal in Dewey's theory of deliberation. Here, I will build on Caspary's interpretations to discuss the characteristics of dramatic rehearsal that are most pertinent to the theatrical debate methodology I develop here. First, the use of the drama metaphor signifies that deliberation always concerns actions that potentially affect other morally significant beings. According to Dewey, *deliberation is dramatic* in the sense of unfolding a narrative or plot and reflecting on potentially emerging conflicts between the parties involved. Through reflective inquiry, we seek answers to questions such as: Who are the actors involved? What are their interests and values? What are possible actions and constraints? This touches upon a second prominent characteristic; dramatic rehearsal is always about *anticipating the responses of other people and oneself*. The content of the imagined lines of action revolves around responses of the actors involved in the morally indeterminate situation. What are people's tendencies? How would they perceive and assess this situation? It also leads to the discovery of one's own tendencies: what is it that I think of this myself? In Dewey's theory of deliberation, *our ability to anticipate and evaluate the responses of others depends on our emotional sensitivity*. As we spin out alternative scenarios in our imagination, we use our emotional reactions, desires, and aversions to judge their acceptability. It is therefore important that the scenarios are vividly imagined, enabling us to grasp the indeterminate situation empathetically and to understand what it would be like to live one way rather than another. Of course, reconceptualization is very central to Dewey's notion of dramatic rehearsal. By imagining different courses of action, we rethink the concepts we use to identify, order, and interpret situations. Such exploration of concepts, Caspary argues (1991, 184), may be the daily work of professional philosophers, but "turns out to emerge in the thinking of anyone engaged in serious ethical deliberation." The final prominent characteristic I mention here is the indeterminate and open-ended character of deliberation. We simply do not know what the next event will be in the narrative plots that we unfold, and what new concepts will help reorganize our understanding. For our deliberations to be productive, we have to open up to the unexpected.

A pragmatist ethics for a technological culture

Adopting a Deweyan perspective, Keulartz et al. (2004) have sketched the contours of a pragmatist ethics for a technological culture. Following Dewey, they maintained that the strongly dynamic and uncertain character of emerging technologies endlessly presents us with new moral problems giving rise to social conflict. They emphasized the need for creative management of those conflicts. In their manifest, they advocated two progressive problem shifts in the tasks of ethics. The first is *a shift from product to process*. Pragmatist ethics should be focused on the process of experimental inquiry and deliberation in order to consider the entire range of values and claims relevant to the resolution of a problematic situation and social conflict. The activities of pragmatist ethics shift to the refinement of the process of inquiry and the development of effective methods of cooperative problem solving (see also Caspary 2000). Furthermore, ethical inquiry is framed as a more creative and dynamic process in which discovery and invention are important characteristics of moral deliberation (Minteer, Corley and Manning 2004). This constitutes the second problem shift Keulartz et al. (2004) have in mind: *a shift from the context of justification to the context of discovery*. In the context of a strongly dynamic and uncertain coevolution between technology and society, new constructs and hypotheses are needed in order to understand and evaluate new moral problems. Keulartz et al. refer to discovery in pragmatist ethics as “the creative capacity for the innovation and invention of vocabularies which provide new meanings and open new perspectives.” These problem shifts result in a new task package for pragmatist ethics (see Table 4.1) Each combination of a specific operational context and focus requires a different task of ethics and a different role for ethicists. The theatrical debate methodology can be understood as a combination of task (c) and (d). Task (c) follows Dewey’s notion of dramatic rehearsal outlined above and involves the imaginative evaluation of the possible consequences of different courses of action. The role of the ethicist here is to provide a critique of existing patterns, customs, and vocabularies and to create new

Table 4.1 The tasks for pragmatist ethics

	<i>Product</i>	<i>Process</i>
Context of justification	Traditional ethics (a) Providing arguments and justifications for courses of action	Discourse ethics (b) Structuring fair public deliberation and decision-making
Context of discovery	Dramatic rehearsal (c) Exploring possible future worlds; criticizing/ renewing vocabularies	Conflict management (d) Aiding open confrontation of moral vocabularies and worldviews

Source: Adapted from Keulartz et al. (2004).

concepts and constructs in order to understand future arrangements. Task (d) relates to the deep-seated fundamental value conflicts that may arise in the development of new technologies. The ethicist plays a role in facilitating the open confrontation of worldviews and the achievement of reflexive awareness.

The use of theater

The metaphor of dramatic rehearsal highlights the imaginative exploration of situations in which someone's actions may affect other morally significant beings, resulting in an imaginative remaking of the future. This resembles the experience theater can provide for its audiences. Science theater is appreciated as a setting that transforms complex and abstract matters into dramatic stories (Odegaard 2003). The enactment of a dramatic narrative makes it possible for the audience to witness the issue at stake, identify with the characters, and learn about scientific practice and its impact on society (Winston 1999). Fesmire (2003), however, points out that the image of the conventional stage drama may evoke misleading associations. He argues moral imagination in Dewey's conception is not a dress rehearsal for a ready-made play. Dewey's moral stage is much more of a continuum with our experience and actions in the real world. Imaginative thinking forms an integral part of the process of moral inquiry and deliberation. As Fesmire (2003, 80) puts it, "scenes are *co-authored* with others and with a precarious environment. The acting is improvisational, the performances open-ended. The drama is experimental, not scripted." This implies that we also have to look for participatory, improvisational approaches to theater if we want to put dramatic rehearsal into practice and create a public space for inquiry and deliberation to shape the moral conversation about emerging technologies. Much of the participatory theater that is performed draws on the work of the Brazilian theater practitioner Augusto Boal (1979, 2002). Inspired by Paulo Freire's *Pedagogy of the Oppressed* (Freire 2000), Boal developed the Theater of the Oppressed, a system of theatrical techniques to empower people to become subjects of their lives and acquire the capacity to realize their needs. Although the theatrical techniques were originally developed in a time of extreme political repression in Brazil, the use of these techniques has now spread across the world and has been adapted to many different contexts by diverse practitioners, such as educators, political activists, therapists, and social workers (Schutzman and Cohen-Cruz 1994). One of the main participatory theater techniques developed by Boal is forum theater. The aim of forum theater is to stimulate the exploration and discussion of problematic issues experienced by the participants in their real-life contexts. Forum theater entails the enactment of a drama around the problems of a protagonist, leading to a crisis; the participants are invited to help resolve the crisis by identifying the key moments of the unfolded scenario and stepping

into the shoes of the protagonist themselves to experiment with alternative courses of action to improve the situation. Boal (1998, 8) argued that forum theater creates the space for citizens to “transgress, to break conventions, to enter into the mirror of theatrical fiction, to rehearse forms of struggle and then return to reality with the images of their desires.” Boal’s figure of the Joker, the neutral facilitator of the forum theater performance, plays a crucial role in shaping the interactions that take place, probing the participants to explain their responses and underlying motives, but especially to alter the situation that is played out on stage, enabling them to investigate their stories and discover deeper layers in their complex understanding of the indeterminate situation (see Dwyer 2004; Diamond and Capra 2007; Perry 2012). Forum theater is a reflexive praxis, encouraging its participants to explore their social conditions critically and re-conceptualize their ideas and routines. Essentially, Boal’s oeuvre is about activating the passive spectator to become, as he called it, “spect-actors:” active performers in the rehearsal for personal and social change (Schutzman and Cohen-Cruz 1994). Enter the game of dialogue: forum theater provides its participants with the dialogical space to shape their own feelings, understandings, and creative solutions to the challenges they encounter in their lives (Boal 2002).

Nano is Big: The theatrical debate about nanotechnology

We took the participatory approach of Boal’s forum theater as the starting point to develop the theatrical debate methodology. We developed the methodology specifically for the production *Nano is Big* in the context of *Nanopodium*. The goal of the *Nano is Big* production was threefold: (1) raising awareness of the ethical and societal aspects of nanotechnology; (2) stimulating reflection and opinion-forming about these aspects; (3) stimulating dialogue and mutual understanding of the different meanings and moral implications of nanotechnology in the future. In this section, I will explain what constitutes the theatrical debate concept and how the methodology works by reviewing the *Nano is Big* experience.

The play was developed as a coproduction of the author (facilitator) and eight professional actors. The process started with preliminary literature research to outline the possible contents of the play. We included scholarly literature on ethical and societal aspects as well as the technomoral scenarios and vignettes that were developed within *Nanopodium* (see Krabbenborg 2013a for an overview of projects). We purposely moved beyond the traditional discourse of promises, risks, and benefits to focus explicitly on broader ethical and societal issues (see Hanssen 2009; Swierstra and Te Molder 2012 for an argumentation). On the basis of their suitability for theater, the priorities of *Nanopodium*, and the relationship to other *Nanopodium* projects, the following themes were selected: dealing with uncertainty; risks, and benefits; effects on social roles, identities, and relationships (such as trust or inequality); societal values (such as privacy); and shifting boundaries.

The scenarios, vignettes, and extracted ethical issues formed the starting point to search for drama and develop the scene formats. In a series of rehearsal sessions, the selected themes were transformed into personalized acts that involved characters, a story line, and a dramatic conflict that expressed the selected opportunities, dilemmas, and concerns. Eventually, the *Nano is Big* theater play consisted of five scene formats that each addressed the societal embedding of a specific nanotechnology application. The five scene formats varied from open to structured formats. Together they assembled a 90-minute interactive theater play. Typically, a module started with a short neutral introduction of the specific nanotechnology application by the facilitator (the author), followed by a semi-scripted scene played out by the actors that explored the societal context of application. The iterative cycle of dramatizations and discussions followed Dewey's model of reflective inquiry. After the status quo of the first dramatization had been demonstrated, the facilitator assisted the audience to explore the possibilities and dilemmas they experienced witnessing the scenes. This sequence of dramatization and discussion transformed an indeterminate situation into a problematic situation (phase one). The facilitator probed the audience members to reflect on their stories, encouraging the discovery of deeper layers of values and assumptions. Participants were invited to perceive the problematic situation from multiple perspectives and conceive possible solutions (phase two). The discussion would result in a variety of ideas and concerns that at some point were translated in directions for new dramatizations. The actors continued the play, trying out alternative scenarios to demonstrate their consequences. The dramatic rehearsal of alternative scenarios altered the status quo of the unfolding story. The audience was again asked for their response to evaluate the scenario that was played out (phase three). The resulting discussion could again lead to new dramatizations, reentering the iterative cycle. In the spirit of forum theater, the focus of the performance was on acting rather than talking. This enabled the participants to explore their experience of the staged scenarios in a more intuitive way, demonstrating aspects that are usually lost out of sight in public debates. Table 4.2 shows what the dynamic exploration of future scenarios by the interaction of play, reflection, and discussion was like.

From April 2010 to March 2011, the theater play *Nano is Big* was performed for different audiences at 23 locations across the Netherlands. Ten performances were located at secondary schools, attended each by approximately 60 students (highest educational level; aged 15–17) with a total audience of 625. The remaining 13 performances were targeted at interested citizens. Attendance rates ranged from eight to 60 attendees with a total audience of 508. Interestingly, the theatrical debate produced a similar variety of ideas and concerns for small and large audiences.

For every performance, we collected data using a brief exit survey. Also, the performances themselves were videotaped and analyzed using a thematic qualitative analysis approach (Braun and Clarke 2006). The surveys

Table 4.2 The “Doctor inside!” scenario

Synopsis

The possible future situation that is explored in this scenario is the application of medical nanobots that patrol the human body, taking measurements of biological parameters inside the body resulting in a diagnosis. The collected information may be transferred to healthcare providers via smart devices in the environment. Optionally, the nanobots may trigger the release of medication inside the body. A potential advantage of this scenario is that people with chronic medical conditions can be monitored and treated more efficiently and effectively. There are also health risks involved, related to the effect of the nanobots on the body and the reliability of this monitoring and treatment application. Wider issues for example relate to the shift in roles and responsibilities in the provision of healthcare, boundary shifts with respect to health vs. disease, man vs. machine, the cultural meaning of suffering, the new normal of this kind of treatment.

Scene format

The chosen scene format was the split-focus scene. In this scene, the stage is split in two. We see two tableaux. In one we see a physician and a patient, representing the healthcare situation now. In the other we see a patient in the future interacting with her nanobot monitoring and treatment device. The audience is asked for a type of vague complaints that both patients suffer from to provide the players with input. The play starts with a dialogue about the vague complaints between patient and physician in the now, if needed supported by physical examination of the patient. The patient in the future receives feedback from the “Doctor inside!” system. The patient is reading the information from a virtual screen. Both scenes develop simultaneously, alternating focus. After the status quo has been established, the facilitator breaks in and solicits the first responses of the participants. The facilitator purposely starts collecting experiences, gradually moving to understandings and evaluations. What do the participants experience? Do they see differences and similarities? How do they perceive the care relationship? Who has responsibility? Are the characters capable and/ or in the position to exercise their responsibility? What could be done to improve the situation? The resulting dilemmas and alternative strategies are transformed into new directions for the players. The scenes may both continue, only one of the scenes may continue or new scenes may be added, depending on the responses of the participants. For example, a scene would focus on the notion of patient autonomy, and the question would become how to realize patient autonomy in both situations. The participants would come up with suggestions that were then tried out by the players enacting the patients. They would meet resistance from the physician (now) and the monitoring and treatment device (future). The facilitator would again break in and ask whether patient autonomy was realized? The participants would come up with new observations and suggestions to continue the exploration.

were analyzed using a simple frequency analysis. To study further how the students experienced the performance, we conducted four focus groups with an average of six students. The focus groups were audiotaped, transcribed, and analyzed using a thematic qualitative analysis approach (Braun and Clarke 2006). Although I do not aim to present an overview of the evaluation

results here, I will use some observations made in this analysis to illustrate my points.

Productive ambiguity and the process of discovery

The open and playful character of the theatrical debate made the performances very accessible for participants. Many participants indicated that they appreciated this form of engagement. The atmosphere was evaluated as open and respectful. The methodology involved participants in the play; for example, many participants indicated that the safe and free environment encouraged them to make their own contribution to the unfolding story. As one participant said, "A surprising amount of people participated, even the ones I didn't expect to." Participants indicated that the neutral and non-judgmental way by which the actors and facilitator experimented with their ideas and concerns encouraged them to reveal themselves and engage in the process of inquiry and deliberation that was taking place. This finding is indicative of the pragmatist shift from the context of justification to the context of discovery. Rather than a competition for the "right" argument, the theatrical debate performances established a process of discovery of the valuable contribution that each participant had to make.

The participants appreciated the vivid dramatizations by the actors because they helped them to visualize how nanotechnology could potentially affect their lives. As a participant formulated it, "You see the examples of what might happen. It made me really start thinking about it." The participants indicated the dramatizations enabled them to see the future embedding of nanotechnology before their eyes and talk about it. It made participants think: "It was presented in such a way that you automatically think along and form your own opinion." By simulating different courses of action on stage, dramatization refined the imagination of the participants in the projection of scenarios and consequences.

The participants evaluated the facilitation of the event as neutral and experienced structuring only with respect to the process and not the content. However, there is also a critical remark to be made. Even though the purpose of the theatrical debate methodology was to accomplish a reflective inquiry process of the widest range of ideas, values, and concerns, participants were sometimes looking for answers rather than questions. One participant commented, "This way you get a better image of what can be done with nanotechnology." Another one stated, "It was very enlightening. It is now clear what the consequences could be." This finding could actually be interpreted as anti-pragmatist, because it could mean that the dramatic exercise, and therefore the process of reflective inquiry, did not start from a question or problem that was felt by the participants but from a response to a problem created by the designers of the play. We anticipated this influence of the designer/ethicist by taking care that the portrayed situations exhibited a productive ambiguity; that is, they contained different possibilities

and dilemmas that could evoke a multiplicity of ideas and concerns, potentially triggering a new understanding of the situation. Indeed, the storylines that developed during the performances appeared to depend considerably on the composition of the audience. Different aspects of the indeterminate situation were perceived as morally problematic. Typically, the dramatizations would pursue the aspects that were perceived as problematic by the participants. Of course, we also encountered differences in perception between different participants of the same performance. In that case, different perceptions were contrasted and their implications explored by new dramatizations. As I will argue below, the collective process of interpretation helped the participants to develop a richer interpretation of the situation that was portrayed. Also, the facilitator and actors cultivated the indeterminacy of the situation by stimulating an atmosphere of open exploration. As one participant put it: "It was O.K. to doubt, and to change your opinion." Most participants considered the format open and flexible with regard to the perceptions of the audience.

Dynamic exploration

The dynamic exploration of indeterminate situations that we have witnessed during the theatrical debate performances are a compelling demonstration of how the notion of dramatic rehearsal can be put into practice in the context of public engagement and the ethical assessment of emerging technologies. This becomes evident in the exploration of the *Doctor Inside!* scenario. The *Doctor Inside!* scenario involves the application of nanobots that are capable of monitoring the course of a specific disease. As I explained above, the scenario was played out in a split-focus format, in which the stage was split into two separate scenes, running in parallel, comparing the future situation with the current situation. The explorative nature of the play is illustrated by the following participant quote: "At first, you thought maybe this technology is better... and then the actor portrayed a dilemma and you would start to doubt if the other situation would be more advantageous." Here is how the unfolding of the scenario would typically go: Upon watching the first sketch of the indeterminate situation, many participants indicated that they expected the doctor-patient relationship to change significantly. At first sight, a part of the participants appreciated the application of these monitoring devices because they believed this could improve both the efficiency and quality of the diagnosis. These participants supposed that nanobots would be faster and more adequate and precise in measuring what is going on, compared to the present-day consult of a physician. However, when the actors continued to play along these lines, now emphasizing the mentioned benefits, other participants started to feel uncomfortable, for example, expressing concerns about the reliability of these monitoring devices. How could they ever trust such a technical system? Wouldn't it be better to trust the judgment of a human practitioner instead of a machine? The actors

then, for example, played out the consequences of a fast device that was not trustworthy, compared to a very trustworthy but slow and ineffective consult with a physician. Such a scene led to a discussion of the issue of trust in this situation. The lack of trust often appeared to be related to different ambiguous features. First, even though many participants believed that the monitoring devices could be adequate and accurate, the technical devices were also thought to oversimplify a complex reality by reducing a disease to a restricted set of biomarkers. They expected this oversimplification to have negative impacts on their health and wellbeing. A human practitioner, the participants argued, would be better equipped to consider the whole situation of a patient. Also, many participants felt that the lack of emotional and social responses would make it difficult to trust the effectiveness of the monitoring devices. At the same time, the monitoring devices were seen as a way to prevent inevitable human failure. Other participants responded that it would be impossible to build a machine that would not make mistakes. Such a deliberation yielded several new questions to explore through a new round of improvisation. How can the advantages of the monitoring devices be utilized in a way that respects the complexity of disease? How can emotional and social aspects be covered in the embedding of such a nanotechnology device? The audience then started to design the future, coming up with alternative strategies for the application of this technology that would be able to satisfy the different values that at first seemed incompatible.

Moral pluralism and collective interpretation

It was explicitly not our intention to reach consensus during the performances. The aim of the theatrical debate was to explore the diversity of ideas for the purpose of making sense of the future of nanotechnology in society and raising reflexive awareness of its associated moral dilemmas. Indeed, the series of 23 performances yielded a rich variety of opportunities, dilemmas, and concerns expressed by our participants in response to the scenes and the discussion. Interestingly, the range of ideas mentioned in the student and adult citizen performances was strikingly similar. The two groups merely differed in the extent to which the participants were able to articulate the background of their ideas and concerns. Moreover, the entire range of ideas was expressed in almost every performance. Differences were primarily found in the relative weight that was attributed to particular ideas and concerns.

One of the central ideas of value pluralism, as it is commonly understood in contemporary philosophy, is the incompatibility of values (Williams 1981; Smith 2003). The incompatibility of values occurs when two (or more) possible ideals cannot be fulfilled at the same time. Conflicts between incompatible values are numerous present in our moral experience. In the context of emerging technologies, they may seem even more unavoidable. For example, one of the scenarios played out in the

performance was the connection of implantable nanobiosensors to smart devices in the living environment that could enable the continuous monitoring of bodily function. Monitoring was appreciated by some participants as a way to improve efficiency in diagnosis, treatment, and performance, but also transparency and fairness, for example, when it comes to using public services. At the same time, participants were concerned about their privacy if their personal information was shared with other actors such as care providers, insurance companies, or employers. The surfacing of values like efficiency, privacy, or transparency in the discussion is, of course, not that new. The added value of the theatrical debate is that it produced a much more contextual view on what these values meant for a specific character in a specific situation. The process of inquiry, as a result, was much more focused on how the pursuit of those values in that situation could improve human life of the characters involved or their wider environment. As Minter, Corley and Manning (2004) pointed out, Dewey's reconstruction of ethics aimed to shift the discussion of moral theory and argument away from a preoccupation with general principles to the development of methods for cooperative problem solving. The theatrical debate methodology has shown to work as such a method. As the abovementioned example illustrates, the conflict between efficiency, transparency, and privacy is not discussed in general, but in relation to a problematic situation that needs to be resolved. We noticed that the contextual and open process of inquiry of the theatrical debate performances produced a number of different insights. First, it was much easier for the participants to stay focused when they were confronted with a situation that had to be resolved compared to the exchange of general arguments in a regular public debate. Second, the context of the indeterminate situation revealed that every decision changes the entire constellation of values that are either propagated or threatened. Third, these conflicts between different values not only manifested themselves between individuals, but also within individuals. Participants indicated that they felt pushed back and forth by different contrasting ideas. They simultaneously appreciated different values that were difficult to reconcile and pulled them in contradictory directions. Seeing that others share the same conflicts of values, be it that they arrive at a different judgment, improved mutual understanding.

Ethical reflection and decision-making tools often needlessly limit the broad range of possible interactions with a moral problem and thereby fail to represent the diversity of moral experiences and values (Minter, Corley and Manning 2004). One of the interesting features of the theatrical debate methodology is that it visualizes the contradiction and complexity of our moral experience. One of the ways by which the theatrical debate methodology encourages inquiry into the diversity of moral experiences and values is via the use of collective interpretation. There is always more to a situation than an individual can tell (or even know). Collective interpretation reveals, as Boal

(1979) called it, the multiple truth of an image. Let us return to the *Doctor Inside!* scenario to understand the role of collective interpretation during the theatrical debate performances. One of the salient changes that the participants recognized in the *Doctor Inside!* scenario was the shift in responsibility for the care provided. In the sketch of the indeterminate situation, the act of care was perceived to move away from the treatment room of the medical practitioner to the living room of the patient. Many participants indicated that they appreciated this development because it would give the patient more autonomy. There were also critical remarks, questioning the benefits of more responsibility for the patient. Such a discussion would evolve into a process of collective inquiry into the meaning of patient autonomy in the context of responsibility for healthcare provision. Different participants would highlight different aspects of the indeterminate situation, making sense of patient autonomy in relation to responsibility from their perspective. This offered a more contextual and complete understanding of the situation at hand. Furthermore, the dramatization of the situation in which a patient was making healthcare decisions in interaction with a monitoring device in the absence of medical practitioners allowed the participants to discover new dilemmas. Direct interaction with the monitoring devices would give a patient access to previously concealed information. At the same time, this would require the knowledge and skills to interpret and use this information. Direct interaction with the monitoring devices would enable the patient to make their own decisions. At the same time, participants recognized that the expected health benefits would then depend on the patient's attitude towards a healthy lifestyle. But what if a patient does not value a healthy lifestyle very much? Would a medical practitioner have to take control in that situation? Or is it in fact a matter of autonomy to decide for oneself whether to pursue the value of healthy living or any other value? The actors would respond to these kind of questions raised in the audience by acting out the consequences of different courses of action that were suggested. The dramatization of the outcomes of alternative decisions invited a reflection on underlying values and assumptions, contributing to a more reflexive understanding of the situation at hand as well as the perspective that participants used to make sense of that situation. As one participant put it, "It is beautiful to see multiple points of view and develop your own opinion." Collective interpretation thus contributed to frame reflection, the process of becoming aware of one's own perspective in relation to the perspectives of others (Schön and Rein 1994).

The theatrical debate methodology as a way of doing experimental ethics

This volume articulates different approaches to experiment with technologies in society. The difficulty of dealing with experimental technologies is mainly related to its uncertainties, especially in an early phase of development. Most approaches to the management and control of emerging

technologies have tried to reduce uncertainty by anticipating possible negative consequences and modulating development trajectories towards positive ends. However, according to Van de Poel (2015), anticipation has its limits precisely because of uncertainties and ignorance. Conceiving of the introduction of novel technologies as a societal experiment provides an alternative way to deal with emerging technologies by accepting their inherent uncertainties and monitoring and adapting their development. The gradual and experimental introduction of technology in society is in line with Dewey's ideas about experimental ethics. Dewey emphasized the need to reconstruct our interpretive frameworks in order to respond more productively to new and unexpected situations (see Dewey 1920). Moral principles should be seen as hypotheses to be tested out in new situations. Indeed, Dewey's notion of dramatic rehearsal is tightly connected to experimentation with the imagined courses of action in real life in order to examine their consequences. In this chapter, I have proposed the theatrical debate methodology as a way of doing experimental ethics from a Deweyan perspective. The theatrical debate performances, however, constitute a form of simulation experiment, not an actual experiment. Does this approach run into the same limits as other anticipatory approaches, for example, a misleading emphasis on unlikely but morally thrilling scenarios (see Van de Poel 2015)? Why conduct simulation experiments if there is also the possibility to design the introduction of nanotechnology itself as a societal experiment?

The first argument to discuss here returns to the dilemma of control, formulated by Collingridge (1980). Whereas nanotechnology may be expected to have severe implications for society, many innovation trajectories are still in a very early phase of development, potential consequences are largely hypothetical, and there are many uncertainties involved. However, now is the time to reflect on potential implications because changes can still be made. How much adaptive space will be left to alter the course of innovation after nanotechnology is experimentally introduced? Will it not be more likely that positions and practices will become fixed by the values and interests of the stakeholders involved? One could even imagine that inquiry and deliberation at this stage is necessary to engage different stakeholders in such a societal experiment. Krabbenborg (2013b) points to the use of techno-moral future scenarios as a way of controlled speculation (see also te Kulve and Rip 2011), diagnosing uncertainties and ambiguities in a narrative that shows possible actions and consequences. These kind of scenarios may be used as part of the repertoire of pragmatist ethicists for conceptual analysis, the renewal of existing frameworks and vocabularies, or during workshops as a platform to structure the interaction between participants. In the past years, various TA scholars and ethicists have taken up this task (see for example Robinson 2009; Boenink, Swierstra, and Stermerding 2010; Lucivero, Swierstra, and Boenink 2011). There is a second issue here, next to the limits of anticipating uncertainties: there is no public yet. As Krabbenborg (2013b) herself indicated, Dewey believed it was the experience of disruption that would trigger people to

organize themselves and start a process of inquiry out of an urge to resolve a problematic situation. In the context of nanotechnology, concrete disruptions may not have been experienced yet. The future scenarios depicting potentially problematic actions and consequences are the result of dramatic rehearsals undertaken by the ethicist/designer. Also, the theatrical debate does not fully escape this point of critique. An important difference is that the scenarios presented to the participants in the theatrical debate performance are still at the level of the indeterminate situation, ambiguous and open-ended. During the performance, the participants are invited to explore the indeterminate situations, diagnose the ambiguities themselves, and co-construct the scenarios by immediate inquiry and deliberation.

The second argument refers to the difference between mental trials and overt trials. Although Dewey emphasizes the real-life experimental testing of hypotheses that emerged as the best solution out of the process of inquiry, he also recognizes the advantages of mental trials. In a discussion of the role of dramatic deliberation in ethical inquiry, Dewey (1932, 275) compares a mental trial with an overt trial:

A mental trial... is retrievable, whereas overt consequences remain. They cannot be recalled. Moreover, many trials may mentally be made in a short time. The imagining of various plans carried out furnishes an opportunity for many impulses which at first are not in evidence at all, to get under way.

Also in the theatrical debate performances, the consequences acted out on stage are simulated and can thus be recalled. It allowed the participants to experiment with alternative imaginations of the same scenario and, importantly, share their interpretations by working together on the same situation that was portrayed on stage. Interestingly, the participatory deliberative forum that was created in the theatrical debate is somewhere between a mental and an overt trial. Its dramatizations have simultaneously been a *real* and *imagined* inquiry into the future societal embedding of nanotechnology applications. Participants tried out not only different courses of action in their imagination, but also in the dramatization of actions and consequences on stage. Hanghøj (2011) observed a similar phenomenon in his study of debate games. Precisely the lack of real-life consequences enable the creation of a safe and free discussion environment where citizens can experiment with different future scenarios, enact their ideas and concerns, make mistakes, and explore their perceptions, understandings, and appreciations.

A third and related argument is that the theatrical debate methodology offers the possibility to experiment with different framings of the indeterminate situation. Even though conceived as gradual and experimental, the actual introduction of nanotechnology in society would also introduce (and reinforce) particular framings of nanotechnology into its practice and discourse. These framings would result in specific problem-setting stories (see Schön and Rein

1994) that would shape ethical reflection. An example of these dynamics of framing is the persistence of the risk discourse in discussions about technologies, despite attempts to broaden the discourse. The task of experimental ethics should not only be to try out different responses to particular problematic situations, but also to try out different interpretations (or framings) of indeterminate situations. The simulation experiments performed in the theatrical debate can accommodate different framings more easily precisely because the experiment is not embedded in a real existing practice. The productive ambiguity of the situations portrayed on stage invites the experimentation with different framings to see how they shape potential actions and consequences. Participants indicated that they appreciated the theatrical debate performances as a very basic exercise in philosophy, challenging habitual ways of thinking by testing their consequences on stage.

The theatrical debate methodology was designed to accommodate the dynamic character of nanotechnology developments and the moral pluralism of society. Indeed, the dynamic exploration of the scenarios that were unfolding on stage contributed to the creative imagination of new solutions to potential controversies. During the process of dramatic rehearsal, it became clear that the different perspectives expressed by the audience all contributed to the understanding of the indeterminate situation that was portrayed on stage. The mechanism of collective interpretation emphasized the multiple truths of these perspectives. As Keulartz et al. (2002, 262) argued, "conflicting parties have to appreciate the fact that they are competing for primacy within the same universe of discourse with others that cannot beforehand be branded as unreasonable." I introduced the theatrical debate methodology as a combination of two different tasks for pragmatist ethics in a technological culture, as outlined by Keulartz et al. (2004): the dramatic rehearsal of new technological worlds and the open confrontation of worldviews. Although the proposal of Keulartz et al. (2004) do commit to a democratic approach directed towards the resolution of conflicts in moral life, they adhere to an understanding of the ethicist as the principal source of knowledge about the structure of moral problems and the rightful candidate for inventing new ethical concepts or ways of understanding the world. The theatrical debate methodology more radically changes the understanding of the role of the ethicist in the moral deliberation of new emerging technologies. In this methodology, the ethicist primarily operates as a designer and facilitator of cooperative processes of inquiry and deliberation. The reconstruction of interpretive frameworks is performed in close interaction with its participants, following their understandings and priorities. Exchanging the armchair for a societal laboratory, the ethicist as a designer/facilitator integrates experimental philosophy with the ideal of participatory democracy. Technology controversies, like the nanotechnology controversy, inevitably have a public character. A reconstruction of nanotechnology ethics will therefore only be effective when it is designed as a democratic dialogue directed at

the reconstruction of public morality. The theatrical debate has the capacity to bring people together and enable them to engage mutually in the process of experimental inquiry and creative deliberation. As such, the theatrical debate methodology can make a both playful and profound contribution to the implementation of creative democracy.

References

- Bijker, Wiebe E., Thomas P. Hughes, and Trevor J. Pinch. 1987. *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge (Ma.): MIT press.
- Boal, Augusto. 1979. *Theater of the Oppressed*. London: Pluto Press.
- Boal, Augusto. 1998. *Legislative Theatre*. New York: Routledge Press.
- Boal, Augusto. 2002. *Games for Actors and Non-Actors*. London and New York: Routledge.
- Boenink, Marianne, Tsjalling Swierstra, and Dirk Stemerding. 2010. "Anticipating the Interaction between Technology and Morality: A Scenario Study of Experimenting with Humans in Bionanotechnology." *Studies in Ethics, Law, and Technology* 4 (2):1–38.
- Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative research in psychology* 3 (2):77–101.
- Caspary, William R. 1991. "Ethical Deliberation as Dramatic Rehearsal: John Dewey's Theory." *Educational Theory* 41 (2):175–88.
- Caspary, William R. 2000. *Dewey on Democracy*. Ithaca, NY: Cornell University Press.
- Collingridge, David. 1980. *The Social Control of Technology*. London: Frances Pinter.
- Delgado, Ana, Kamilla Lein Kjølberg, and Fern Wickson. 2010. "Public Engagement Coming of Age: From Theory to Practice in STS Encounters with Nanotechnology." *Public Understanding of Science* 20 (6):826–45.
- Dewey, John. 1920. *Reconstruction in Philosophy*. New York: H. Holt and Company.
- Dewey, John. 1922. *Human Nature and Conduct: An Introduction to Social Psychology*. New York: Modern Library.
- Dewey, John, 1932. "Moral Judgement and Knowledge." In *The Later Works of John Dewey, Volume 7, 1925–1953: 1932, Ethics*, edited by Jo Ann Boydston, 262–284. Vol. 7. Carbondale: SIU Press.
- Diamond, David, and Fritjof Capra. 2007. *Theatre for Living: The Art and Science of Community-Based Dialogue*. Victoria: Trafford Publishing.
- Dwyer, Paul. 2004. "Making Bodies Talk in Forum Theatre." *Research in Drama Education* 9 (2):199–210.
- Felt, Ulrike, Simone Schumann, Claudia G. Schwarz, and Michael Strassnig. 2014. "Technology of Imagination: A Card-Based Public Engagement Method for Debating Emerging Technologies." *Qualitative Research* 14 (2):233–51.
- Fesmire, Steven. 2003. *John Dewey and Moral Imagination: Pragmatism in Ethics*. Bloomington, IN: Indiana University Press.
- Freire, Paulo. 2002. *Pedagogy of the Oppressed*. New York: Continuum.
- Grin, J., and A. Grunwald. 2000. *Vision Assessment: Shaping Technology in 21st Century Society. Towards a Repertoire for Technology Assessment*. Heidelberg: Springer.

- Hanghøj, Thorkild. 2011. *Playful Knowledge. An Explorative Study of Educational Gaming*. Saarbrücken: Lambert Academic Publishing.
- Hanssen, Lucien. 2009. *From Transmission towards Transaction. Design Requirements for Successful Public Participation in Communication and Governance of Science and Technology*. Nijmegen: University of Twente.
- Hickman, Larry A. 1998. *Reading Dewey: Interpretations for a Postmodern Generation*. Bloomington, IN: Indiana University Press.
- Hickman, Larry A. 2001. *Philosophical Tools for Technological Culture: Putting Pragmatism to Work*. Bloomington, IN: Indiana University Press.
- Irwin, Alan. 2001. "Constructing the Scientific Citizen: Science and Democracy in the Biosciences." *Public understanding of science* 10 (1):1–18.
- Jasanoff, Sheila. 2003. "Technologies of Humility: Citizen Participation in Governing Science." *Minerva* 41 (3):223–44.
- Keulartz, Jozef, Michiel Korthals, Maartje Schermer, and Tsjalling Swierstra. 2002. *Pragmatism in Action*. In: *Pragmatist Ethics for a Technological Culture*, edited by J. Keulartz, Korthals, M., Schermer, M., and Swierstra, T., 247–64 Dordrecht: Kluwer Academic Publishers.
- Keulartz, Jozef, Maartje Schermer, Michiel Korthals, and Tsjalling Swierstra. 2004. "Ethics in Technological Culture: A Programmatic Proposal for a Pragmatist Approach." *Science, technology & human values* 29 (1):3–29.
- Krabbenborg, Lotte. 2013a. *Involvement of Civil Society Actors in Nanotechnology: Creating Productive Spaces for Interaction*. Groningen: RU Groningen.
- Krabbenborg, Lotte. 2013b. "Dramatic Rehearsal on the Societal Embedding of the Lithium Chip." In: *Ethics on the laboratory floor*, edited by Simone van der Burg and Tsjalling Swierstra, 168–87. Basingstoke: Palgrave Macmillan.
- Kupper, Frank, and Tjard De Cock Buning. 2011. "Deliberating Animal Values: A Pragmatic—Pluralistic Approach to Animal Ethics." *Journal of agricultural and environmental ethics* 24 (5):431–50.
- Lucivero, Federica, Tsjalling Swierstra, and Marianne Boenink. 2011. "Assessing Expectations: Towards a Toolbox for an Ethics of Emerging Technologies." *NanoEthics* 5 (2):129–41.
- Macnaghten, Phil, Matthew B. Kearnes, and Brian Wynne. 2005. "Nanotechnology, Governance, and Public Deliberation: What Role for the Social Sciences?" *Science communication* 27 (2):268–91.
- McGee, Glenn. 2003. *Pragmatic Bioethics*. Cambridge, MA: MIT press.
- McKenna, Erin, and Andrew Light. 2004. *Animal Pragmatism: Rethinking Human-Nonhuman Relationships*. Bloomington, IN: Indiana University Press.
- Minteer, Ben A., Elizabeth A. Corley, and Robert E. Manning. 2004. "Environmental Ethics beyond Principle? The Case for a Pragmatic Contextualism." *Journal of Agricultural and Environmental Ethics* 17 (2):131–56.
- Odegard, Marianne. 2003. "Dramatic Science: A Critical Review of Drama in Science Education." *Studies in Science Education* 39:75.
- Palm, Elin, and Sven Ove Hansson. 2006. "The Case for Ethical Technology Assessment (eTA)." *Technological Forecasting and Social Change* 73 (5):543–58.
- Pappas, Gregory F. 1998. "Dewey's Ethics: Morality as Experience." In *Reading Dewey: Interpretations for a Postmodern Generation*, edited by Larry A. Hickman, 100–123. Bloomington: Indiana University Press.
- Perry, Adam J. 2012. "A Silent Revolution: 'Image Theatre' as a System of Decolonisation." *Research in Drama Education: The Journal of Applied Theatre and Performance* 17 (1):103–19.

- Pidgeon, Nick, and Tee Rogers-Hayden. 2007. "Opening Up Nanotechnology Dialogue with the Publics: Risk Communication or 'Upstream Engagement'?" *Health, Risk & Society* 9 (2):191–210.
- Rip, Arie, and Rene Kemp. 1998. "Technological Change." In *Human Choice and Climate Change*, edited by S. Rayner and E.L. Malone, 327–399. Columbus, Ohio: Battelle Press.
- Robinson, Douglas K.R. 2009. "Coevolutionary Scenarios: An Application to Prospecting Futures of the Responsible Development of Nanotechnology." *Technological Forecasting and Social Change* 76 (9):1222–39.
- Roco, Mihail C., and William S. Bainbridge. 2001. *Societal Implications of Nanoscience and Nanotechnology*. Arlington: National Science Foundation.
- Schön, D., and Martin Rein. 1994. *Frame Reflection: Resolving Intractable Policy Issues*. New York: basic books.
- Schutzman, Mady, and Jan Cohen-Cruz, eds. 1994. *Playing Boal: Theatre, Therapy, Activism*. Oxon: Routledge.
- Shelley-Egan, Clare. 2011. Ethics in Practice: Responding to an Evolving Problematic Situation of Nanotechnology in Society. PhD diss., University of Twente.
- Smith, Graham. 2003. *Deliberative Democracy and the Environment*. Oxon: Routledge.
- Steen, Marc. 2013. "Co-Design as a Process of Joint Inquiry and Imagination." *Design Issues* 29 (2):16–28.
- Swierstra, Tsjalling, and Arie Rip. 2007. "Nano-Ethics as NEST-Ethics: Patterns of Moral Argumentation about New and Emerging Science and Technology." *Nanoethics* 1 (1):3–20.
- Swierstra, Tsjalling, and Hedwig te Molder. 2012. "Risk and Soft Impacts." In *Handbook of Risk Theory*, edited by Sabine Roeser, Rafaela Hillerbrand, Per Sandin and Martin Peterson, 1049–1066. Springer Netherlands.
- Swierstra, Tsjalling, Dirk Stemerding, and Marianne Boenink. 2009. "Exploring techno-moral change: the case of the obesity pill." In *Evaluating new technologies*, pp. 119–138. Springer Netherlands.
- Swierstra, Tsjalling. 2013. "Nanotechnology and Technomoral Change." *Etica & Politica / Ethics & Politics* 15 (1):200–219.
- Te Kulve, Haico, and Arie Rip. 2011. Constructing Productive Engagement: Pre-engagement Tools for Emerging Technologies. *Science and engineering ethics* 17 (4):699–714.
- Van de Poel, Ibo. 2015. "An Ethical Framework for Evaluating Experimental Technology." *Science and Engineering Ethics*: 1–20.
- Verbeek, Peter-Paul. 2009. "Moralizing Technology: On the Morality of Technological Artifacts and Their Design." *Readings in the Philosophy of Technology*: 226.
- Williams, Bernard. 1981. *Moral Luck: Philosophical Papers 1973–1980*. Cambridge University Press.
- Wilsdon, James, and Rebecca Willis 2004. *See-through Science: Why Public Engagement Needs to Move Upstream*. Demos.
- Wilsdon, James, Brian Wynne, and Jack Stilgoe. 2005. *The Public Value of Science. Or How to Ensure that Science Really Matters*. London: Demos.
- Winston, Joe. 1999. "Theorising Drama as Moral Education." *Journal of Moral Education* 28 (4):459–71.
- Wynne, Brian. 2001. "Creating Public Alienation: Expert Cultures of Risk and Ethics on GMOs." *Science as culture* 10 (4):445–81.

5 Social learning in the bioeconomy

The Ecover case

Lotte Asveld and Dirk Stemerding

Introduction

Developing new technologies to achieve sustainability is a complex endeavor because the notion of sustainability is complex and perceived differently by different societal actors. The general notion of sustainability is deemed desirable by most people. Nevertheless, in specific cases such as the assessment of a new technology, differences in perception often come to the fore.

This is what a producer of sustainable cleaning products, Ecover, discovered when it started using oil derived from genetically engineered algae, and was criticized by a coalition of environmental NGOs that stated that anything derived from engineered algae cannot be considered sustainable.¹ Ecover and Solazyme, the producer of the algae oil, considered oil produced by genetically engineered algae to be more sustainable than alternatives such as palm oil. This judgement was based on a Life Cycle Assessment (LCA), which is a tool to determine the sustainability impact of technologies. The LCA of the algae oil turned out to be better than that of palm oil. This outcome was due to the fact that sugar cane, which is the feedstock for algae oil, is supposed to have a lower environmental footprint than palm oil. Additionally, the algae can switch to another feedstock, should it appear to be more sustainable. Even though this innovation was deemed sustainable to Solazyme and Ecover, they met with an unexpected critical societal response and, subsequently, Ecover abandoned the innovation. We will elaborate on this below.

We argue that an experimental approach would have enabled all actors involved to contribute to societally desirable innovations more effectively while reducing uncertainties about what sustainability entails. New technologies that emerge under the banner of sustainability bring about new uncertainties. Questions arise about the exact environmental impact of these new technologies, about whether new regulations are required to control new potential risks or to stimulate innovation, and about what sustainability amounts to and what innovation trajectories should be stimulated to achieve sustainability (Asveld, Est, and Stemerding 2011). Because of these uncertainties, the introduction of sustainable technologies can be

considered *de facto* social experiments, i.e., experiments that are not intentional but rather occur without explicit planning and without being named as such (Asveld 2016).

A major source of uncertainty surrounding sustainable technological developments lies in the different frameworks that actors use to assess a technology. In making sense of the world around us, everyone relies on specific frameworks of meaning that may lead to different perceptions of various issues. They will, for instance, determine our perception of a technology (Kupper et al. 2007; Brom et al. 2011). They will lead to specific answers to general questions such as: what purpose does this technology serve? How should it be managed? Is it risky? Divergent frameworks will produce divergent answers to such questions.

The existence of diverging frameworks gives rise to moral ambiguity about the desirability of a new technology, implying moral ambiguity in a social sense. While individual actors may have a clear sense of how to assess a new technology, the variety of coexisting perceptions of the technology in society means there is unlikely to be consensus, which creates social, moral ambiguity. Moral ambiguity in this chapter hence refers to the lack of a societally dominant frame to assess the sustainability of a new technology.

The main question addressed in this chapter is: How can we overcome moral ambiguity in the introduction of new sustainable technologies? We propose that the introduction of sustainable technologies should be approached as deliberate experiments, suggesting that the actors organize themselves to optimize social learning effects. Social learning can be directed at reducing various uncertainties, but often, reducing moral ambiguity requires the most attention. Social learning exercises can be modelled after commonly used technological development trajectories involving qualities such as careful scaling up and adaptability to learning effects. To ensure that an exercise results in effective social learning, it should both focus more broadly than on only the specific technology and take into consideration the particularities of those specific technologies.

To illustrate how social learning might be organized for sustainable innovation trajectories, we explore the Ecover case described above. The innovation took place in the context of the transition to a sustainable bioeconomy, which further increases complexity. We claim that an experimental approach would have reduced moral ambiguity and would have enabled all actors involved to contribute to societally desirable sustainable innovations more effectively while reducing uncertainties about what a sustainable bioeconomy entails in general.

Social learning

As several authors have pointed out (Glasser 2007; Armitage, Marschke, and Plummer 2008; Van Mierlo et al. 2010), there is no widely shared definition of social learning. Some definitions focus on changes in institutions,

whereas others focus on the effects learning should have on individual participants. However, many authors do agree on the need for social learning in the context of achieving sustainable societies.

The claim that social learning is a requirement for a sustainable society derives from adaptive management where the sustainable management of natural resources often involves conflict resolution between many diverging claims regarding the resource at stake (Armitage, Marschke, and Plummer 2008). The lessons derived from such conflicts provide valuable input for new policies and are essential to any societal transition (Glasser 2007). Additionally, the connection between social learning and sustainability entails the idea that individual actions and preferences will fall short of achieving a sustainable society. Sustainability requires a shared perspective that guides many actors in society as a whole, and social learning is a means to arrive at such a shared perspective (*ibid.*).

We consider reducing moral ambiguity to be the main element of social learning, especially in the context of the bioeconomy, our focus here. The bioeconomy is an economy in which biomass is the main resource for energy, materials, and chemicals, and this resource is used with the utmost efficiency. Biomass can be obtained from food and feed crops, non-food crops, woody forest-based sources, and various types of wastes and residues, including the biodegradable fractions of municipal or industrial wastes. The use of biomass for the replacement of fossil resources has been hailed as a sustainable solution that might help combat both climate change and pollution (OECD 2009; EC 2012). Because the transition to a sustainable bioeconomy involves a wide variety of societal actors and sectors, moral ambiguity can be expected to be considerable.

Three levels of uncertainty

In general, uncertainties on moral, physical, and institutional aspects can be considered relevant when learning about new technologies in society (Van de Poel 2017, Grin and Van de Graaf 1996) and when managing natural resources (Armitage, Marschke, and Plummer 2008), all relevant for the bioeconomy.

Physical impact refers to the risks and benefits of a specific technology, such as environmental impact or health effects. The institutional impact refers to the kind of social structures and (legal) rules needed to adequately embed a technology in society. Institutions can be understood as the formalization of explicit or tacit agreements between multiple actors (*cf* Bachmann and Inkpen 2011) with sufficient legal or moral status to influence the further development and use of relevant technologies. Examples are the regulation of risky technologies, subsidy schemes, or voluntary schemes for the monitoring of sustainability effects.

We argue that in the bioeconomy, social learning will, to a large extent, concern moral learning and thereby entail the reduction of uncertainty regarding the norms and values we use to evaluate a technology. We believe

that moral learning requires the exploration of various frameworks present in society. These can make us more aware of our own assumptions and values. Furthermore, arriving at a shared perspective on the desirability of a new technology requires knowledge of the frameworks of other relevant actors.

Moral learning is fundamental to the development of a bioeconomy because biobased innovations not only encompass the introduction of new technologies, but also may involve a restructuring of value chains and accompanying professional roles and dependencies. This implies a redistribution of economic benefits and burdens which gives rise to questions about fairness and justice. Furthermore, the bioeconomy touches upon values that hold wide societal appeal such as naturalness, food security, and global justice (Asveld et al. 2014).

The development of a bioeconomy will not solely require moral learning. Other relevant types of uncertainty that need to be addressed are related to the physical impact of new technologies and to the shaping of institutions (Asveld 2016). Because the bioeconomy involves wide-ranging societal effects involving a wide range of actors, it can be expected that moral learning is fundamental. When moral uncertainty persists, physical and institutional uncertainty also endure. This is related to the kind of uncertainty that typifies moral uncertainty.

Epistemological, indeterminate, and ambiguous uncertainty

Van de Poel (2017) distinguishes between two kinds of uncertainty: indeterminate and epistemological. Both can be undesirable when we want to establish the sustainability of a specific innovation. Epistemological uncertainty can usually be resolved by producing more knowledge, for instance, by monitoring the effects of a specific innovation. Indeterminate uncertainty exists when “multiple causal chains toward the future are still open” (Van de Poel, 2017). This type of uncertainty can only be reduced by choosing one specific course of action.

We would like to add a third type of uncertainty—ambiguity. Ambiguity arises from diverging perspectives on a specific matter where it is unclear which of these perspectives take precedence. Take for instance the introduction of biofuels (Asveld 2016) in which the uncertainty was initially indeterminate. It could not be resolved with more knowledge because the causal chains could only become determinate when biofuels were used on a larger scale. From the onset of their introduction, different assessments were made about the amount of CO₂ biofuels produced. These assessments differed because many aspects of their production had not materialized yet. How would they be grown? How would they be transported? Once biofuels were used on a larger scale, the measurements became more reliable.

With wider use of biofuels, the uncertainty became epistemological, suggesting that uncertainty arose from a lack of knowledge, which was resolved

with improved monitoring (cf. Van de Poel 2017). However, while the measurements and data have improved, there is still uncertainty because the knowledge generated can be interpreted in different ways. Actors disagree about what aspects of biofuel production should be included in the calculation of CO₂ production. Does the production of biofuel have indirect effects that should also be calculated, such as the clearing of land somewhere else to replace the land that is now being used for biofuels? This so-called “indirect land use change” is a major point in the debate on how to establish the CO₂ production of biofuels. This uncertainty is due to diverging frameworks of reference and cannot be solved by learning by doing or producing more knowledge. It should as such be considered a third type of uncertainty: ambiguity.

Ambiguous uncertainty is to a large extent moral in nature. The diverging frameworks of reference can be technological in nature, for instance, between competing design frameworks (Grin and Van de Graaf 1996) that might be resolved by putting the designs into practice and comparing performance. However, when diverging frameworks are partly moral in nature, resolution is more difficult to achieve, because diverging moral frames of reference comprise not only values but also assumptions and beliefs difficult to falsify, for example, regarding macroeconomic systems and societal structures. We will return to these frameworks later on (section 3).

Different types of uncertainty in the Ecover case

How did these different uncertainties figure in the Ecover case?

As recounted above, Ecover and Solazyme aimed to introduce a new sustainable biobased product. However environmental organizations such as the ETCGroup and Friends of the Earth US disputed this claim to sustainability. These NGOs are critical of a range of specific technologies including biotechnology, nanotechnology, and geo-engineering. To support their position, they put forward arguments that are not covered in an LCA, such as economic equality. They voiced concerns of a macroeconomic nature, stating that genetically, or synthetically, engineered organisms allowed for a misbalance in economic power due to patents and a concentration of knowledge with powerful companies. They furthermore expressed concerns that farmers of other forms of sustainable oil such as coconut oil would suffer from the algae oil production. As an additional concern, they stated that the engineered algae posed great risks to the natural environment: even though the algae are contained, they may still escape, for instance, when they latch on to someone's lab coat and undermine the stability in nearby ecosystems (FOE 2011).

These diverging lines of argumentation show the effect of diverging frameworks. Ecover and Solazyme supported their position with arguments related to measurable impact as captured by an LCA. The arguments of the ETC group and associates regarded the indeterminate physical impacts

(the risks), the possible future effects of institutions (the patents), and the moral impact (economic equality). Thus, the arguments of the environmental coalition concern the institutional embedding of this innovation and its long-term economic effects on inequality, whereas Ecover and Solazyme employ the measurable physical impact of the algae in their LCA argument. However, most of the concerns of the environmental coalition cannot be captured by an LCA (Flipse 2014).

In the Ecover case, uncertainty mainly consists of ambiguity. When we operate in a context of uncertainty that cannot easily be resolved, we rely on our own frameworks to assess the desirability of a technology (Asveld 2008). The actors have diverging frameworks of reference and moral orientations. They agree on the importance of sustainability, but what sustainability entails is a matter of debate. Whether the engineered algae will have a positive sustainability effect or will pose a threat to the environment remains to be seen. The various assumptions the actors make determine their assessment of the risks and benefits of this innovation, as well as the macroeconomic effects of patents and the impact on small-scale producers of competing oils. These issues involve persisting uncertainties which lead to social ambiguity featuring diverging moral points. How can social learning be organized in such a case?

Moral ambiguity due to different frameworks

Frameworks, worldviews, and social learning

As said above, individuals each have their own frame of reference consisting of values, beliefs, and convictions that determine how they perceive and evaluate the world around them (Schön and Rein 1994). The existence of divergent frameworks is a source of uncertainty, namely social ambiguity. However, for the individual adhering to a specific frame, it is not a source of uncertainty, but instead helps him or her to reach a specific perspective.

Social learning does not necessarily resolve social ambiguity. Moreover, social ambiguity might not be resolved at all, but making ambiguity explicit can nonetheless lead to learning effects that benefit all actors and society more generally (Grin and van de Graaf 1996) because such social learning helps us articulate what contested issues such as sustainability imply and what we might do to bring it about (Loeber et al. 2007), even if we do not achieve a general consensus.

Reflection on one's own framework and comparison to those of others can increase awareness of a personal framework (Kupper et al. 2007; Loeber et al. 2007). As Loeber et al. (2007) state:

We need others to help us notice not only what we fail to observe because of practical reasons (lack of information due to time lapse or distance) but also because of “what [we] have worked to avoid seeing” (Schön 1983, 283).

(88)

For instance, in the Ecover case, the environmental associations compelled Ecover and Solazyme into an (unintentional) social learning process. The criticism of Ecover and Solazyme actually brought forward uncertainties not recognized before. Ecover and Solazyme were surprised by questions the environmental organizations posed. They thought of their innovation as incremental, not as fundamental, change. Learning about opposing views made these actors aware of their own assumptions. Additionally, learning about the frameworks of others may offer a base for a true consensus, thus going beyond a mere compromise (Van de Poel and Zwart 2010; Doorn 2010; Rölöing 2002).

To ensure that the frameworks considered in a social learning exercise sufficiently cover all effects on institutions, values, and physical environment, representation of all relevant frameworks is key (Cuppen 2012). A consideration of worldviews will ensure both sufficient representation and an understanding of the salient issues in a prospective learning exercise. Frameworks leading to diverging perspectives on technological developments can be linked to worldviews. Worldviews are culturally dominant frameworks of meanings that are shared by a wide range of people in society (Kahan 2012; Hedlund-de Witt 2013). They consist of coherent structures of values, beliefs, and attitudes. An indication that representation is sufficient is when all dominant worldviews held in a society are included in a social learning exercise.

In Figure 1, we depict our understanding of the relation between frameworks, perspectives, and worldviews. An individual's framework will never completely coincide with one worldview, but likely will be consistent with one worldview more than another. Individual frameworks are more complex and nuanced than worldviews, but the individual frameworks often take shape with reference to the overarching sociocultural worldviews. Worldviews help us to grasp some of the recurring tensions in debates on sustainability even if we haven't explicated the depth of each individual frame involved in such tensions. Figure 5.1 shows the relationship between worldviews, frameworks, and perceptions.

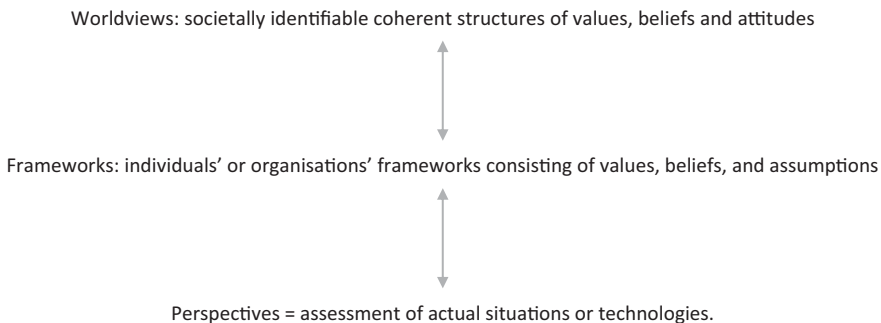


Figure 5.1 Relationship between worldviews, frameworks and perceptions.

A wide body of literature exists with regard to worldviews in which these worldviews are usually brought down to four distinct types. Defining characteristics often include issues such as an orientation on the local or the global (Douglas and Wildavsky 1982, De Vries and Petersen 2009) perceptions of the role of the government (De Vries and Petersen 2009), the vulnerability of the natural environment (Douglas and Wildavsky 1982; Kahan 2012), our responsibilities towards that environment (Dryzek 2013), and expectations about the possible beneficial effects of technology (Hedlund-de Witt 2013).

Worldviews in the Ecover case

In existing descriptions, worldviews are often depicted in a quadrant. An example of such a quadrant that appears relevant to the Ecover case is depicted below. We think this particular quadrant is relevant because it includes both views on nature and views on the desirable organization of society. We rely on the work of Douglas and Wildavsky (1982) and Thompson, Ellis, and Wildavsky (1990) in which they distinguish four prevalent views on nature.

In this case, the main clash appears between a worldview of “nature as a resource” and that of “vulnerable nature.”

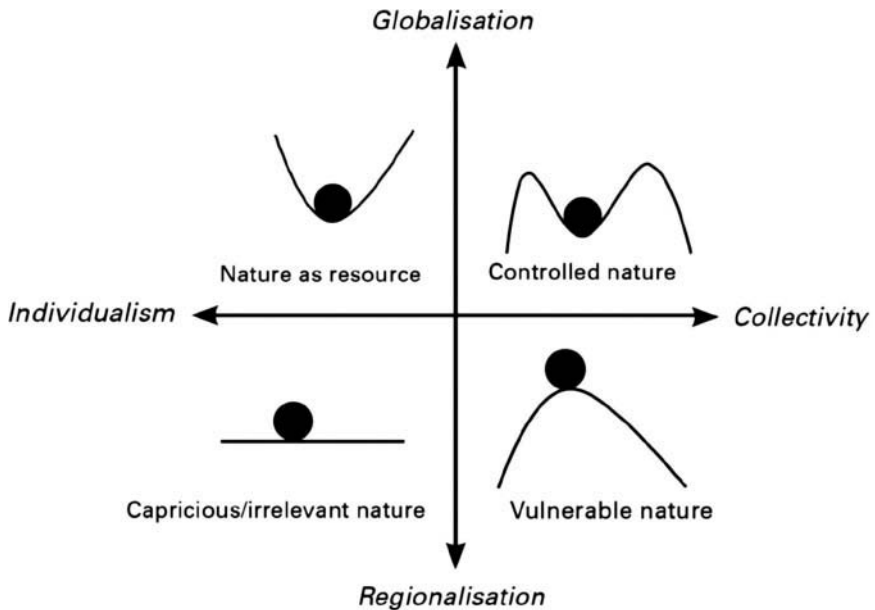


Figure 5.2 ‘Four prevalent views on Nature’.

Nature as resource

In the first worldview, nature is perceived of as resilient. Nature can therefore be exploited (“nature as resource”) without problems. Additionally, market forces are expected to produce beneficial (sustainable) effects for everyone.² This group may include multinationals (for instance, the Dutch Sustainable Growth Coalition), but also pioneers of new technology, such as small start-ups and venture capitalists. This group does not support regulation, in line with its individualistic attitude and aversion to collectivism. Members of this group might³ state the following in relation to the algae case:

We need to move away from our old production mechanisms for environmental and marketing reasons. Tinkering with existing natural production mechanisms opens up vast and increasingly efficient production avenues for beautiful, renewable, and sustainable resources that might lead us out of our current environmental predicament.

Innovations are key to progress. Not all innovations will be perfect immediately, but we will learn and develop as we go along. We need innovations to increase the efficiency of our existing economic infrastructure. Risks associated with engineered biological entities are controllable because we can engineer organisms more precisely due to new engineering technologies and containment.

A sentence from the Solazyme website can be understood as fitting within this worldview: “We are harnessing the power of an ancient resource, transforming microalgae into solutions for the world’s biggest problems.”

Vulnerable nature

The second worldview can be described as “vulnerable nature.” Adherents believe that nature is in a precarious balance. A push in any direction could be devastating. In the vulnerable nature worldview, only some technologies are beneficial, and only when bound by solid social and legal frameworks. Some other technologies pose great, potentially uncontrollable and irreversible risks. This group has a strong preference for local economies in which buyers and producers know one another and in which large companies are either absent or have only limited influence. Since there are presently large inequalities between economic players, open markets are considered detrimental because they preserve or increase existing inequalities. According to this worldview, we need to be constantly wary of imbalances in power. Therefore, this group embraces a collective decision-making process in which everyone has their say. Members include some NGOs and “dark-green” consumers and citizens.

A proponent of this quadrant might state the following in relation to the algae case:

We should fear the sugar economy in which everything is turned into a monoculture. In such an economy, feedstock becomes irrelevant and with that, natural diversity, local identities, and local needs become irrelevant. It will lead to the ultimate destruction of natural, economic, and social ecosystems. Also, large-scale companies patenting natural resources are wrong in themselves. We should only invest in small-scale, diverse solutions that prevent knowledge monopolies and the concentration of power over resources. With many so-called sustainable innovations, the economic benefits often accrue to a single actor while the promised ecological and health benefits remain unproven.

We need innovations, but only those that are truly sustainable and do not reinforce current inequalities and unsustainable practices. We need to consider the direction a particular innovation is taking: what other pathways are we neglecting by investing resources into a particular innovation? We should reject some disastrous innovations, such as synthetic biology, and focus instead on promising innovations, such as solar cells.

New tools for engineering biological entities might be risky because whole new organisms are created that we do not know the ecological effects of. To think that these organisms can be controlled is a dangerous illusion. We cannot control nature.

A quote from a Friends of the Earth report on synthetic biofuels echoes this worldview: “It is short-sighted to create new and unpredictable life forms that fit with our current infrastructure instead of investing in a new, clean, and sustainable infrastructure.” (2011, 20)

Controlled nature

A third worldview relevant here is the “controlled nature” perspective. This worldview lies between the two described above: its adherents see opportunities to exploit nature but also recognize limitations. These actors usually see the benefits of new technologies, but are also aware of the possible risks and, therefore, desire regulation. Market forces are considered beneficial if properly regulated to assure fair and open trading conditions. In this worldview, it is possible to arrive at global agreement on the values underlying regulatory frameworks for either technologies or markets. Members of this group include some NGOs, some companies, and government actors. A proponent of this worldview might state the following:

Whatever solution we can think of to support sustainability is worth considering. We should not condemn any innovations upfront. We should try to steer innovations in the right direction. The main criteria

for judging an innovation should be whether something is sustainable or not. Efficient use of resources can be expected to contribute to sustainability, be it large-scale or small-scale. We can (and we have) developed indicators to define sustainability. For a large part, we can already rely on LCAs to determine the sustainability of various feedstocks. Once we agree on all the indicators, such as sustainability criteria, we will know how to proceed. Therefore, everyone has to join in the conversation for establishing these indicators.

With any new innovation, risks are inevitable. As long as we take the right measures, we will be able to minimize risks to an acceptable level.

Ecover fits into this description because the company is critically assessing new technologies to increase sustainability on the basis of measurable indicators, while inviting a wide range of actors for input.

Capricious/irrelevant nature

Finally, the “capricious nature” world view group does not see the point of any form of control. This group is usually uninterested in new technology and its regulation unless it produces obvious benefits in the local area, local region, or individual daily life. Members of this group may be poorly educated people in Western societies or subsistence farmers in developing countries. These populations have no recognizable voice in this debate, although they do have a role. NGOs often voice concerns about the rights of these people. However, in terms of representation, this group is notably absent. This is not only true in our particular case but also common in societal debates on the bioeconomy. Despite the logistical complications of including them, this group is crucial for optimal social learning. While this group is often represented in natural resource management discussions, many bio-economic innovations sever the spatial connection between application and production of resources, thus marginalizing it.

These diverging worldviews cause ambiguity. If for instance one actor claims that engineered algae are a desirable technology because they might contribute to sustainability and we need to try every possible solution, and another says that engineered algae are undesirable because they present great ecological and economic risks, this can be taken as ambiguity about how we should evaluate the algae, i.e., moral uncertainty. Should we use an LCA to determine the sustainability of a specific innovation, or should we consider the institutional embedding and the macroeconomic effects of an innovation?

Moral ambiguity and social learning

With regard to social ambiguity, several outcomes of social learning processes are possible. Rölöing (2002) describes two different results in cases where diverging frameworks, what he calls “multiple cognition[s],”

are present. When actors engage in social learning, the result of multiple cognitions can be either “collective cognition” or “distributed cognition.”

Collective cognition refers to a shared understanding of crucial issues held by actors who agree on salient values and beliefs. It is most likely to arise within a homogenous set of actors, for instance, a group of employees of the same organization (Van Mierlo et al. 2010). When participants in a social learning exercise arrive at distributed cognition, they may not have a shared understanding of crucial issues, but they may have an overlapping understanding in a way that enables fruitful collaboration. Similarly, other authors (Van de Poel and Zwart 2010; Doorn 2010) inspired by Rawls (1993) speak of “overlapping consensus” based on the notion that most people share some basic assumptions that allow them to engage in reasonable debate.

For instance, Ecover holds the controlled nature worldview, whereas Solazyme maintains the nature as resource worldview. They each have a specific perspective on nature and the way humans should deal with nature, yet the algae oil is compatible with each of these perspectives and hence can be considered to represent a distributed cognition. Such a shared perspective on a specific item, arising from different frameworks, allows the actors involved to achieve an outcome that complies with each of their values, even though these are different. When actors arrive at collective or distributed cognition, uncertainty has been reduced.

To specify what distributed cognition might entail, we add the distinction in frameworks and perspectives to the notion of social learning as described above. In the case of distributed cognition, there may be agreement on the level of perspectives, i.e., on a necessary course of action to take in a specific case, but not on the underlying beliefs and values, i.e., the frameworks. However, if the social learning involves both first- and second-order learning, the shared perspective that the actors achieve can be justified within all of the relevant diverging frameworks (Van de Poel and Zwart, 2010). In this case, the actors will not have opted for a quick fix but will have co-constructed a mutually justifiable outcome.

For instance, all actors involved in the Ecover case agree, but for different reasons, that algae oil can be considered sustainable when taking the impact on local and other smallholders into account. Some actors want to produce

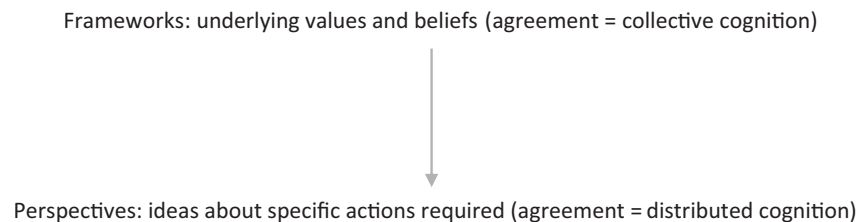


Figure 5.3 Frameworks/perspectives.

as sustainably as possible, others primarily want global economic equality, and others want to preserve a green image for their company. If they all agree on the need and focus on burdens and benefits for smallholders, they may enact institutional change in the form of social programs for smallholders in a way that respects all diverging frameworks.

In case of social learning processes where ambiguity plays a large role, outcomes of social learning processes can also amount to an identification of the kind of uncertainty that ought to be reduced. Is it productive to focus on impact uncertainty, or is it more useful to focus on moral uncertainty?

Overcoming moral ambiguity through a more deliberate experimental approach

Because the applications of new biobased technologies are surrounded by uncertainties, it is reasonable to plan them as social experiments designed for social learning (Gross 2015). Using the framework of experiments, two related strategies to facilitate learning emerge: namely scaling up and adaptability. If we accept that new innovations are accompanied by uncertain consequences, starting the experiment on a small scale to arrive at initial results and then proceeding to increasingly larger scales of application can be fruitful. Additionally, the design of the experiment requires that the technology be adaptable, thus able to accommodate learning effects (Van de Poel, 2017). These two strategies will be translated to the specific context of biobased innovations and the Ecover case specifically. We will first discuss what the possible outcomes of such an exercise might be.

Scaling up

The process of conscious scaling up—gradual increase in size of application—occurs in the development of most commercial technologies. Technologies are first developed or designed in a laboratory or research facility, then they are tested in a pilot plant or demonstration plant, which may take on different sizes. The final stage is commercialization. These different stages serve to test the technical behavior of an innovation. They might also be used to test out the *social* behavior of the innovation. Socially conscious scaling up is necessary because it allows for deliberate learning instead of ad hoc learning. Conscious scaling up embeds a process of social reflection through deliberations with a wide range of actors representing various perspectives of the existing uncertainties.

Social ambiguity can be reduced by assessing a technology at different scales of application. Some morally relevant effects only occur on a large scale, but application on a small scale can increase knowledge about possible consequences of a technology, for instance, by submitting it to initial safety tests. Additionally, at different scales of application of a technology, different social effects are likely to take place. This process reduces moral ambiguity.

When the application is still in the laboratory phase, the social process can be organized as a small sheltered space where a select group of people exchange views on the immature design. Sheltered exchanges remain confidential and discussions are not communicated to any nonparticipants.

If a social learning process is to result in a shared result, individuals may be required to adjust their positions or at least be willing to reflect on them. In a sheltered space, individuals may be more willing to abandon their entrenched positions because their social roles have become less solid. Especially professional social roles, such as CEO or spokesperson for an NGO, include considerable social expectations. The CEO is expected to safeguard the interests of the company s/he represents. The NGO spokesperson is expected to stand up for the values his/her constituency cherishes. The individual in such a role is released from his/her pressure somewhat once no longer in the public eye (Merton 1968, 428–29). This can impact his/her willingness to reflect on his/her own frame.

Additionally, social learning benefits from a sense of mutual dependency between the participants because they have a common stake (Röling 2002) and/or feel that they can accomplish more together than they would on their own (Süsskind, McKearnen, and Thomas-Lamar 1999; Loeber et al. 2007). Such mutual dependency does not always come about naturally. In this case, ETC is not dependent on Solazyme and Ecover. Rather, ETC is dependent on its constituency, and it must remain credible in the eyes of its supporters to flourish. If ETC abandons its highly critical position, even in a sheltered setting, this may jeopardize its position with its constituency. It may be possible, however, that when actors are involved in an early stage of technology development and can contribute to the outcome of the innovation trajectory, it results in mutual dependency.

Willingness to reflect on one's frame and adjust it might, in some instances, occur more easily in the public domain rather than in a small, closed setting that is sheltered from the public eye. A CEO might not be compelled to reflect on his/her frame solely because s/he interacts with someone who thinks differently, such as a representative of an NGO. But because s/he is expected to safeguard the interests of the company and these interests may be hurt by an NGO's allegations that its products are unsustainable, the CEO is publicly committed to reconsider his/her assumptions because the allegations of the NGO might affect his/her company's commercial prospects. Such an effect is most likely to surface at a large scale of application.

It may be questioned if effective social learning can begin only when an actual public controversy exists or if learning exercises before controversies arise can be constructive. Major controversies both motivate actors to learn and provide insight into relevant frameworks of reference and actors, thus yielding useful information. This information can help with selecting the right participants for more directed small scale exercises (Cuppen 2012). Additionally, once a controversy exists, it provides compelling reasons for actors to engage actively in social learning. If an exercise is planned before a

controversy has emerged, actors may not be aware of diverging perspectives. Social learning might seem artificial in such a case, as if the social learning exercise does not address any “real” problem (Bogner 2012).

However, to make social learning dependent on the emergence of social controversy is not productive. Firstly, once a technology moves beyond its development stage and is introduced in society, it may be more difficult to change its design based on new lessons learned. Secondly, controversies often incur great social costs, such as tainted reputations and hostility between actors. A more useful approach relies on the worldviews relevant in related innovations to start learning exercises at a relatively early stage of the innovation process. Relevant actors would therefore have a sense of potential disagreements before an actual controversy arises. These disagreements can then inform the further social learning process, for instance, by providing guidelines for the selection of participants.

Moreover, it is feasible to start social learning exercises not by focusing on a specific technology, but rather on the problem that the technology is supposed to solve. If the focus is on the technology, the outcome of the learning exercise usually answers the question, “is this an acceptable technology and if so, under what conditions?” Social learning exercises can be expected to reflect the plethora of perspectives better if they are tuned to a problem that might be solved by a specific technology, but for which alternative solutions exists. In the Ecover case for instance, the engineered algae provide a solution to the problem of unsustainable vegetable oils. When the debate immediately is limited to the algae, they will be the center of all discussions. This risks dividing the participants into supporters and critics. Instead, when the debate focuses on a common problem, actors are dependent on each other to reach a common solution, comparable to processes in natural resource management.

Adaptability

Adaptability is another required quality to ensure deliberate learning in a social experiment and to avoid large scale damage due to uncertainties. When an innovation as well as the associated trajectory are adaptable, the actual uptake of learning effects in the design become possible.

The future trajectory of the innovation was one major point of dispute in the Ecover case. Was it a step toward more sustainable applications, or would it form a hurdle for more sustainable applications because it redirected resources from more optimal solutions? If innovation trajectories can be flexible, the debate about the future need not strand in a deadlock. Instead, useful insights can emerge during the development of the trajectory, with all parties continuing to influence it.

Adaptability can mean different things for different innovations. Small-scale application can make adaptability possible because the application has not yet become socially entrenched and the costs of changing it are

relatively low. But for some innovations, any unwanted effects only become clear once the technology has reached a considerable scale of application. A Life Cycle Assessment becomes, for instance, more realistic once it is possible to assess the actual flow of energy and resources needed to produce a product. Should any unwanted effects occur, adaptability implies that the artifact can be altered in such a way that unwanted effects are either reduced or completely avoided. Biobased production processes can, for instance, be designed to allow for a variety of feedstocks so that the most sustainable one can be used.

The engineered algae Solazyme produced are adaptable to a point: they can be grown on different feedstocks. Switching to an alternative feedstock can be a partial accommodation to the concerns of the environmental organizations. Should representatives of Solazyme, after reflecting on their own framework and being convinced by the perspectives of other actors involved, choose to, they can consider switching to alternative feedstocks.

Actually, in the view of Ecover, a switch in feedstock is indeed the next logical step. More sustainable feedstocks might, for instance, consist of non-edible plant material, or any other local feedstock that has no use. In the view of Ecover, the Solazyme plant is basically a first step in a sustainable, decentralized natural oils production method. Because this argument accommodates the concerns of environmental organizations, it is a form of partial distributed cognition, a step further than multiple cognition.

However, it can be questioned whether Solazyme can indeed adapt to concerns over its feedstock. The technology to breakdown cellulosic biomass (the indigestible parts) has not matured. Additionally, the main concerns of the environmental organizations reside with the technology itself. Such criticism is, of course, much more difficult to accommodate, since this technology is their core business. The most they can do is to take all the other concerns of the critics into account and develop their business plan accordingly or switch to another product entirely. But they have shareholders behind their product who might not readily accept such a radical shift. Hence, options for adaptability are limited for Solazyme.

Adaptability can also imply resilience through diversity. For the effective management of natural resources as well as the realization of sustainability-related policy goals, authors in the fields of evolutionary economics, evolutionary policies, and adaptive management have proposed the condition of resilience through diversity (Rammel and van de Bergh 2003; Nill and Kemp 2009). Such resilience implies that in any given case, a wide range of solutions is applied to avoid a lock-in, i.e., the system at hand is made adaptable through diversity. If one solution does not work out, another is readily available. The various solutions are executed on a small scale in the first instance. This resilience enhances the ability to adapt to changing circumstances while it provides at the same time insights into the consequences of these varying solutions. These insights apply to the bioeconomy, as this is a field that deals with natural resources for which environmental circumstances

will often vary. It makes sense to always call on biobased companies to invest in multiple innovation trajectories. This may seem costly, but may create greater resilience once the innovations are brought onto the market.

Ecover has more options to adapt. The oil from engineered algae is one of many possible ingredients for their cleaning formula. It might arrive at complete distributed cognition by switching to coconut oil if it is consistent with both the Ecover values, such as always choosing the environmentally best available resource, and the values of environmental activists, namely that global economic equality deserves priority. If one of the actors would switch their frame to the frame of another worldview, then it would be collective, rather than distributed, cognition. For instance, if Ecover switched to coconut oil because it considered engineered algae a threat to global economic equality rather than because it has a better environmental performance, its frame will have changed.

These different approaches to adaptability echo the distinction made by Ansell and Bartenburg in this volume (Chapter 2) between evolutionary experimentation and generative experimentation, in which the first is experimentation through diversity and the second is experimentation by iteration.

From de facto to deliberate learning in the Ecover case

How might a more experimental approach have led to a constructive form of social learning that would have reduced moral ambiguity and hence produced a more societally acceptable innovation in the Ecover case?

First, relevant stakeholders should have been invited in an earlier phase of the innovation in a closed, sheltered setting. This might have helped both Ecover and Solazyme either to articulate their own perspective better or to make changes in the technical design or the economic infrastructure of the innovation. Changes in the economic infrastructure might, for instance, have included more economic benefits for local farmers. Technical changes might have involved smaller plants in a finer maze of distribution to avoid concentration of economic power. Learning should not stop at the sheltered level, however. Taking the innovation and the associated debate into the wider society can deliver additional insights.

Additionally, it might have compelled ETC to reconsider its evaluation of this technology because it was still in an early phase. ETC could have influenced the design of the technology rather than attack it as an outsider. Also, because the setting would have been closed and sheltered, ETC would have had more options to abandon its public role as a critical watchdog because they would have been less visible.

The actors might not have reached collective cognition, but might have reached distributed cognition resulting in a design of the technology that was more acceptable to all parties, even if this acceptance came from different frameworks. It might have been possible, for instance, to increase the benefits for local farmers and laborers at the sugar cane plantation. This adaptation

would be consistent with ETC's call for equal distribution of economic benefits as well as Solazyme's and Ecover's need for a sustainable product. Also, additional measures might have been taken to safeguard against environmental harm, if such measures suited all of the relevant frameworks of reference.

It needs to be noted, however, that this occurrence of distributed cognition remains speculative. There may be a higher chance of achieving distributed cognition in a sheltered setting. However, in this case with no mutual dependency, social learning might have been ineffective. Moreover, effective social learning need not necessarily result in distributed or collective cognition. If actors gain insight into the sources of uncertainty and adapt their strategies accordingly, even if only individually, this is a gain. Individuals who gain insight as well as organizations that reach a collective agreement both reduce moral ambiguity.

Additionally, actors involved in this debate should have prepared themselves for flexibility. Learning can only be effective if the learning effects can be accommodated. Thus, Ecover and Solazyme should invest in diversity and flexibility. Environmental organizations should also be open to reconsider their positions.

Deliberate learning in the bioeconomy

What lessons can be derived for the bioeconomy? We consider the Ecover case as a signal for salient uncertainties in the bioeconomy in general. Since the bioeconomy is based on natural resources, worldviews in which nature and technology are central elements can be expected to play a role in the uncertainties of the bioeconomy in general. In such cases, moral learning, i.e., learning about underlying values and worldviews motivating the perceptions of involved actors, is of paramount importance to reduce harm caused by uncertainties and lack of knowledge.

How can we organize such moral learning? We conclude with principles of governance for the introduction of new biobased applications. Firstly, to ensure that the debate results in effective social learning, it should on the one hand be broad enough to consider the challenges that motivated the development of the technologies at hand, while on the other hand the debate should include the particularities of those specific technologies. There should be opportunity to reflect on worldviews and frameworks, i.e., to go beyond particular perspectives on specific technologies. Additionally, there should be room to explicate the problems that specific technologies seek to address and the alternative solutions to these problems. This reflection should be linked to controlled scaling up of the innovation so that learning about impact, values, and institutional embedding can take place on different scales while minimizing harmful effects or the risk of a deadlock.

Worldviews and frameworks can be a sources of divergence, but they may also offer opportunities to form shared perspectives, i.e., distributed cognition. This is not to say that actors involved will necessarily change

their frame when they reflect on it, but, by gaining insights into the values, beliefs, and assumptions behind ambiguous uncertainties, these uncertainties can be reduced. If the values that actors hold dear are clear, then these values can be used as a foundation to draft socially acceptable solutions. If such solutions turn out to be unattainable, learning will still have taken place. Additionally, clarity about frameworks and worldviews can feed into the learning about institutional uncertainty and impact uncertainty, for instance, because it becomes clear what kind of impact uncertainty is deemed relevant to actors. Many participants in the field will not consider the uncertainties that they encounter in terms of frameworks. It may therefore be useful to have a facilitator present who can infuse the learning process with awareness about frameworks and worldviews.

Secondly, the designers and users of the technology at stake, in this case Ecover and Solazyme, should consider how they might change to an alternative design if they learn that their current design is not morally acceptable. In addition to looking for an optimal solution, they should also consider “ways out” of that solution if required, for instance, by thinking about how a production plant might be altered or how a line of supply might be changed. Both a substantive process of reflection and measure for adaptability might have repercussions for the way they engage shareholders, because there can be costs associated with being adaptable. These costs can, however, be considered an investment in long term success.

Thirdly, an important lesson from adaptive management applied to the bioeconomy as a whole is to involve in the process local actors most affected by the commercial activities associated with the natural resource. In the Ecover case, those determining the discussion were either representatives from global organizations, such as NGOs, or from commercial or academic organizations, such as Solazyme, Ecover, and the academics we spoke to. But the people actually making their living on sugar cane and coconut plantations did not have a voice in the discussion. This is a severe omission if we want to learn about sustainability and reduce uncertainty about how to use biomass, since the people closest to the resource at stake will have the best knowledge about it. Additionally, sustainability also involves social and economic welfare, and this is generally achieved best when those who need better living conditions are represented in relevant fora. In many discussions about a sustainable bioeconomy, this group is missing, which might seriously hamper the design of a sustainable bioeconomy in the long run.

Acknowledgments

We would like to thank Ibo van de Poel and Donna Mehos for their valuable feedback on this chapter. This chapter was written as part of the research program “New Technologies as Social Experiments,” which was supported by the Netherlands Organization for Scientific Research (NWO) under grant number 277-20-003.

Notes

- 1 This case has been extensively described elsewhere (Asveld and Stermerding 2016) and is based on interviews, literature, and observations at several meetings.
- 2 As voiced, for instance, by the Dutch Sustainable Growth Coalition in their latest report (2014).
- 3 We formulate fictitious statements for each of the groups to explicate their views better. These statements are derived from actual statements made in the debate on algae oil and from statements on website and reports. We did not do this for the capricious nature view because this group has no distinct voice in the debate.

References

- Armitage, Derek, Melissa Marschke, and Ryan Plummer. 2008. "Adaptive Co-Management and the Paradox Of Learning." *Global Environmental Change* 18 (1):86–98. doi:10.1016/j.gloenvcha.2007.07.002.
- Asveld, Lotte. 2008. Respect for Autonomy and Technological Risk. PhD Thesis. Delft University Press.
- Asveld, Lotte. 2016. "The Need for Governance by Experimentation: The Case of Biofuels". *Science and Engineering Ethics*, doi:10.1007/s11948-015-9729-y.
- Asveld, Lotte, Quirine van Est, and Dirk Stermerding. 2011. Getting to the core of the bio-economy: A perspective on the sustainable promise of biomass. The Hague: Rathenau Instituut.
- Asveld, Lotte, Jurgen Ganzevles, Patricia Osseweijer, and Laurens Landeweerd. 2014. *Naturally Sustainable? Societal Issues in the Transition to a Sustainable Bio-economy*. Delft: Delft University of Technology.
- Asveld, Lotte and Dirk Stermerding. 2016. *Algae Oil on Trial. Conflicting views of technology and nature*. The Hague: Rathenau Instituut.
- Bachmann, Reinhard, and Andrew C. Inkpen. 2011. "Understanding institutional-based trust building processes in inter-organizational relationships." *Organization Studies* 32.2: 281-301.
- Bogner, Alexander. 2012. "The Paradox of Participation Experiments." *Science, Technology & Human Values* 37 (5):506–27.
- Brom, Frans, Antoinette Thijssen, Gaston Dorren, and Dieter Verhue (red.). 2011. Beleving van technologie en wetenschap - Een segmentatieonderzoek. Den Haag: Rathenau Instituut.
- Cuppen, Eefje. 2012. "Diversity and Constructive Conflict in Stakeholder Dialogue: Considerations for Design and Methods." *Policy Sciences* 45 (1):23–46. doi:10.1007/s11077-011-9141-7.
- De Vries, Bert J. M., and Arthur C. Petersen. 2009. "Conceptualizing Sustainable Development: An Assessment Methodology Connecting Values, Knowledge, Worldviews and Scenarios." *Ecological Economics* 68 (4):1006–1019. doi:http://dx.doi.org/10.1016/j.ecolecon.2008.11.015.
- Doorn, Neelke. 2010. "A Procedural Approach to Distributing Responsibilities in R&D Networks." *Poiesis & Praxis* 7 (3):169–88.
- Douglas, Mary, and Aaron Wildavsky. 1982. *Risk and Culture: An Essay on the Selection of Technical and Environmental Dangers*. Berkeley: University of California Press.

- Dryzek, John S. 2013. *The Politics of the Earth: Environmental Discourses*. Oxford: Oxford University Press.
- European Commission. 2012. *Innovating for Sustainable Growth: A Bioeconomy for Europe*. Brussels: European Commission
- Flipse, Steven M. 2014 "Environmental Life Cycle Assessments as Decision Support Systems within Research and Development Processes: Solutions or Confusions for Responsible Innovation?." *International Journal of Business and Management* 9 (12):210–20.
- Friends of the Earth US.2010. *Synthetic solutions to the Climate Crisis*.
- Glasser, Harold. 2007. "Minding the Gap: The Role of Social Learning in Linking Our Stated Desire for a More Sustainable World to Our Everyday Actions and Policies." *Social Learning: Toward a More Sustainable World*. 35–61. Wageningen, The Netherlands: Wageningen Academic Publishers.
- Grin, John, and Henk Van de Graaf. 1996. "Technology Assessment as Learning." *Science, technology & human values* 21 (1):72–99.
- Gross, Matthias. 2015. "Give Me an Experiment and I Will Raise a Laboratory." *Science, Technology & Human Values*. doi:10.1177/0162243915617005.
- Hedlund-De Witt, A. 2013. "Worldviews and the Transformation to Sustainable Societies." PhD thesis, Vrije University.
- Kahan, Dan M. 2012. "Cultural Cognition as a Conception of the Cultural Theory of Risk." In *Handbook of Risk Theory*, edited by Sabine Roeser, Rafaela Hillerbrand, Per Sandin and Martin Peterson, 725–759. Springer Netherlands.
- Kupper, Frank, Linda Krijgsman, Henriette Bout, and Tjard de Cock Buning. 2007. "The Value Lab: Exploring Moral Frameworks in the Deliberation of Values in the Animal Biotechnology Debate." *Science and Public Policy* 34 (9):657–70. doi:10.3152/030234207x264944.
- Loeber, Anne, Barbara van Mierlo, John Grin, and Cees Leeuwis. 2007. "The Practical Value of Theory: Conceptualising Learning in the Pursuit of a Sustainable Development." In *Social learning towards a sustainable world*. 83–98. Wageningen, The Netherlands: Wageningen Academic Publishers.
- Merton, Robert King. 1968. *Social theory and social structure*. New York: Simon and Schuster.
- Nill, Jan, and René Kemp. 2009. "Evolutionary Approaches for Sustainable Innovation Policies: From Niche to Paradigm?" *Research policy* 38 (4):668–80.
- OECD. 2009. The Bioeconomy to 2030: Designing a Policy Agenda. In *International Futures*: OECD.
- Rammel, Christian, and Jeroen CJM van den Bergh. 2003. "Evolutionary Policies for Sustainable Development: Adaptive Flexibility and Risk Minimising." *Ecological economics* 47 (2):121–33.
- Rawls, John. 1993. The Domain of the Political and Overlapping Consensus. *The idea of democracy*, 246.
- Roling, Niels. 2002. "Beyond the Aggregation of Individual Preferences." In *Wheelbarrows full of frogs: Social learning in rural resource management*, edited by Cees Leeuwis and Rhiannon Pyburn, 25–48. Assen: Van Gorcum.
- Schon, Donald A..1983.. *The reflective practitioner*. New York: Basic Books
- Schön, Donald A., and Martin Rein. 1994. *Frame Reflection: Toward the Resolution of Intractable Policy Controversies*: New York: Basic Books.

- Susskind, Lawrence E., Sarah McKearnen, and Jennifer Thomas-Lamar. 1999. *The Consensus Building Handbook: A Comprehensive Guide to Reaching Agreement*. London: Sage Publications.
- Thompson, Michael, Richard Ellis, and Aaron Wildavsky. 1990. *Cultural Theory*. Boulder, CO: Westview Press.
- Van Mierlo, Barbara, Cees Leeuwis, Ruud Smits, and Rosalinde Klein Woolthuis. 2010. "Learning towards System Innovation: Evaluating a Systemic Instrument." *Technological Forecasting and Social Change* 77 (2):318–34. doi:<http://dx.doi.org/10.1016/j.techfore.2009.08.004>.
- Van de Poel, Ibo, and Sjoerd D. Zwart. 2010. "Reflective Equilibrium in R & D Networks." *Science, technology, & human values* 35 (2):174–99.
- Van de Poel, Ibo. 2017. "Society as a Laboratory to Experiment with New Technologies" In *Embedding New Technologies into Society: A Regulatory, Ethical and Societal Perspective*, edited by Diana M. Bowman, Elen Stokes and Arie Rip, 61–87. Singapore: Pan Stanford Publishing.

6 Cognitive enhancement

A social experiment with technology

Nicole A. Vincent and Emma A. Jane¹

Introduction

People from different walks of life—for example, video gamers, students, academics, entrepreneurs, classical musicians, and public servants—are increasingly experimenting on themselves with putative cognitive enhancement (CE) technologies. The lure includes promise of superior memory, focus, reflexes, calmness, clarity of thought, problem-solving ability, mental stamina, and ability to function well with little sleep. Much of this experimentation, though, involves risky repurposing of medications and devices normally used to treat mental disorders. Examples include pharmaceuticals such as Ritalin (methylphenidate) and Adderall (amphetamine salts), and “electroceuticals” such as transcranial direct current stimulation (tDCS) devices. Yet despite the risks, these self-experimenters seldom seek out scientific and medical advice and supervision.

Against this background, in March 2015, the Presidential Commission for the Study of Bioethical Issues (henceforth Presidential Commission) published its report *Volume 2, Gray Matters: Topics at the Intersection of Neuroscience, Ethics, and Society*. Among its many recommendations to President Barack Obama were the following that pertain specifically to emerging CE technologies:

Funders should support research on the prevalence, benefits, and risks of novel neural modifiers to guide the ethical use of interventions to augment or enhance neural function. If safe and effective novel forms of cognitive enhancement become available, they will present an opportunity to insist on a distribution that is fair and just.... Limiting access to effective enhancement interventions to those who already enjoy greater access to other social goods would be unjust. It also might deprive society of other benefits of more widespread enhancement that increase as more individuals have access to the intervention. In addition, more widespread enhancement might help to close some gaps in opportunity that are related to neural function, such as educational attainment or employment.

(Presidential Commission 2015, 4)

Although we applaud the Presidential Commission's concern for equality as well as safety and effectiveness, in our view, this recommendation and its underlying reasoning are so problematic that we will eventually in this chapter describe their recommendation as a reckless form of social experimentation. The narrow conceptions of "safety" and "effectiveness" in play mean that serious social and normative hazards are overlooked. Even when these hazards are recognized, their importance is played down because: they are temporally distant; it is uncertain whether they will manifest and how people will feel about them; and because navigating around them would allegedly require an abandonment of political neutrality and diminish our liberty. Consequently, important design and regulatory decisions are not being made. Perhaps most worryingly, safeguarding equality of access to CE may, paradoxically, actually increase the likelihood that the social and moral hazards that trouble us will manifest. Given that the Presidential Commission overlooks these hazards, our case will be that the above recommendation constitutes a reckless *moral* and *social* experiment. And although we think that social experimentation with technology is ultimately indispensable, it can and should be conducted in a more responsible manner.

In what follows, we begin by providing some examples of cognitive enhancers and make some observations that provide the conceptual background to our discussion. We then reconstruct what we take to be the Presidential Commission's reasoning—i.e., why it has such enthusiasm for CE—and explain what this enthusiasm overlooks. In the last section, we propose our own seven-point methodology for thinking about the design and regulation of emerging technologies, which frames the CE topic as one that inevitably involves social experimentation with an emerging technology, though not necessarily a neuro-technology.

Cognitive enhancement: examples

"Cognitive enhancer" is a broad term for any of a number of different techniques aimed at improving the mind, and the phrase "neural modifiers" in the Presidential Commission's recommendation could technically refer to anything that effects changes in the brain. At the *familiar* end of the spectrum, this includes you reading these words, which presumably influence your thoughts, presumably with underlying changes in your brain. At the *novel* end of this spectrum, "neural modifiers" can also refer to means of modifying the brain more directly—e.g., through surgical means.² In between these two extremes lies a range of more and less novel techniques.

For instance, although we do not usually think of them as such, education, sleep, power naps, and the caffeine present in tea, coffee, and a range of commercially produced beverages are all familiar forms of CE. Through education, we improve our knowledge and skills, including our discipline and ability to analyze novel problems and to systematically seek out solutions. Sleep and power naps restore wakefulness, attention, and other mental

functions that gradually degrade the longer a person stays awake, and caffeine's ability to restore wakefulness and attention is also due to its action within the brain.

However, in the above context, the Presidential Commission uses the term “novel neural modifiers” in reference to a range of psychotropic medications and transcranial electrical and magnetic stimulation devices for modifying the brain that have attracted much scholarly (e.g., Farah et al. 2004; Glannon 2008; Greely et al. 2008; Coffman, Clark, and Parasuraman 2014; Meinzer et al. 2014; Sahakian et al. 2015) and public (Wise 2013; Ojiaku 2015; Thomson 2015; Schwarz 2015) interest, and which are thought to improve people's minds in various respects. While, on the one hand, they are similar to brain surgery in that they also directly modify the brain, on the other hand, they are not quite as invasive.³

Common examples of pharmacological CEs, especially those that attract media attention, are central nervous system stimulants like Adderall, Ritalin, and Provigil, which are normally prescribed for the treatment of conditions such as narcolepsy, attention deficit and hyperactivity disorder, and shift work sleep disorder. However, when taken by healthy individuals these medications reportedly improve memory, focus, reflexes, clarity of thought, task enjoyment, and ability to function well with little sleep. Another example of a pharmacological CE is Donepezil, a medication normally prescribed for treatment of Alzheimer's disease, dementia, and mild cognitive impairment. Although Donepezil does not slow down the progress of Alzheimer's disease, dementia, and mild cognitive impairment, it does produce modest cognitive benefits in people with these disorders, and these mild benefits are also what healthy people who take Donepezil for CE hope to obtain.

Transcranial electrical and magnetic stimulators are two examples of non-pharmacological CE techniques that have also been investigated and that comprise another important class of techniques and technologies that the Presidential Commission had in mind by its reference to “novel neural modifiers” (Coffman, Clark, and Parasuraman 2014; Meinzer et al. 2014). Transcranial magnetic stimulation is a technique in which a powerful electromagnetic field is generated and focused (at varying intensities, durations, and pulse rates) on a selected region of the brain to induce temporarily increased or decreased neural activation in that region. Transcranial direct current stimulation involves the placement of electrodes on selected areas of the scalp and then passing a very small electrical current through those electrodes—typically in the range of 0.5–2.0 milliamps, so small that these devices can draw their power from conventional nine-volt batteries—which passes through the scalp and creates an electrical charge in the underlying brain tissue. Contemporary “electroceuticals” are largely unregulated and under-tested, yet they are manufactured and marketed alongside promises relating to curing depression and chronic pain, as well as improving brain plasticity, hand-eye coordination, sports performance, memorization, and learning.

Attractions of cognitive enhancement

In the cited passage from the Presidential Commission's (2015) report, medically safe and effective CE is claimed to have the potential to improve our lives by delivering (a) *universally valuable* things like knowledge and freedom, (b) *social goods* like improved science, technology, and medicine from which everyone can benefit, and (c) *individual goods* that differ from person to person by providing mental resources that can help secure them.

To see how medically safe and effective CE might be thought to deliver these things, consider again what people cite as some of CE's attractions: improved wakefulness, attention, clarity of thought, creativity, problem-solving ability, mental stamina, memory, reflexes, motivation, and a range of executive functions such as planning, self-insight, self-monitoring, and the ability to direct one's own attention (Vincent and Jane 2014).⁴ The social goods that CE may be thought to promote are such things as advances in science, technology, medicine, and other fields that a society full of smarter, more productive, and more resilient humans has the potential to create. Through advances in science and technology, we might understand better the causes of problems that afflict humanity including infectious disease, acquired medical disorders, and environmental degradation. With improved understanding of their causes and appropriate technical breakthroughs, we might also develop better technologies and medical treatments to address these problems. Who knows, we might even develop better—that is, even more effective and even more medically safe—CE medications and devices!

We might also suppose that if everyone has greater mental resources—that is, if all people are smarter, more productive, and have greater mental stamina—then everyone will be in a better position to pursue whatever other things are valuable to them because everyone will have better mental resources (Bostrom and Roache 2008, 137–38). For example, if you were more productive, perhaps you could get your work done sooner and free up time to engage in other pursuits you find valuable—whatever things make for a better life *in your view*—like spending time with family or friends, volunteering for charitable organizations, developing your talents and hobbies, learning new things, or just resting and relaxing. Alternatively, perhaps some people might prefer to cash-in on their greater productivity and stamina, and work and produce more, in order to get a better education or a more rewarding or better paying job. Some people's conception of a good life might even be such—one of achievement, attainment, or excellence—that they would work more not just as a means to obtain other ends (e.g., better job or higher income), but as an end in itself.

As things stand right now, currently available novel neural modifiers are neither sufficiently effective nor sufficiently medically safe—or at least there is insufficient scientific evidence to support the positive claim that they are effective and safe (Outram and Stewart 2012)—enough to warrant

recommending their use for CE purposes. However, this is precisely why, in the Presidential Commission's view, research funding should be allocated to develop better CEs—ones that are medically safe and truly effective—and then to ensure that everyone has equal access to them.

Problems with cognitive enhancement

Naturally, concerns about CE medications and devices might linger. For instance, perhaps some medical side effects may go undetected for a long time and only manifest themselves after years of wide-scale use by the public, i.e., maybe CEs won't be as *medically safe* as they initially seem. Or perhaps the actual benefits of CE will turn out to be so mild that investing government funding to develop them and then to make them widely available will not yield a good return on society's investment, i.e., maybe they won't be as *effective* as we hope. Or maybe the same funding could be put to better use elsewhere, e.g., to develop cures for diseases rather than further improving the lives of people who are well-off, or by giving prioritized access to CE to people with learning disabilities and to the cognitively under-privileged rather than giving everyone *equal access* to it? However, although such concerns (and potentially many others) are important, in this section we would like to set them aside in order to show that even if they prove to be unfounded—that is, even in the Presidential Commission's presumed *ideal* scenario where we have medically safe, highly effective, and equally distributed CEs—important problems would still remain.

We have argued elsewhere (Vincent and Jane 2014; Jane and Vincent 2015) that medically safe and effective CEs may still have unexpected and unintended effects on society—what we referred to as “social side effects”, and what Tsjalling Swierstra (2015) calls “soft effects”—that may, quite apart from any medical side effects, still adversely impact our quality of life. Specifically, we believe that there is a very real threat that in a competitive society like ours, wide-scale use of certain kinds of CE—in particular, of CEs that produce precisely the kinds of effects that CE enthusiasts most often seek, namely increased productivity, stamina, and motivation—may eventually result in our society becoming even more oppressively competitive and demanding. That is, our concern is that use of such CE in competitive contexts will produce a society in which everyone is expected to be more productive, to work longer, possibly even to work in more narrowly-defined and mundane jobs, and consequently to be less rather than more able to do the broad range of activities (which differ from individual to individual) that contribute to a good life.

To avoid the charge of mere scaremongering, though, in this section, we shall spell out our thinking in greater detail to show precisely what our concern amounts to, what would be bad about the state of affairs that we worry about, and through what mechanisms we believe such a state of affairs would come about.

Competition and mutual social coercion absorb the benefits of CE and make life worse

In this subsection, we argue that even if the sorts of CE that people seem to want could deliver universally valued and social goods, nobody would individually have any control over which of these things they actually get, and in competitive societies, these things would come at the price of sacrificing (not promoting) our ability to live life according to our own conception of the good. To make this argument, we will employ a thought experiment to anticipate some of the possible social side effects of CE. Although the reliability of our narrative approach to anticipating social side effects has been debated (see, for example, van de Poel's contribution to this collection), we agree with Swierstra who argues that narratives are better suited to the task of anticipating potential soft impacts of emerging technologies on account that, in contrast with hard impacts, they are usually qualitative (not quantifiable), ambiguous (not indisputably harmful), and co-produced by humans (not straight-forward effects of technology) (2015, 11–17).

To see how the state of affairs that troubles us could arise, imagine if there were a pill that anyone could take that would be perfectly medically safe, that was provided free of charge to anyone who wanted it, and that would make those who used it more productive. For simplicity's sake, suppose that anyone who took it would experience the same quantitative increment⁵ in productivity as anyone else who took it—call it the *work faster pill*. People would still get tired after the same number of hours spent working, and they would still get bored of doing boring tasks after the same number of hours, but in that same amount of time they would produce more than what they would have produced if they had not used the work faster pill.

Presumably, at first some people would be skeptical, and many might prefer to abstain from taking it. After all, there is a long history of pharmaceuticals being marketed that are claimed to be medically safe but which subsequently turn out to be dangerous.⁶ But at least some people—maybe early adopters of technology, perhaps one of us, let's say, Nicole—would give it a try. At work, Nicole's productivity would soar. She would now write and publish significantly more papers, grade student assignments and tests more quickly, and with more time on her hands, she could sit on more university committees and clock up extra points for service to list on her professional CV. Consequently, like in any other competitive workplace situation, Nicole would likely get a competitive advantage over her colleagues and commensurate rewards, whether in salary, promotion, invitations to deliver talks, or other ways in which good performance is recognized and rewarded in academia.

How do we think others would be likely to respond? Well, whether from a desire to make a greater contribution to knowledge, to become more accomplished, to advance their own careers, or just to ameliorate that they would no longer stand out as much as they once did given the new higher standard

that early adopters like Nicole were setting, other academics—let's say, Emma—would probably also consider using the work faster pill. However, with increasingly more academics using work faster pills, what constitutes the *normal* level of performance would itself gradually shift upwards until yet more people who had originally resisted the urge to use work faster pills would now start feeling pressure to use them.

With increasingly more people using work faster pills in academia, whatever positional advantage Nicole and Emma may initially have had when they started using those pills would gradually be eroded, and at the same time the competitive pressure on those who still were not using them (perhaps you, dear reader) would gradually mount. With increasingly greater pressure to use them, more people would buckle, and by the time that most people were using work faster pills, Nicole and Emma (and anyone else who joined them along the way) would have returned to whatever relative position performance-wise they occupied prior to this miracle pill becoming available. Every user of the work faster pill *would* get more work done in a day than they once did, *but so would their colleagues*. If what Nicole and Emma had hoped for was that they could get their work done sooner, and then have time to take up sport, a hobby, or start a family, then these hopes would be dashed unless everyone else also decided not to use the productivity advantage of the work faster pill for competitive purposes. Same too if their hope had been to rise in their respective careers *and stay there*.

Notice, also, that by this stage everyone would have surreptitiously signed themselves on into a new collective pact to keep using work faster pills just to stay in their professional positions, because what constitutes normal functioning would have shifted upwards in line with our work faster pill-enabled greater abilities. The freedom *not* to use these pills and to *remain* in our occupations would no longer be there, because the shape of those occupations—i.e., what it takes to be a university professor, for instance—would have been altered by the practice of using them in those occupations, with the end result that now people in those occupations would *need* to use them to be competitive in those occupations (as would new entrants into those occupations).

Similarly, a pill like modafinil that also made work more enjoyable (or perhaps more palatable, or just less unpalatable) so we could push through the boredom barrier associated with mundane tasks⁷—call it the *increased motivation pill*—would also in a competitive environment be used by some people to gain a competitive edge by being prepared to engage in more mundane and boring tasks.⁸ There might even be reason to suppose that in a more demanding and competitive environment, people's jobs would become more specialized, and that with increasing specialization jobs might indeed become more narrowly defined, less varied, and more...boring.

A similar narrative could be developed for a pill that, perhaps like Adderall or Ritalin, enabled its users to remain wakeful for longer—call it the *work longer pill*. Once some people start using the work longer pill in

a competitive environment to gain advantage over others, soon others will follow in their stead, and the more people use it to work longer hours, the more pressure will mount on others to do likewise or fall behind the others.

The obvious objection to the above discussion is to ask why any of this should be of concern to anyone? After all, we are assuming that these pills would have no medical side effects and many benefits, so why worry?

One reason to fret is simply that the loss of freedom to say “no” to using even a medically safe and effective CE *is* a loss. Another reason for complaint about losing the freedom to say “no” to using medically safe and effective CE is that presumably nobody in this scenario actually intended to make their entire profession more competitive. What they probably wanted was to make their lives easier, or to get an *individual* competitive edge. Some people might also have wanted to make more breakthroughs, because that is their conception of a good life. What people would actually have brought about, however, is a more demanding and competitive profession for themselves and others because of their collective actions. Finding ourselves in this sort of society, especially when that is distinctly the opposite of what we wanted, an anticipated and unwelcome social side effect seems to be a legitimate ground for complaint.

The precise effects of CE matter

In our view, a society that placed even greater demands on our productivity, on our time, and on our ability to engage in mundane tasks would not be a better society. In the scenario that we described above, we would all end up running faster, for longer, and keeping pace with a more monotonous beat. We do not deny that the use of CE with such qualitative effects—that enabled us to work faster and longer with increased motivation, and especially if our example is broadened to include more than just academic philosophers but other jobs too—might deliver a range of social benefits and universally valued advantages. Rather, our worry is that these other benefits would be purchased at the price of a loss of freedom and of individual goods. We would have lost the ability to say “no” to using medically safe and effective CEs unless we were prepared to suffer setbacks to our welfare due to falling behind.

Individually chosen goods would not only fail to be advanced by providing everyone with equal access to CEs, but our ability to obtain and enjoy individual goods may even end up being marginalized and eroded depending on precisely what qualitative effect we imagine CEs will produce. Consequently, to ascertain whether we have reason to support creating CEs and making them available to everyone on demand, we need more information about precisely what qualitative effects a given CE would produce.

This is important because in a competitive context where we could potentially use various advantages to compete with one another, we must be very careful not to design or make available CEs that make it possible for *some* people to sacrifice things—for example, spare time or a meaningful

job—that we actually value so much that we would not want those things to be sacrificed by others and used as pawns in competitions with us. CE technology that enables people to force each other to sacrifice in the course of free market competition the very things that we need to lead good lives is not, in our view, a desirable CE technology. A pharmaceutical CE or CE device that would increase productivity, lengthen the work day, and make mundane tasks feel less mundane (i.e., a pill with precisely the qualitative effects that are celebrated by CE enthusiasts) should in our view either not be created, or its use should be tightly regulated. It certainly should not be made available to everyone on demand.

What went wrong with the presidential commission's reasoning?

In this section, we will argue that what went wrong with the Presidential Commission's reasoning is that there are significant oversights in how they understood the notions of safety, effectiveness, and equality (and hint at what we think led to these oversights), and that their conception of regulation was insufficiently nuanced.

Different kinds of safety, effectiveness, and equality

Notice that the Presidential Commission's focus on safety, effectiveness, and equality has a very distinctive flavor. Namely, *safety* is framed as a *medical* issue; *effectiveness* is rendered as *competitiveness in a market economy*; and *equality* is portrayed as *strict equality in the distribution of a resource*. However, these ways of understanding these notions are deeply problematic.

In regards to safety, unexpected and possibly unwelcome consequences *on society*—i.e., that the workplace may become even more competitive and demanding, or what Swierstra calls “soft impacts”⁹—are simply not considered. But if we are concerned about ensuring that a new technology like CE improves our lives rather than makes them worse, we see no reason, in principle, why potential social side effects should have been disregarded.

In regards to effectiveness, consider how improved productivity, stamina, and motivation compare to alternative effects that could potentially be engineered into CE pills or devices—e.g., fostering greater connection with your fellow citizens, more empathy for all creatures, or deeper emotional connection with those around you. When these two sets of potential effects are contrasted, what stands out for us is that the effectiveness referred to by the Presidential Commission is, at least implicitly, effectiveness *as a worker in a competitive market*, not effectiveness along a number of other dimensions on which humans could potentially be made better or worse. However, we think it unhelpful to focus solely on—and to attempt to foster in people—such commercially useful forms of effectiveness, especially

now that we realize that this kind of effectiveness is likely to make competition in the marketplace fiercer and our lives harder. Why should we think that making ourselves better as commercial gladiators would necessarily improve our lives? In short, the Presidential Commission's notion of effectiveness expressed a very specific and narrow substantive comprehensive conception of the good—one which, even if only because of its narrowness, is not obviously conducive to improving everyone's life as judged by their own substantive comprehensive conception of the good.

Finally, in regards to equality, we also think it plausible that if what we want is equality in a more nuanced and satisfying sense, then what we might have reason to strive for vis à vis how CE is distributed is not *strict* equality—i.e., that everyone gets the same level of access—but that some people (for instance, the cognitively under-privileged) get greater access to CEs than others (for instance, those who already have many social and intellectual advantages). The topic of equality is far too complex for us to do justice to it here, but that is precisely the point that we are driving at. Namely, that if the Presidential Commission's concern was indeed with improving equality in society, then simply giving everyone equal access to CE is not necessarily the best way to do it. In fact, giving everyone unbridled access to CE in a competitive environment is precisely—and somewhat paradoxically, in that the effect produced is not what was intended by the Presidential Commission—what is likely to lead to the fiercer competition that we described above.

Factors that explain the oversights described above

In this subsection, we suggest that a number of factors are likely to have influenced how the Presidential Commission understood the notions of safety, effectiveness, and equality, and thus why it overlooked the serious threat that we described above. Two of these factors relate to how this topic has been framed, namely as a debate about *state regulation of medical devices*, and the other three factors relate to very real but (as we shall argue in this chapter's final section) not intractable challenges with predicting, evaluating, and devising measures to navigate around potential social side effects (what we shall call “goal-setting” henceforth), which seem politically, normatively, and computationally intractable when viewed from within the incumbent framing.

The current public and scholarly debate about cognitive enhancement has been framed as a *bioethics and neuroethics* topic concerned with *how the state should regulate* the associated pharmaceuticals and devices. However, this framing is unhelpful for a number of reasons. Firstly, the *bioethics and neuroethics* (i.e., medical) framing predisposes us to overlook potential social side effects simply because they do not necessarily have a “medical”¹⁰ dimension. True, medicines and medical devices are involved, but this does

not make CE an exclusively medical topic. Medications and medical devices can, after all, have more than medical effects. There are more ways for our lives to go badly than by developing aches and pains, high blood pressure, blurry vision, an upset stomach, diarrhea, restlessness, insomnia, fatigue, drowsiness, or any of a number of other typical examples of side effects that might be classified as medical. We could also, for instance, end up living in a more demanding society with fewer leisure hours to pursue other activities that contribute to a good life.

What's worse, we might not even notice that this new state of affairs, where we spend even more hours at work in even more specialized and narrowly focused jobs, is bad. We would now enjoy tasks that previously seemed like drudgery, partly because our values would have shifted imperceptibly and insidiously alongside gradually changing social norms (Jonas 1973; Swierstra 2015), and partly because the medications or devices may have transformed our evaluations (e.g., see note 8 regarding modafinil's effects). Within a bioethics or neuroethics framing, once we hear that a given pharmaceutical or medical device has no medical side effects, it is too easy to stop asking questions about what other non-medical side effects that medication or device might have—e.g., on our values or on society—and how such non-medical side effects might adversely influence our ability to live what we currently consider to be a good life.

To be clear, we are not presupposing that whatever unexpected social changes occur will necessarily be bad. We cannot say this any more than anyone else can say that the social changes will clearly be good. This is, indeed, our very point. A large part of the problem here is that social changes *are* very difficult to predict and evaluate. But this should not mean that we allow the invisible hand of competition to determine what social changes shall come about and what society will come to look like. Still, it is helpful to reflect on the nature of the predictive and evaluative challenges.

On the prediction side, when we expand the scope of our concern from caring about only medical side effects, we are potentially faced with a plethora of other possible effects that demand scrutiny. Furthermore, each of these potentially large number of effects may be produced by a potentially large number of factors, including the choices of human agents. And given the role that human psychology plays in generating social side effects, the vagaries of human psychology will make the task of predicting social side effects more complex than predicting medical side effects of a medical device because there are fewer things to predict and fewer factors to consider. Lastly, whereas we already have methodologies for attempting to predict medical side effects—e.g., laboratory studies, clinical trials, and analysis of epidemiological data, each with their own ways of controlling for numerous variables—there simply is as yet no recognized methodology for predicting social side effects.¹¹ This lack of existing methodology makes the predictive

challenge vis à vis social side effects particularly steep. Such predictive challenges help explain why social side effects are not usually taken into account in the CE debate.

On the evaluative side, if we cannot predict what social side effects might manifest, how can we begin to think about evaluating them? (Swierstra 2015, 6) Worse still, even if we could predict some social side effects, their evaluation would still face steep challenges. For one, we may simply lack the imagination to decide whether living in that kind of society would be good or bad. In the discussion above, we boldly asserted our own substantive comprehensive view when we stated that living in a more competitive and demanding society would not be pleasant. However, we recognize that this is ultimately our own current substantive comprehensive view, and our intention is not to substitute our current evaluations for anyone else's or to suggest that our evaluations might not change over time. Furthermore, perhaps someone who grew up in an increasingly competitive and demanding society would adjust to that level of competition and those demands because they would not have known a different way of life, and so perhaps they would not care as we do for whether they had less time to spend with friends and family.

This possibility of being transformed as evaluators presents a very serious challenge to evaluating social side effects, especially by comparison with medical side effects. For one, while it is probably universally viewed as bad to develop a festering wound or to break a bone, views among people living now and those who have lived in the past probably differ greatly on what kind of society would be good.¹² Furthermore, for us, the scenario described above did not sound appealing. From our current standpoint, we have no reason to endorse becoming like that—that is, people who would no longer care about the things that we currently care about, like having time to spend with family and friends, to pursue our own individually-chosen ends, and to have a meaningful occupation judged by our current standard of what is meaningful.¹³

Our suggestion, though, is that although the evaluative challenge is indeed also steep, just as the predictive challenge, we nevertheless have solid reasons to express our views about what kind of society we would like to live in, and what kinds of people we would like to become, not to treat these difficulties as reasons to remain neutral in regards to such important matters. From our current evaluative standpoint—one which we have no reason to ignore—CE medications and devices that would deliver a society that we now find deeply unappealing, and that would warp our values such that we no longer even find that society unappealing, are not ones that we should view as attractive.

A flow-on challenge is that if we can neither predict nor evaluate the potential social side effects of CE medications and devices, then this will also present challenges to deciding how CE medications and devices

should be regulated. In part, this is because choosing to regulate a given technology in one way rather than another in the absence of data about what effects that technology may produce or how those effects will be viewed may simply rob us of the opportunity to discover useful things (e.g., cures for diseases or useful new gadgets) simply because we are afraid of possible but ill-defined social side effects. Another flow-on problem appears to be more political in nature. Namely, there are good reasons to suppose that government regulations should be guided by accurate data about effects of different technologies, and how citizens might view those effects (van de Poel 2013, 353). If we suppose that governments represent their citizens, then citizens' informed preferences should surely play an important role in guiding the state's regulatory decisions when those decisions will impact the citizens' lives. But if data about what will likely happen and whether it will be viewed as good or bad is unavailable (because of the foregoing predictive and evaluative challenges), then a government that attempted to regulate CE to produce a particular social outcome (e.g., not allow CE of the sort that we discussed above in order to ensure that society does not become even more competitive and demanding) could seem to verge on politically non-neutral social engineering.

For such reasons, we think that the Presidential Commission opted for what it viewed as a light-handed regulatory regime. It proposed equal access to medically safe and effective CE technologies for everyone, without taking a stance on what social effects CE should or should not produce. These matters would be left to scientists and consumer choice.

In a recent opinion piece in *The Boston Globe*, Steven Pinker advances a more polemical though, in our view, similar line of argument for decreased regulation (2015). In it, he argues that the primary moral goal for today's bioethics can be summarized in a single sentence: "Get out of the way." His position is that all that is achieved by "paus[ing to] consider the long-term implications of research before... rush[ing] headlong into changing the human condition" is that progress is slowed down. In his view, biomedical research is too valuable to be thwarted by ill-defined concerns about such things as "dignity," "sacredness," and "social justice" by speculation about possible but temporally distant future harms, or by analogy-driven scaremongering with dystopic "Brave New World" and "Gattaca" scenarios. While he agrees that "individuals must be protected from identifiable harm," in his view, "we already have ample safeguards for the safety and informed consent of patients and research subjects [and] slowing down research has a massive human cost." Furthermore, in his view, "technological prediction beyond a horizon of a few years is so futile that any policy based on it is almost certain to do more harm than good." And in any case, he points out, "[t]his ignorance... cuts both ways [since] few visionaries foresaw the disruptive [sic] effects of the

World Wide Web, digital music, ubiquitous smartphones, social media, or fracking” (2015).

We recognize the attraction of remaining politically neutral in the face of the predictive, evaluative, and goal-setting challenges discussed above, and that it is unattractive to forego potential benefits for fear of ill-defined concerns. However, above we showed that the threat we envisage is neither ill-defined nor mere scaremongering but well-defined and very concrete. Furthermore, as things stand now, we do not have ample safeguards to secure safety, at least not when safety is understood in more than just the arbitrarily-bounded medical sense. Furthermore, because of the problems with predicting and evaluating the potential social side effects of CE, it is far from clear how informed consent could even be obtained from the population. Worse still, what is troubling about social side effects is that when one person uses CE to become more competitive, other people feel the side effects—namely, the greater pressure to compete, and the resulting changes in the social environment—and this, too, creates important problems for reflecting on who should even be asked to provide informed consent. Furthermore, because we all live in society together, if medically safe and effective CEs are made available to everyone, individuals could not easily opt out of society to avoid the effects of this technology. And who, in any case, would a person who wishes to opt out even go to in order to request that *they* no longer take part in this social experiment? Who precisely would be in charge of noting the ways that society is being changed, how people’s views on matters are being influenced, and ensuring those who do not wish to participate in this social experiment *can* opt out?

For such reasons, our case is that if CE is to improve rather than worsen the quality of our lives then, pace Pinker and other opponents of regulation, this technology needs to be very carefully regulated (Alternatively, we may need to design CE technologies that produce qualitatively different effects).

Regulation and freedom

Admittedly, regulation tends to be viewed as a bad thing, as a state imposed restriction of individual freedom. Hence, in this subsection we wish to recount some ways in which regulation of the right sort can be freedom-expanding.

Firstly, we suspect that one reason why regulation gets such a bad rap—why it is often viewed as a restriction of freedom—is simply because the term “regulation” is misunderstood. For instance, when regulation is viewed as a matter of either prohibition or permission, it is easy to assume that greater regulation means more prohibition, and that more prohibition entails greater infringements on individual freedom (Swierstra 2015, 18). However, not all prohibitions reduce freedom—for instance, when you are prohibited from taking what is ours, our freedom is expanded not contracted. Furthermore, regulation comes in many more shades than just permission

and prohibition—these are merely two *modes* of regulation on a scale that includes at least three others:

prohibit – discourage – permit – encourage – require

Some modes of regulation can empower rather than disempower and thus enhance rather than diminish freedom. Also, precisely who would be the subject of regulations—e.g., who would be permitted, or encouraged, required, discouraged, or prohibited from using CE—is something that can also be nuanced. A nuanced approach to regulating CE could apply one mode of regulation (e.g., permission) to one group of people and a different mode of regulation (e.g., requiring) to another group. Furthermore, instead of viewing regulation as an instance of the state (conceptualized as a separate entity from the citizens whom it governs) dictating to its citizens what they may and may not do, in a democracy, regulation can also be conceived of as citizens telling appointed officials how to arrange things on their behalf, and this, too, creates the potential for a similar degree of nuance vis à vis who should do the regulating. Our point here is just this: conceiving of regulation as a matter of prohibition or permission is unhelpful because it depicts regulation as a crude and coercive instrument which it actually need not be.

Secondly, another way in which regulation can be freedom-expanding is by providing us with a way to protect ourselves from the coercive factors of a prisoner's dilemma-like situation, which, if we do nothing about it, would lead us to make choices in competitive contexts that will produce unwanted results. What we face in the scenario that we described is not unlike the coercive prisoner's dilemma regarding, for example, doping in sports where regulations aim to protect us from feeling pressured to use prohibited substances because others might be doing so (World Anti-Doping Agency n.d.; Heathers 2012; Mazanov 2012). In the context of sports, we put regulations into place so that everyone may assume safely that others are not using such substances, and hence nobody will have reason to begin using them. But even if doping were perfectly medically safe, we might still put regulations in place that ban doping to prevent the nature of sports from becoming unappealing, for instance, by shifting its focus in a more technical direction (Santoni de Sio, Robichaud, and Vincent 2014; Santoni de Sio et al. 2016).

Thirdly, regulation is also freedom-expanding in another distinctly positive way—that is, regulation not only protects our *freedom from* coercion, but it also expands our *freedom to* do things by providing us with a way to take charge of how we shape the future social environment in which we and our descendants will live and in which we shall exercise our agency. Through regulations, we can shape the future environment so that it is both qualitatively more to our liking (e.g., not too competitive and demanding), as well as more conducive to expanding rather than constraining our opportunities (or, put

another way, so that our environment constrains us in ways we wish to be constrained, and does not offer us opportunities we would rather not be offered). It is a way for us, in the moment, to exercise control over our future, and in so doing to secure circumstances that will be qualitatively attractive and conducive to our ongoing exercise of agency. Viewed from this perspective, regulation is an incredibly valuable tool for creatures like us—creatures who have a future and a past as well as a present, and who live together in societies that are shaped by how we live, and that shapes our values—because, through intelligent regulation, we can enjoy the fullest exercise of agency over time with others.

Through regulation, we can extend our agency temporally so that our in-the-moment choices become more than mere *reactions* to whatever pushes and pulls are present. Through regulation, we borrow some of the reasons for our choices from the future and from the past by considering the distal as well as proximate causes and effects of our choices in both a “Have you thought about what might occur?” sense and an “I’ve thought about it and here’s what I would like to bring about” sense. Through regulation (or perhaps “*regulating*”), we also extend our agency in a social dimension. Instead of relinquishing control over the future shape of society,¹⁴ we take control over its shape by expressing our substantive preferences, and then put in place mechanisms to secure conditions conducive to those preferences being realized. Rather than allowing a lack of forethought—both myopia for the future and for the collective nature of the choice situation in which we find ourselves—to result in outcomes that are at best unexpected and at worst undesirable, regulation protects our freedom from coercion, and it enables us to take charge by expanding our agential powers in temporal and social dimensions.

Framing this as a debate about *state regulation* is thus extremely unhelpful because it suggests that the issue at stake is whether your individual liberty may be curtailed by government officials who think that they know, better than you do, what is good for you. This debate is really about *self-regulation* (not others regulating us) and *extending individual freedom* (not curtailing it).

A methodology for the design and regulation of emerging technologies

We have argued that framing the discussion about CE as a debate about state regulation of pharmaceuticals and medical devices is very unhelpful. Framing CE as a *medical topic* renders invisible, or downplays, important non-medical concerns such as seriously adverse social side effects. Further, the *state regulation* framing treats challenging practical problems—ones which, as we shall now argue, we believe are amenable to practical solutions—as if they were a fixed part of the moral and political landscape

and about which we can do nothing (even though doing nothing effectively relinquishes our control over the future shape of society in which we and our descendants will live). For this reason, we conclude this chapter by proposing that the CE topic needs to be reframed as an *ethics of technology* topic (as opposed to a medical topic, i.e., bioethics or neuroethics) in which the debate addresses how to secure *diachronic self-control* (as opposed to state regulation).

As we argued above, within this reframing, the seemingly political challenges involved with predicting and evaluating potential social side effects,¹⁵ as well as the goal-setting challenges,¹⁶ become individual not political matters. We believe that the remaining challenges for prediction, evaluation, and goal-setting can be addressed through the following nuanced, gradual, and iterative methodology for the design and regulation of emerging technologies that attempts to give due recognition to our previous observations in this chapter. The seven steps of our proposed methodology are as follows:

- 1 reflect on what constitutes human flourishing;
- 2 study past emerging technologies to understand better their possible outcomes;
- 3 informed by (2), anticipate the presently emerging technology's likely outcomes;
- 4 work out if those outcomes would be conducive or antagonistic to human flourishing as per (1);
- 5 regulate and design to promote flourishing-conducive outcomes and avoid flourishing-antagonistic outcomes;
- 6 deploy the technology slowly, gradually, to a restricted population, in accordance with regulations from (5);
- 7 periodically reappraise our changing views on human flourishing at (1), and adjust regulations and design to account for them and the actual outcomes.

This methodology has several features which we wish to highlight. First, it implicitly presupposes that regulation is what citizens engage in—it is what we do to ourselves, or how we instruct the government to regulate technology on our behalf, not something that governments inflict on their citizens. Second, and related, precisely because regulation is what we do to ourselves, it is something that requires *public* reflection. Hence, continual public engagement and reflection on values is a core component of this methodology. Third, instead of trying to do the impossible—i.e., to predict unpredictable consequences and to evaluate the unevaluable—we explicitly recommend a gradual and iterative approach. We should try to predict what we can, do the best we can at evaluating what we predict, design and regulate as best we can, and then, after some time has passed, come back to reappraise results. Fourth, given that regulation is a finely nuanced instrument, we

recommend that regulations be as nuanced as this fine instrument allows (i.e., not just blanket prohibition or permission, which puts us in the untenable position of having to either recklessly embrace unknown dangers or to sheepishly forego the possibility of discovering new technologies that benefit us because we were too afraid to try something new). Fifth, because CE is not the first technology that has the potential to alter society, we think that valuable insights can be gained from studying past technologies and past regulations—e.g. insights about such things as what effects technologies can produce on society and how different regulations of those technologies can modulate what effects those technologies produce. Sixth, because technologies can directly (e.g., via brain changes) and indirectly (e.g., via social change) alter us as evaluators, we propose that the evaluative tasks should be performed with awareness of not only current values, but also past values, as well as a vision of what sort of people and society we wish to become. We must challenge ourselves to provide an explanation for why these changing values are, from our current standpoint, defensible as a basis from which to make the evaluations that we make. Seventh, we propose two reasons why a new technology should be deployed slowly, gradually, and to a restricted population: one, to ensure that our values do not change rapidly due to the direct or indirect impact of a technology on us as evaluators such that we lose our ability of adequate self-critical reflection, and two, to ensure that if unpredicted serious problems manifest, they will not result in a massive public disaster.

Finally, we wish to relate our seven-point methodology and the preceding discussion specifically to the theme of the volume for which this chapter has been written. The mis-framing of CE as a topic that concerns state regulation of medical technology has obscured the real issues in this debate. In the design and regulation of emerging technologies, social experimentation with technology is *indispensable*, and we wish to highlight five points that relate our chapter directly to this claim about the indispensability of social experimentation with technology.

One, this social experimentation is inevitable because no amount of arm-chair reflection can overcome the predictive, evaluative, and goal-setting challenges that we described. We can, at best, make only short-range predictions and evaluations and set short-range goals, and then we must periodically revisit these predictions, evaluations, and goals.

Two, this experimentation is inherently social in nature for at least two reasons: one, it is social because technology is deployed in society and changes the material conditions of our existence (e.g., that we can travel greater distances through motorized forms of transport, that food can be grown in far off locations to where we live, that diseases that once plagued us no longer do, etc.) and has clear social effects that we outlined, and two, it is social because one of the most important effects of technology is to alter us as evaluators. Scientific and technological developments, when deployed in society, shape our lives, often imperceptibly, by gradually changing our

values and the moral, legal, and social landscape in which we operate as agents. The opportunities that technologies make available to us can have weighty moral implications—for what we have reason to do, for what others have reason to do, and more generally, for what rights and duties people have (Vincent 2013; Santoni de Sio, Faulmüller, and Vincent 2014). This relationship between technology, society, and the moral, legal, and political landscape is so intricate that we cannot imagine modeling it in an armchair and figuring things out *ex ante*. Rather, it requires in-the-flesh social experimentation with technology.

Three, restricting the scope of the debate about CE to neuroethicists and bioethicists has been detrimental to the aim of making progress on this topic, and not only because it has shielded from view the fact that important potential social side effects have not been considered in decisions about what effects CE should have and how its use should be regulated. It has also, more broadly, kept a wider audience of scholars—in the philosophy and ethics of technology, and in political philosophy more generally—from making contributions to this debate. CE is not inherently a bioethics or neuroethics topic. It is not inherently a medical topic. It is a philosophy of technology topic, an ethics of technology topic, and a political philosophy topic. Consequently, discussions of this topic should not be confined to bioethics and neuroethics journals, but rather they should occur in mainstream philosophy journals with special contributions from political philosophy, philosophy of technology, and ethics of technology. These discussions should also occur in scientific and engineering journals.

Four, direct brain intervention-based methods—e.g., pharmaceuticals and transcranial electrical and magnetic stimulators—are not the only forms of CE that we should be discussing. Indeed, the term “cognitive enhancement” makes it awkward to discuss a range of other technologies, such as the plethora of information communication technologies, from within this framing. However, other technologies do have significant CE effects—they enhance our ability to make decisions, to do things, and to interact with one another—and experimenting with these other technologies poses similar problems to the ones that we have discussed in this chapter. Hence, our discussion should not be viewed as a discussion of special and narrow relevance to CE medications and devices, but rather as a discussion with broad implications for the design and regulation of all technologies.

Five, and lastly, when we say that the design and regulation of emerging technologies involves people experimenting with technology in society, we also wish to highlight that the people involved are not only scientists and engineers but also consumers and politicians. People design technologies, use them, and regulate them. In light of the point made in the previous paragraph, it is unproductive to view CE as a niche topic that concerns something designed in laboratories by neuroscientists. The public needs to be included in this debate.

Because critically important values are at stake when we engage in social experimentation with technologies—namely, how humans shape themselves and their future environments—we think that a socially responsible approach to the design and regulation of emerging technology needs to be as critical and as self-reflective as we recommend in our seven-point methodology. Given that the technologies we create and deploy will change us as evaluators, we need to be sufficiently self-reflective and critical to notice that we have a responsibility to our future selves and to those who come after us—a responsibility to think about how the decisions we make now may have an impact on how we or our descendants make decisions in the future, and thus whether the decisions that we make right now are as defensible as they might appear from a more narrow present-focused perspective. Our point is not just that the technologies we create might pollute our natural environment, but that they may change us as evaluators in possibly indefensible ways. That is indeed the central reason why our seven-point methodology involves as much navel-gazing and re-gazing to continually reappraise our current values within a broader diachronic context. The recognition that we are malleable creatures who are influenced in our views and values by the technologies and social arrangements we put into place (or allow to come into being) carries with it, we think, an obligation to treat this malleability as something to be monitored and reflected upon, to ensure we do not shape ourselves in ways that we would, with the benefit of hindsight, have reason to regret.

Notes

- 1 Nicole A. Vincent is an Honorary Fellow in the Department of Philosophy at Macquarie University and a visiting researcher in the Department of Philosophy at Technische Universiteit Delft. Emma A. Jane is Senior Research Fellow in the School of Arts and Media at the University of New South Wales. Nicole A. Vincent's work on this chapter was partly funded by grants from Macquarie University and Technische Universiteit Delft, as well as the very generous support of a grant from the John Templeton Foundation via The Enhancing Life Project. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.
- 2 The spectrum on which we have placed our example brain modification techniques (or neural modifiers, as the Presidential Commission calls them) is sometimes described as capturing the degree of *directness* or *indirectness* with which a given technique effects changes in the brain. While the traditional techniques (e.g., conversation or written text) are said to modify the brain *indirectly*, in contrast novel techniques (e.g., brain surgery) are said to modify the brain *directly*. The distinction between direct and indirect brain modification techniques—and in particular, its normative significance—is not unproblematic. However, we are not the only ones who think that it still captures the intuition that these ways of changing people's minds are importantly different, and not just because medical interventions like surgery (as opposed to conversation) carry a range of medical risks like infection. In what follows, we shall adopt Jan Christoph Bublitz and Reinhard Merkel's suggestion that "indirect interventions are inputs into the

cognitive machinery our minds are adapted to process, whereas direct interventions change the cognitive machinery itself” (2014, 70).

- 3 While the distinction between more and less invasive brain interventions is also contested, for the purpose of the ensuing discussion we shall treat the invasive–non-invasive distinction as being roughly equivalent to the direct–indirect intervention methods distinction. For discussion, see Bostrom and Roache (2008, 121–22).
- 4 For a similar list of anticipated benefits, see Bostrom and Roache (2008, 138–39).
- 5 We do not discuss potential *qualitative* improvements—like that users of this pill might not just (e.g.) read, learn, and think *faster*, but also develop deeper or more profound insights, or have more earth-shattering breakthroughs in their thinking—because such qualitative improvements are very difficult to define (people can more easily disagree about what is a qualitative improvement than what is a quantitative improvement) and to measure. We can make our point just as well without introducing these additional complications.
- 6 Diethyl Stilbestrol (Dunea and Last 2001) and Thalidomide (National Cancer Institute 2011) are typical examples.
- 7 Barbara Sahaikan and colleagues argue that “modafinil seems to affect motivation in a manner that makes unappealing tasks more appealing and therefore they can be undertaken and completed more easily. In other words, overall task-related pleasure is increased by modafinil” (Sahakian et al. 2015, 9).
- 8 That is, to be prepared to engage in tasks that were increasingly mind-numbingly boring, and/or to be prepared to engage in a greater number of boring tasks.
- 9 Swierstra writes that “technologies don’t exclusively have ‘hard impacts’ like poisoning, exploding, polluting, and depleting”, but also “soft impacts” like the ones that we highlight here which “are *qualitative* rather than quantitative; the core values at stake are *unclear or contested* rather than clear instances of harm; and the results are *co-produced by the user* rather than being caused solely by the technology” (2015, 7). Nick J. Davis (2017, 5–6) has also recently argued that CEs may produce societal, not just neurobiological and ethical, harms.
- 10 We set aside the question of what precisely “medical” means because our point is not to endorse drawing on the distinction between medical vs. social side effects as if it provided a useful way of distinguishing between important and unimportant side effects, but to show why this assumption is problematic.
- 11 Swierstra writes “[w]e are still in the process of devising methods to deal with the new normative challenges that come with ... soft impacts” (2015, 5). In section 6 below we propose a new methodology to fill this gap.
- 12 We take it that these will count as examples of medical maladies on any plausible definition of “medical”.
- 13 For a discussion of how transformative experiences present a challenge to evaluation and goal setting, see Laurie A. Paul’s book, *Transformative Experience* (2014).
- 14 That is, instead of allowing society’s future shape to be decided by however the invisible hand of competition, fuelled by blind (or financial-reward-responsive) technological progress, orchestrates individual choices into concrete outcomes.
- 15 Here are two examples of specifically *political* challenges to prediction and evaluation. Precisely which side effects should be of concern? And by whose standards should they be evaluated?
- 16 An example of a seemingly-political goal-setting challenge is who should decide what sort of society—e.g., a highly competitive one or a less demanding one—is promoted through CE technology?

References

- Bostrom, Nick, and Rebecca Roache. 2008. "Ethical Issues in Human Enhancement." In *New Waves in Applied Ethics*, edited by Jesper Ryberg, Thomas Petersen, and Clark Wolf, 120–52. Basingstoke: Palgrave Macmillan.
- Bublitz, Jan Christoph, and Reinhard Merkel. 2014. "Crimes against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination." *Criminal Law and Philosophy* 8:51–77.
- Coffman, Brian A., Vincent P. Clark, and Raja Parasuraman. 2014. "Battery Powered Thought: Enhancement of Attention, Learning, and Memory in Healthy Adults Using Transcranial Direct Current Stimulation." *NeuroImage* 85:895–908.
- Davis Nick J. 2017. "A Taxonomy of Harms Inherent in Cognitive Enhancement." *Frontiers in Human Neuroscience* 11:63. doi:10.3389/fnhum.2017.00063.
- Dunea, George, and John M. Last. 2001. "Thalidomide." In *The Oxford Illustrated Companion to Medicine*, edited by Stephen Lock, John M. Last, and George Dunea, 807–12. 3rd Edition. Oxford: Oxford University Press.
- Farah, Martha J., Judy Illes, Robert Cook-Deegan, Howard Gardner, Eric Kandel, Patricia King, Eric Parens, Barbara Sahakian, and Paul Root Wolpe. 2004. "Neurocognitive Enhancement: What Can We Do and What Should We Do?" *Nature Reviews. Neuroscience* 5:421–25.
- Glannon, Walter. 2008. "Psychopharmacological Enhancement." *Neuroethics* 1:45–54.
- Greely, Henry T., Barbara Sahakian, John Harris, Ronald D. Kessler, Michael S. Gazzaniga, Philip Campbell, and Martha J. Farah. 2008. "Towards Responsible Use of Cognitive Enhancing Drugs by the Healthy." *Nature* 456:702–705.
- Heathers, James. 2012. "Lance Armstrong Charged with 'Blood Doping' and EPO-Use... So How Do They Work?" *The Conversation*, June 12. Accessed April 25, 2017. <http://theconversation.com/lance-armstrong-charged-with-blood-doping-and-epo-use-so-how-do-they-work-7666/>.
- Jane, Emma A., and Nicole A. Vincent. 2015. "The Rise of Cognitive Enhancers Is a Mass Social Experiment." *The Conversation*, June 15. Accessed April 25, 2017. <https://theconversation.com/the-rise-of-cognitive-enhancers-is-a-mass-social-experiment-40072>.
- Jonas, Hans. 1973. "Technology and Responsibility: Reflections on the New Tasks of Ethics". *Social Research* 40:31–54.
- Mazanov, Jason. 2012. "The Lance Bomb Has Blown, But Is Doping Really Cheating?" *The Conversation*, October 17. Accessed April 4, 2014. <http://theconversation.com/the-lance-bomb-has-blown-but-is-doping-really-cheating-10183>.
- Meinzer, Marcus, Sophia Jähnigen, David A. Copland, Robert Darkow, Ulrike Grittner, Keren Avirame, Amy D. Rodriguez, Robert Lindenberg, and Agnes Flöel. 2014. "Transcranial Direct Current Stimulation over Multiple Days Improves Learning and Maintenance of a Novel Vocabulary." *Cortex* 50:137–47.
- National Cancer Institute. 2011. "Diethylstilbestrol (DES) and Cancer." *National Cancer Institute Factsheet*, October 5. Accessed April 25, 2017. www.cancer.gov/about-cancer/causes-prevention/risk/hormones/des-fact-sheet.
- Ojiaku, Princess. 2015. "'Smart Drugs' Are Here — Should College Students be Allowed to Use Them?" *The Washington Post*, November 3. Accessed January

- 25, 2017. www.washingtonpost.com/news/speaking-of-science/wp/2015/11/03/smart-drugs-are-here-should-college-students-be-allowed-to-use-them/.
- Outram, Simon, and Bob Stewart. 2012. "Smart Pills: Magic Bullets or Benign Slugs?" *The Conversation*, July 22. Accessed April 25, 2017. <http://theconversation.com/smart-pills-magic-bullets-or-benign-slugs-7628>.
- Paul, Laurie A. 2014. *Transformative Experience*. New York: Oxford University Press.
- Pinker, Steven. 2015. "The Moral Imperative for Bioethics." *The Boston Globe*, August 1. Accessed February 2, 2017. www.bostonglobe.com/opinion/2015/07/31/the-moral-imperative-for-bioethics/JmEkoyzITAu9oQV76JrK9N/story.html.
- Presidential Commission for the Study of Bioethical Issues. 2015. *Volume 2, Gray Matters: Topics at the Intersection of Neuroscience, Ethics, and Society*. Accessed January 25, 2017. http://bioethics.gov/sites/default/files/GrayMatter_V2_508.pdf.
- Sahakian, Barbara J., Annette B. Bruhl, Jennifer Cook, Clare Killikelly, George Savulich, Thomas Piercy, Sepehr Hafizi, Jesus Perez, Emilio Fernandez-Egea, John Suckling, and Peter B. Jones. 2015. "The Impact of Neuroscience on Society: Cognitive Enhancement in Neuropsychiatric Disorders and in Healthy People." *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* 370 (1677):20140214. doi:10.1098/rstb.2014.0214.
- Santoni de Sio, Filippo, Nadira Faulmüller, Julian Savulescu, and Nicole A. Vincent. 2016. "Why Less Praise for Enhanced Performance? Moving beyond Responsibility-Shifting, Authenticity, and Cheating to a Nature of Activities Approach." In *Cognitive Enhancement: Ethical and Policy Implications in International Perspectives*, edited by Fabrice Jotterand and Veljko Dubljević, 27–41. Oxford: Oxford University Press.
- Santoni de Sio, Filippo, Nadira Faulmüller, and Nicole A. Vincent. 2014. "How Cognitive Enhancement Can Change Our Duties." *Frontiers in Systems Neuroscience* 8:131. doi:10.3389/fnsys.2014.00131.
- Santoni de Sio, Filippo, Philip Robichaud, and Nicole A. Vincent. 2014. "Who Should Enhance? Conceptual and Normative Dimensions of Cognitive Enhancement." *Humana. Mente: Journal of Philosophical Studies* 26:179–97.
- Schwarz, Alan. 2015. "Workers Seeking Productivity in a Pill Are Abusing A.D.H.D. Drugs." *The New York Times*, April 18. Accessed April 25, 2017. www.nytimes.com/2015/04/19/us/workers-seeking-productivity-in-a-pill-are-abusing-adhd-drugs.html?emc=edit_th_20150419&nl=todaysheadlines&nid=64524812.
- Swierstra, Tsjalling. 2015. "Identifying the Normative Challenges Posed by Technology's 'Soft' Impacts." *Etikk i praksis* 9:5–20.
- Thomson, Helen. 2015. "Narcolepsy Medication Modafinil Is World's First Safe 'Smart Drug'." *The Guardian*, August 19. Accessed February 2, 2017. www.theguardian.com/science/2015/aug/20/narcolepsy-medication-modafinil-worlds-first-safe-smart-drug.
- Van de Poel, Ibo. 2013. "Why New Technologies Should be Conceived as Social Experiments." *Ethics, Policy & Environment* 16:352–55.
- Vincent, Nicole A. 2013. "Enhancing Responsibility." In *Neuroscience and Legal Responsibility*, edited by Nicole A. Vincent, 305–33. New York: Oxford University Press.
- Vincent, Nicole A., and Emma A. Jane. 2014. "Put Down the Smart Drugs – Cognitive Enhancement is Ethically Risky Business." *The Conversation*, June 16. Accessed

January 25, 2017. <https://theconversation.com/put-down-the-smart-drugs-cognitive-enhancement-is-ethically-risky-business-27463>.

Wise, Brian. 2013. "Musicians Use Beta Blockers as Performance-Enabling Drugs." *WQXR*, August 16. Accessed April 25, 2017. www.wqxr.org#!/story/312920-musicians-use-beta-blockers-relieve-stage-fright/.

World Anti-Doping Agency. n.d. *The Code*. Accessed April 25, 2017. www.wada-ama.org/en/what-we-do/the-code.

7 Living a real-world experiment

Post-Fukushima imaginaries and spatial practices of “containing the nuclear”

Ulrike Felt

Introduction

On the afternoon of March 11, 2011, an earthquake of a size that Japan had never experienced before and a follow-up tsunami hit the area of the Daiichi nuclear power plant in the prefecture of Fukushima. The resultant flooding disabled the power supply of the nuclear power plant, and the emergency core cooling system was not able to stabilize the reactor temperature, leading to a meltdown. Between March 12 and March 15, 2011, three of the reactors blew up, releasing radioactive material into the air and water, leading to one of the world's largest technonatural disasters. More than 150,000 people were evacuated from the area. At the time of writing this chapter, it has become clear that some will never be able to return to their homes, whereas others feel that they no longer want to inhabit the irradiated spaces. Additional areas have been declared as inhabitable, and people have begun to move back.¹

For Japan, the “Fukushima disaster” came at a time when nuclear technologies had become well entrenched in its technopolitical culture (Hecht 2009; Felt, Fochler, and Winkler 2010). From the late 1950s onwards, the Japanese government had made considerable efforts to develop a strong pro-nuclear agenda, both as a compensatory reaction to the trauma suffered through the atomic bombing during WWII and as a demonstration of the country's technological capacity (e.g., Jones, Loh, and Sato 2013; Fujigaki 2015). Public promotion campaigns, schoolbooks disseminating the narrative of a safe nuclear future, and Japan's engagement in the International Atomic Energy Agency (IAEA) were all part of integrating “the nuclear” into Japan's “technopolitical identity” (Felt 2015; Fujigaki 2015) as a technologically advanced nation. Nevertheless, Hecht's (2012) notion of “nuclearity” reminds us that the manner in which “places, objects, or hazards are designated as ‘nuclear’” is by no means clear and uncontested. This is essential to my story because what is perceived of as problematic with regard to “the nuclear” never was, is not, and never will be as clear-cut as is often assumed. Although housing, schools, hospitals, leisure locations (see Figure 7.4), and farmland seemed to inhabit the same space as the nuclear power plant without major friction, this relation was disrupted with the accident

in the Daiichi nuclear power plant. Even in the aftermath, accounts of the nuclear disaster were never solely about radiation. Instead, they were multi-layered, diverse, and entangled. They were about the political system's lack of transparency, failures in the functioning of nuclear industry, a lack of appropriate information diffusion, the manner in which the clean-up was handled, and so on (Hecht 2013). Therefore, when this chapter investigates the strongly expert-driven spatial reorderings that occurred after the nuclear disaster, the story will be as much about space lost to the nuclear as about belonging, attachment, and displacement, the loss of trust and disappointment in governmental and industrial actors, disrupted sociotechnical imaginaries of control and containment built around nuclear technology, and unclear futures. The nuclear is thus configured and comes to matter in multiple ways in the many narratives about Fukushima.²

In the early 1960s, the town councils of Futaba and Okuma, towns that are very close to the Daiichi nuclear power plant (NPP), passed resolutions to invite the construction of nuclear power plants; at that time, the promise and the expectations of this technology were great. To welcome the technological site meant to develop the region, to realize a better future, and to bring income granted by the state to the regions that would host the NPPs (Jones, Loh, and Sato 2013). Pictures of the huge crossroad sign at the entry of the town of Futaba, located inside the 20-kilometer exclusion zone drawn around the NPP after the accident, show one reminder of this past, an expression of those hopes: "Nuclear Power: Energy for a brilliant future." Five years later, as the government prepared the ground to move people back into these areas, pictures of Futaba's crossroad sign were again shown in the media. This time, the dismantling of this material witness of a strong techno-optimism was shown, representing a changing relation to nuclear technology. This does not mean that there were no critical voices in the early nuclear period, only that there were no clear political signs that the technology could and would be called in question.

In the aftermath of the disaster, analysts in science and technology studies (STS) (e.g., Pfotenhauer et al. 2012; Jones, Loh, and Sato 2013; Fujigaki 2015; Yamaguchi 2016) and beyond have begun to reflect on and investigate many facets of the unfolding technonatural and above all human disaster, studying emergency response procedures, public communication efforts and their narrative framings, the history of nuclear development in Japan, risk assessments, anti-nuclear movements, and many more. This chapter will complement these studies, following "the nuclear" once escaped from the reactor, thus having disrupted the "sociotechnical imaginary" (Jasanoff and Kim 2009, 2015) of both control and containment and of progress and development that had been so carefully constructed, nourished, and protected by not only the nuclear industry, but also governmental and other pro-nuclear actors in Japan in the post-WWII period. Such imaginaries are essential elements of any technopolitical culture because they represent "collectively held, institutionally stabilized, and publicly performed visions

of desirable futures, animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology” (Jasanoff and Kim 2015, 4). Destabilizing such an imaginary through a major accident would thus call into question much more than a single technological site; indeed, it could potentially threaten a technology-centered societal order.

Below, I will argue that the period after the Fukushima Daiichi NPP accident should be understood and investigated as a huge “real-world experiment” (Gross and Krohn 2005), with both the human subjects and the environment exposed to numerous technological, social, and scientific interventions with unclear long-term outcomes. Although some have conceptualized the employment of nuclear energy technology broadly speaking as a social experiment given that it “retains an experimental nature even after its implementation in society” (van de Poel 2011, 287), this chapter will focus on a specific experimental episode, namely, the spatial interventions performed after the accident with the hope of lowering human and environmental risk, upholding the image of the region, and assuring the survival of the sociotechnical imaginary built around nuclear energy production.

To fully grasp the experimental character of the post-disaster interventions, it is important not only to highlight the open-endedness and tentativeness of the diverse actions taken and interventions made, but simultaneously to follow the construction of the “laboratory walls,” i.e., the making of the space(s) in which the experiment would be taking place. The latter should allow—so it is hoped—for the production of stable and satisfactory outcomes within the confines of the paradigm of nuclear controllability. I will investigate how diverse actors began to design and implement a range of sociotechnical interventions to re-envision and sustain a continued imagination of control and containment. In particular, my attention will turn to the techniques of the (temporal) reordering of space around the Daiichi nuclear power plant as a key element of the experimental setup.

The chapter begins both by introducing the notions of experiment and laboratory, and by outlining the conceptual understanding of space and its production as used in this chapter. I will then focus on the three techniques of space (re)creation that are deployed: map-making, physical demarcations, and the displacement of contaminated radioactive soil. More concretely, I will examine different radiation and evacuation maps, analyze practices of physically defining/imposing/self-constructing zones of non- or limited habitation, and reflect on the efforts of collecting and “re-localizing” radioactive soil, thus redefining the space it had covered before. This will lead me (1) to argue that the retention of power and control over the post-nuclear-accident experiment relies on the capacity to shape the production and distribution of space; (2) to show the adaptation of the sociotechnical imaginary of nuclear containment through spatial redefinition; and (3) to reflect on the understandings and distributions of responsibility for and within such experimental spaces.

The following analysis and reflections draw on two site visits to the Fukushima prefecture in November 2013 and July 2014, where I participated in two international conferences organized by the Medical University of Fukushima and the IAEA,³ an IAEA working group in Vienna addressing health issues around Fukushima, numerous visual materials used in debates and talks on Fukushima by local experts, extensive conversations with medical practitioners at Fukushima Medical University and other experts visiting the area, and scholarly (STS) work produced in the aftermath of the disaster.

Experimental societies and the role of spatial arrangements

In recent decades, we have witnessed the emergence of a growing body of literature addressing the deep entanglement of contemporary societies with technoscientific developments. It has become obvious that it is impossible to separate the technical and social, the natural and the cultural—the Fukushima disaster being an outstanding case of this hybridity. Technological orders and social orders would be coproduced in the same move (Latour 1993; Jasanoff 2004). According to this understanding, the nuclear power complex is conceptualized as a network of diverse sets of human actors, technological entities, regulatory regimes, risk-benefit calculations, promises of progress, and much more, which cannot be separated from one another either for the sake of analysis or when making policy choices. Together, they perform certain societal values and practices while simultaneously being their product; similarly, national identity, risk regulation, and corporate culture are materialized in the production and operation of nuclear power plants.

Notions such as the “society as the laboratory” (Krohn and Weyer 1994), “self-experimental society,” and “real-world experiments” (Gross and Krohn 2005) aim at capturing these shifts. Underlying these discussions is the observation that “social practices increasingly present themselves as experiments via a willingness to remain open to new forms of experience,” and that this type of experiment “appears to be precariously perched both on the actions of its participants and on the structures created by the members of society” (ibid., 79). Using the notion of an experiment in such a wider societal context directs us to the fact that here the “laboratory” has no clear walls, that humans are affected, but not necessarily identified as experimental subjects, and that the underlying hypothesis is not stated very clearly and evidence is not collected systematically. It often also means that in such “social experiments [...] it [is] more difficult to control experimental conditions,” it might prove somewhat complex to terminate them, and they might have not only unintended but also (and above all) irreversible consequences (van de Poel 2011, 287). What these writings have in common is their diagnosis that experiments are being conducted in a range of new types of hybrid settings, often related to technoscientific developments, which pose new and

challenging questions. Who has the authority to design and perform experiments? Who can participate in establishing the protocol for experiments? How is informed consent assured in highly complex environments? How is the validity of outcomes assured? And, what are the regimes of valuation that decide the success of these experiments? These are but a few of the questions that must be asked given the hybrid form of the setting in which these technosocial experiments take place and choices must be made. (Felt et al. 2007; Callon, Lascoumes, and Barthe 2009; van de Poel 2011)

In such an experimental setting, we must also attend to the fact that there is no clear consensus, neither about the status of the phenomenon at stake nor about who should be the actors involved in reaching such a consensus. We cannot expect—as is the case in a classical laboratory experiment—that ontological ambiguities of the phenomenon under investigation and the temporal dimensions of its production can be fixed. (Pickering 1992) Indeed, at the core of any real-world experiment, we find permanent alternative configurations, i.e., ontologies always remain plural. Neither do they precede technoscientific practices; rather, they are shaped by them. Understanding ontology as that which belongs to “the real,” they also crucially define the conditions of possibility in which we live. This is very much in line with Mol’s understanding of ontologies as always “brought into being, sustained, or allowed to wither away in common, day-to-day, sociomaterial practices” (Mol 2002, 6). In the case of this chapter, such an approach also requires attention to the ontological politics at work when reconfiguring space around the NPP and to look at the ways, to paraphrase Mol (2002, vii), in which problems are framed, how spaces are continuously shaped and reshaped, and how, as a related issue, lives are pushed and pulled into one shape or another.

Exploring the concept of the experimental society and social experimentation in order to better comprehend the developments in the post-nuclear-accident period, a second notion requires closer scrutiny: that of the laboratory (see e.g., Knorr-Cetina 1981). In what is to be understood as “laboratory” in the context of this chapter, I draw on Michael Guggenheim’s (2012, 101) argument that a laboratory “is not so much a closed space, but a procedure that often results in a space with the properties to separate controlled inside from uncontrolled outside.” The understanding of the laboratory is thus “procedural and praxeological;” it is a space that must be brought into being, and it must be imagined and practiced. The “interventionist character of the laboratory” is thus essential, as is “the idea that the laboratory is an assemblage of technologies and practice” (ibid., 102). In what follows, the laboratory will be a region within a wider national and global context and, therefore, we will be confronted with the potential meaning of the distinction of the controlled inside and the uncontrolled outside in the context of this experiment. Within this laboratory, I will follow the reorganization of space in response to the dramatically changed conditions after the nuclear accident and reflect on the types of interventions

performed. I will ask how boundaries are being drawn and upheld within the laboratory and toward the outside, and how classification is performed. Moreover, I will reflect on the ontological politics at work.

Before engaging with this perspective, I would like to clarify briefly my understanding of space. My analysis has been guided by conceptualizing space as being brought into being through relations and practices, in particular drawing on the work of David Harvey (1990) and Martina Löw (2016). What their respective theories of space share is that space is defined neither by its physical dimensions nor by classical representational practices; instead, space is something that must be continuously brought into being through practices. However, the two authors' theories of space also have important differences. Harvey, building on Lefèbvre's work (e.g., 1991), develops a more structural concept of space in the Marxist tradition, reminding us how deeply the capacity of exercising power and domination is intertwined with the capacity to influence and control the production and distribution of space. For the following analysis, I was primarily inspired by Harvey's "grid of spatial practices" (1990, 220), which sensitized me to the multiplicity of spatial work performed. In developing the grid, Harvey actually begins from Lefèbvre's *The Production of Space* (1991), which identified three key dimensions: (1) material spatial practices, which draw attention to physical and material flows, transfers, and interaction; (2) the representation of spaces, in that we pay attention to the signs, codes, and knowledge that permit people to talk about space and make sense of it; and (3) spaces of representation that enable us to reimagine and re-perform space in new ways. He then complements these three dimensions with a set of four aspects that characterize spatial practices: (1) "accessibility and distantiation," which speaks to delimitations, barriers, and their effects; (2) the "appropriation of space," which focuses on how space is/can be occupied by objects and people, allowing for some activities but not others; (3) the "domination of space," which points at who can decide how space is occupied; and (4) "the production of space," which draws attention to how spaces are changed, e.g., through the introduction of new systems of transportation, communication, and representation. This approach offers fine-grained observational support to grasp the diversity and the complexities of spatial work.

In more general terms, however, I embrace a more action-oriented approach to space, similar to that elaborated by Löw (2016). Here, material placing and the perception of spaces are at the core of understanding any spatial arrangement. In such a view, space is produced through action and in turn shapes the possibility of action. Löw classifies spatial practices as analytically falling into two major groups: *spacing* and *synthesis*. Whereas spacing points to the continuous (re)arrangement of different entities (humans and material objects), synthesis draws our attention to the connections that can be made between these entities. To fully understand the making of space, we thus must understand the processes of synthesis that are deeply entangled with "processes of perception, ideation, and memory"

(L w 2016, 135). Making space and reordering space thus always means not only reassembling human and non-human entities, but also redistributing agency and recreating networks and meanings that can be stabilized—and it is a deeply political activity, displaying and reconfiguring power relations.

Modes of ordering post-Fukushima environments: A laboratory always in the making

Below, I will look into three different modes of reordering space after the Fukushima Daiichi accident: map making, material demarcation practices, and displacement. Whereas the first mode focuses on representational practices, attentive to how space is perceived in new ways, the latter two modes are more concerned with material spatial practices, how people (are made to) experience space. Each of these modes will begin with a short ethnographic narrative from the field visit to capture the situatedness of these spatial practices, the imaginations attached to them and some of the complexities and emotions at stake. Furthermore, questions of accessibility, (re) appropriation, control, and the means to produce space will be discussed.

MAP MAKING

In the spring of 2013, a meeting on medical perspectives on the Fukushima disaster brings me to the offices of the International Atomic Energy Agency (IAEA) in Vienna, where approximately 25 people gather for a meeting. They are medical doctors from Fukushima University who cared for injured and traumatized people immediately after the disaster; they are radiation specialists, but they are also historians, ethicists, and science and technology studies scholars. We are receiving first-hand information on the aftermath of the disaster and the challenges that this posed, always asking what can we learn from it. As the talks continue and the slides flip by, I realize that most of them start with maps showing Japan, Fukushima, and the “problem space.” They are generally colored, with red pointing to the most dangerous spaces, orderly representing and differentiating the zones that needed evacuation, the borderlands that had an unclear status between unproblematic and potentially in danger, and the zones declared as unproblematic. Brian Wynne’s (1993) paper on sheep farmers after Chernobyl and the difficult relationship between radiation maps and farmers’ real-world experiences immediately crosses my mind. I would see many more of these maps during this meeting and its follow-up conferences in Vienna and Fukushima.

Watching the different maps pass by, I begin to compare them with the many others I had seen in newspapers, on the Web, and on

(Continued)

television. There are significant differences: they tell different stories about space and radiation, about actions to be taken, about solutions to be found.

I ask myself: Who had the power to draw these maps? Based on what knowledge? Were they widely embraced or contested? What would they mean to people in their everyday lives? Would the maps represent their reality, or simply provide an illusion of clarity in a sea of uncertainty?

As these thoughts continue, so do the slides, desperately reiterating a specific form of spatial reality and a specific kind of order and control, convincing the observer that the problem must be addressed exactly within and through spatial reordering.

How should we think about maps and their work in the aftermath of the Fukushima disaster?

When looking at the proliferation of diverse maps produced and circulated, it becomes obvious that we are observing the creation of a new geography of the “the nuclear.” Before looking at the different types of maps that we encounter, it seems essential to remember Anderson’s (1991, 178) argument that maps do not merely represent what already exists objectively. Instead, “a map anticipate[s] spatial reality, not vice versa. ... a map [is] a model for, rather than a model of, what it purported to represent.” Maps are made to do work. Using the physical definition of work as “the application of a force through a distance, and force is an action that one body exerts on another to change the state of motion of that body,” Wood (2010, 1) argues that maps “apply social forces to people and so bring into being a socialized space. The forces in question? Ultimately, they are those of the courts, the police, the military. In any case, they are those of... authority.” Maps—and, in particular, the radiation maps that I will address—are intended to be such an authoritative voice through which political actors can speak. They achieve this status through performing what Daston and Galison (1992) have termed “mechanical objectivity:” they pretend to represent a reality in the world objectively. Nevertheless, we also know that maps are necessarily always partial, challenged, and temporal. Moreover, maps never only refer to specific spatial arrangements and the radiation measures attached to them; they often also suggest and legitimize specific actions to (not) be taken: avoiding spaces, not being allowed to enter spaces, being safe in spaces, having to return to spaces.

What types of maps circulated in the post-Fukushima disaster period? What representational techniques did these maps use, and what work were they meant to perform in this crisis situation? The spectrum of the type of work they would have to do was broad: some were supposed to inform;

others prescribed actions for those who should “rationally” behave in an evidence-based manner; still others focused on a classificatory task, forming the basis of policy choices over which territories would be defined as inhabitable and to what degree; and finally, projective maps attempted, through modeling, to show where the radioactive plume would move, and when and how much radiation change could be expected. A single map could sometimes fulfill several of these tasks at once; different communication techniques were used and multiple versions of reality (Law 2011) were produced simultaneously. I will reflect on three types of maps that were meant to do specific work. In addition, I will shortly reflect on how they took meaning away from preexisting maps.

The **first group of maps** can be characterized by the importance attributed to the **distance of any location** from the power plant; such maps proliferated in the first days after the accident. They were the first indications of potential radiation intensity and were primarily intended to instruct people to behave in specific ways. They always showed concentric circles with the Daiichi nuclear power plant at the center, assuming that—as a first estimate—this could represent the idealized spread of radiation. This type of map redefined space through its distance to the power plant, dividing it into circular zones starting with an evacuation zone from which people had to leave urgently and would probably never be able to return, to zones where people were advised to seek shelter and stay inside, to territories beyond the circle that were defined as safe. Whereas the maps would look (one way or another) quite similar to the one presented in Figure 7.1,⁴ they supported somewhat different policy recommendations for the populations. On March 11, 2011, the Japanese government began to evacuate a three-kilometer circular zone around the power plant. This was expanded to a 20-kilometer radius the next day, after the explosion of the first nuclear reactor unit. On March 15, in the zone between 20 and 30 kilometers from the power plant, people were asked to stay in their houses and be prepared to evacuate, if necessary. They were evacuated on March 25, 2011. Therefore, these maps did change the ontological status of both specific spaces and the people inhabiting them, depending on their physical distance from the NPP. Farmlands and villages were redefined as irradiated zones, and it was recommended in different ways that these areas no longer be inhabited. People were reconceptualized as “at-risk-subjects” expected to behave in specific ways.

It is revealing to observe in comparison the communication of the US Embassy in Tokyo concerning the evacuation of their citizens from the zone around the Daiichi power plant.⁵ Under the heading “Estimates of Possible Exposure Define U.S. Evacuation Zone”, the *New York Times* reported shortly after the accident that:

[t]he American Embassy recommended on March 17 that Americans within 50 miles of the Fukushima reactors evacuate. The recommendation

was based on an analysis by the Nuclear Regulatory Commission that predicts possible radiation levels assuming conditions at the plant degrade. It is not based on current radiological conditions. It includes factors such as whether containment vessels remain intact and weather patterns, among others. Here are the results of the analysis on March 16.

This short news item is of interest for several reasons. It clearly addresses the multiple uncertainties involved in any type of map (weather patterns, the developments at the reactor), it figures in quite some detail possible radiation exposure in numbers with accompanying consequences for human health, it points to the population that is currently living in these spaces, and above all, it has a clear temporality embedded in that it anticipates potential futures not yet admitted as possible by the company running the plant (see Figure 7.1).

Thus, “thinking in circles” was a first quick way to define some spaces as no longer inhabitable, others as a nuclear borderland with an unclear status that could quickly change, and still others as safe. Therefore, these latter spaces could keep their ontological status as a “safe space,” whereas others had to change their status. The distance from the location of the accident became the key element in most of the first assessments. All of this activity embodied a clear temporal uncertainty that was embedded but not yet spelled out. What these neat circular maps had not captured was that weather conditions had changed, with winds causing radioactive particles to spread northwest on March 15.⁶ Communities far beyond the zones observed as in danger were thus exposed to large doses of radiation without being warned (e.g., Iitate). This very clearly shows the difficulty with this type of map. A map, ideally considered as a stable guide for action, had become fluid, continually changing as the accident and the weather developed.

These circular maps quickly were complemented by more detailed radiation maps (see Figure 7.2) intended to represent “the real” spread of radioactive material. Although they often maintain the circles to show the physical distance from the location of the NPP, they use specific color codes for the radiation intensity—generally from red, orange, and yellow to different shades of green and blue—to define the status of spaces. Distance from the power plant (with the exception of very close vicinity) clearly can no longer significantly describe the radioactive contamination of space. We thus observe an effort to coproduce maps, risk, and territory in the same move (November, Camacho-Hübner, and Latour 2010), and in the end, to entangle through them social orders and knowledge orders in significant ways (Jasanoff 2004). These maps show small color-coded pixels that indicate the radiation a person would be exposed to when staying in this place for an hour. The colors cut the radiation intensity spectrum into distinct intervals that somehow correlate with the risk that a person would assume when staying in specific spaces. The outcome of this representational exercise is space on a map classified as either more or less risky. This, in turn, creates the expectation that the map provides an objective basis for both human action and policy intervention.

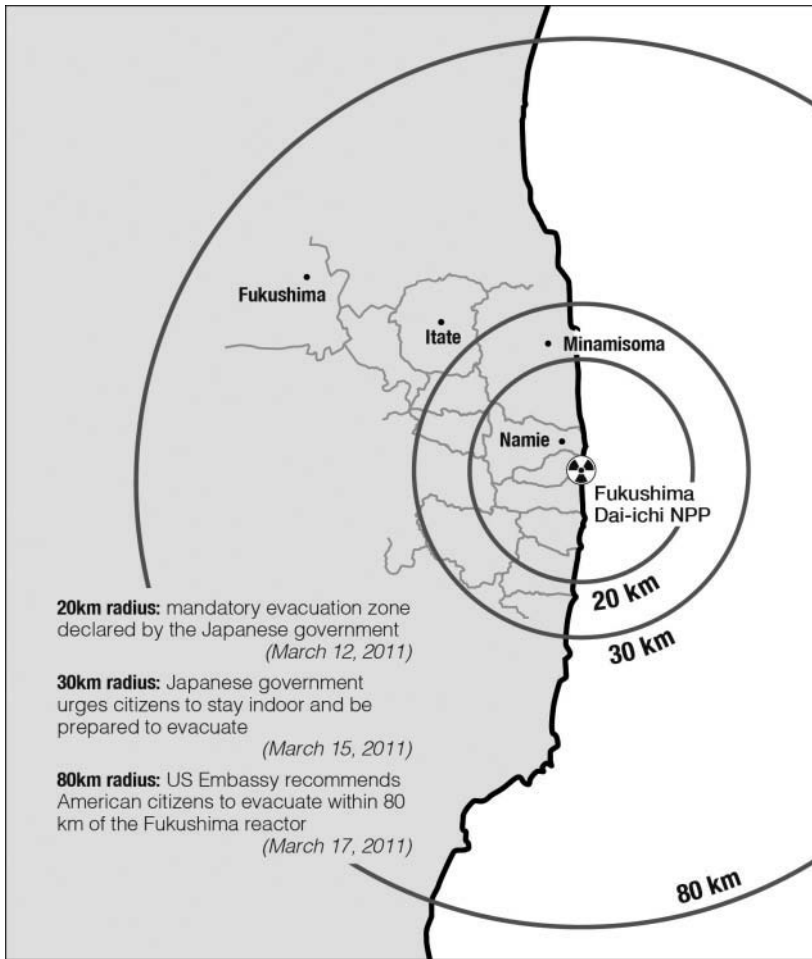


Figure 7.1 Typical map published after the Fukushima accident.⁷

As time passes, maps have to change. As this occurs, maps' dependency both on the most recent data and on their trustworthiness becomes increasingly obvious. These maps are always connected to numbers that participate in giving meaning to something that had either no name or that had a totally different meaning before the disaster redefined it. This meant that the definition of space was supposed to be simultaneously rigid (to call for sometimes drastic action, such as leaving one's home) and fluid, causing serious difficulties for people to act based on these maps. Particularly in the early phase, radiation could shift at any moment because of wind or rain and be unequally distributed across any given space represented by a colored pixel on the map. Thus, these maps materialized a specific

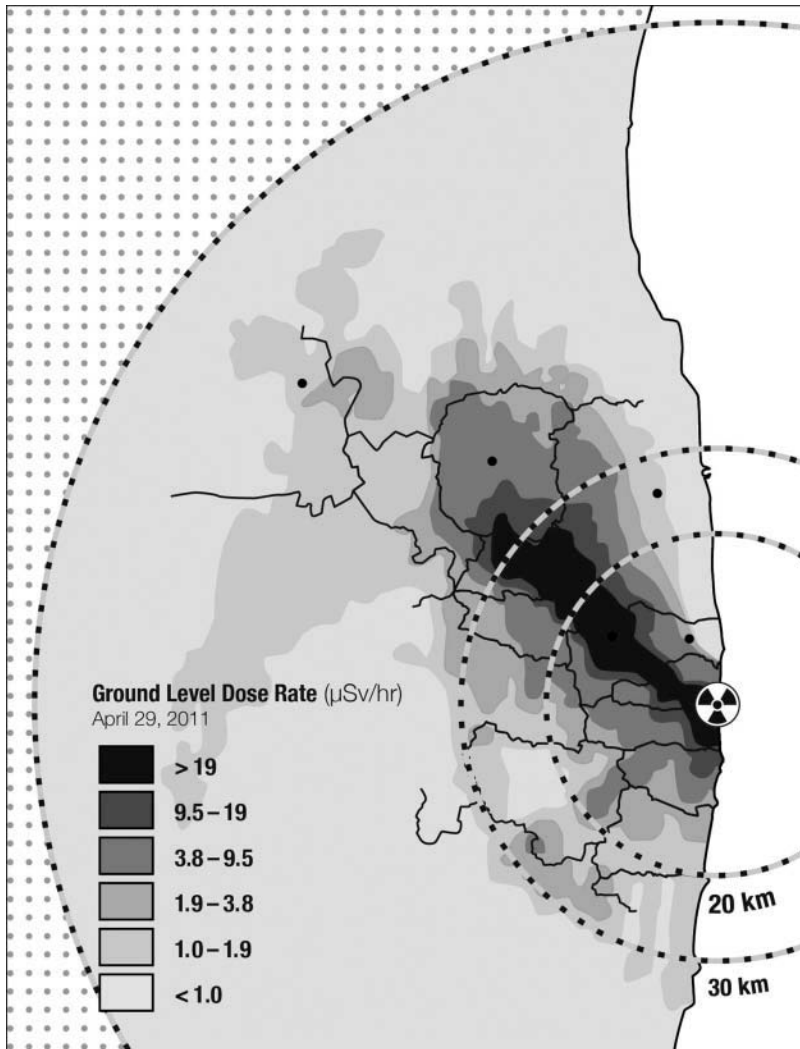


Figure 7.2 Radiation map⁸.

dimension of the disaster: the space people lived in was transformed into evacuation zones, the homes that meant safety for them were transformed into places that endanger their health and lives; furthermore, a new type of temporal imaginary was superimposed on their space, one that would allude to when, if ever, people could hope or would be obliged to come back. This latter dimension transformed the map, which previously had been largely perceived as a stable representation, into a continuously changing one; it never said anything about the now, but always something about the

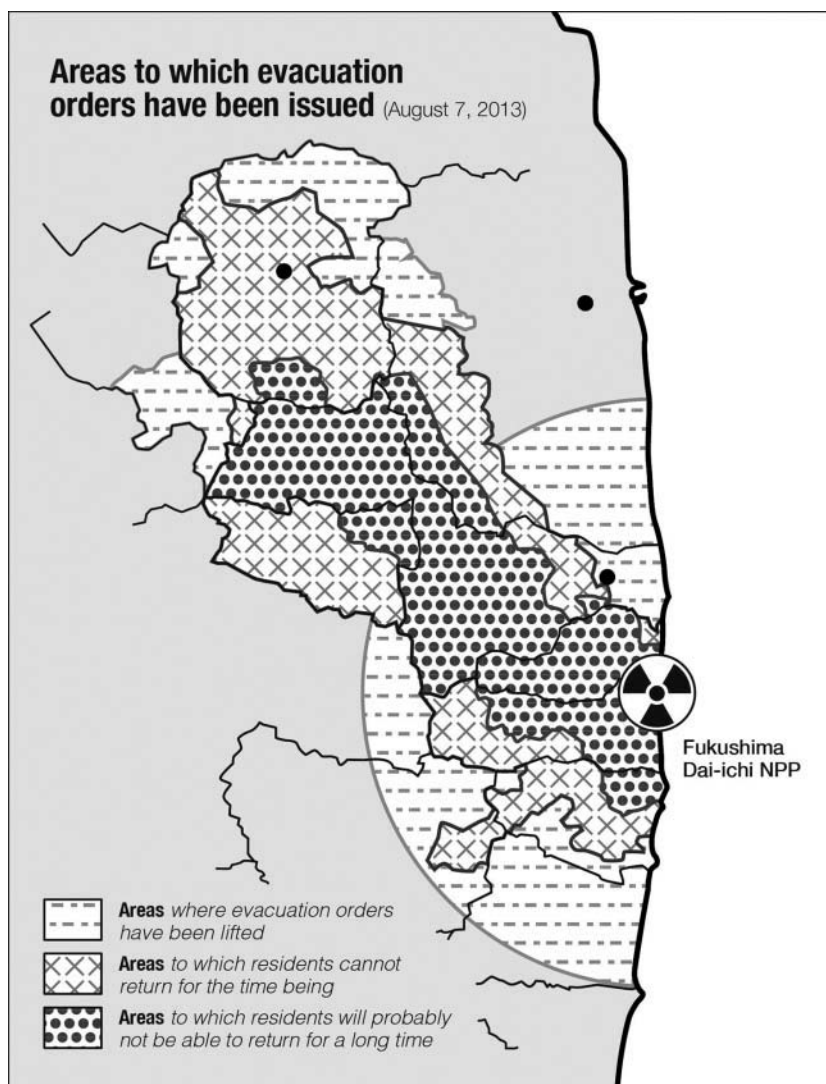


Figure 7.3 Map identifying the status of the territory⁹.

past and potentially about the future. Intended to promote certainty and to represent a stabilized order, maps unwillingly became part of a narrative of change and uncertainty.

This is particularly true for the map shown in Figure 7.3. It is a “translation” of radiation maps into explicit policy maps (by the appropriate Ministry), classifying territories according to their status of being an evacuation

zone that people were forbidden to enter; zones classified as showing different degrees of high radiation expressed through shades of orange and yellow; and green zones, which could be inhabited. As with the radiation maps in Figure 7.2, this map very nicely shows how the perception of what is inhabitable territory changed over time, not always in the direction of handing space back to people: sometimes territory that was considered safe turned out to have excessively high radiation levels. We will pick up this point later.

Who makes these maps?

Soon after the first shock, this actually became a key question. As trust in the state authorities and the nuclear industry started to deteriorate, not only did a strong anti-nuclear protest movement emerge, but also a growing number of individual and collective initiatives became engaged in collecting and sharing radiation data and constructing their own alternative maps. Plantin (2014) describes these activities in some detail, pointing to the increasing importance of new information and communication tools in the Fukushima disaster, and how these maps came to have significance in the complex experiment of orchestrating “the nuclear” and “the social” through these maps. We thus witness numerous actors—from engaged individuals to scientists and hackers to web-industry companies such as Yahoo!—who used “mapping technologies to take part in the assessment of the nuclear situation in the country, creating alternative resources of information that competed with official reports of the government and the Tokyo Electric Power Company (TEPCO).” (Plantin 2015) People either used their own data or combined different data to develop new protocols of how to order and represent data in the form of maps and to change the rules of this form of experimentation by delivering their own interpretation of space and its inhabitability. (see e.g., Saito and Pakh 2016) New media played an essential role, as did alternative actors who developed information. The authority of the maps thus was constantly upheld through powerful performances while being contested in multiple places and practices. Numerous personal Geiger counters were used to create alternative maps, showing that the official maps did not match people’s measurements. Official maps suddenly no longer appeared as robust objects, but instead were subjected to a comparative gaze between the participatory maps created on the Web and those offered by the government and TEPCO. These radiation maps could participate in the controversies concerning the assessment of the relation between radiation intensity and place. These counter-maps enabled a different interpretation, showed the variations within seemingly homogeneous territories, and allowed the creation of a shared understanding. Although I had heard about these maps during my two stays in Fukushima, they always remained invisible in official discourses. They were neither used on slides nor were they printed; they seemed to exist solely on the Web.¹⁰

Finally, it is essential that in one way or another, all the new maps overwrote preexisting ones. Indeed, before the disaster struck the region, the space in the Fukushima province had been imagined and practiced in specific, culturally entrenched ways. The map in Figure 7.4 is one such example. It stands in the center of Minamisoma, reminding the viewer that this space around the city had once been imbued with a very different meaning: it was used for leisure activities, with hiking, sunbathing, and surfing somehow cohabitating with the nuclear power plant approximately 30 kilometers south. With the radiation maps entering the arena, many of these places had experienced an ontological redefinition from a leisure zone to a risk zone.



Figure 7.4 Public map in Minamisoma (showing leisure activities in the region).
© Ulrike Felt 2013.

MATERIAL DEMARCATION PRACTICES

On a gray November day in 2013, I am driving down the road that many Japanese people had taken in the opposite direction when fleeing the invisible danger of radiation in March 2011. It was virtually the only road that had remained intact after the tsunami in what would be later named “the evacuation zone.” I am told that people were stuck here in endless lines of cars, taking only the most important things with them—probably still hoping that this was only temporary. I try to make sense of what I hear, to capture my many small impressions, to imagine what this meant to people.

Occasionally, we see large Geiger counters in public spots. We are told that many people bought personal Geiger counters because they were unsure if they could rely on the maps. There were “hot spots” with high radiation, even within the territory that was declared inhabitable on the map. We also have our own radiation-measuring device with us.

As we drive towards Namie, a town inside the evacuation zone, we pass fewer and fewer cars. The road transforms into a corridor; to the left and the right of the road are barricades, and sometimes guards ensure that no cars trespass. Cars and boats are in the middle of the fields, with grass grown over and through them; they seem to have become an integral part of the post-tsunami landscape. Their “being out of place” is just one of the many signs of disruption.

We stop and walk around, passing houses that were empty, displaced by the floods or torn apart. An entire wall of one of the houses had been swept away by the power of the tsunami, showcasing peoples’ private space, their personal belongings—all covered with dry mud and dust. We drive on until we arrive at a control point. No entering the evacuation zone without a permit, we are told. I would return a year later and witness the first acts of decontamination of the empty town of Namie, representing an effort to redefine the status of a town’s territory.

Whereas the maps were intended to represent and classify space as well as to rearrange it in zones, material demarcation practices were intended to implement or impose this classification, while also partially challenging the neat order presented by the maps. To understand better the spatial work of demarcations, I will use three moments during which this boundary drawing work becomes visible and the meaning of space is experimented with and explored in the struggle of making sense.

I first noticed the spatial work because of the growing density of signs as we approached the evacuation zone. Little flags in the empty fields indicate that territories have been decontaminated. Drawing closer to the evacuated space, the open road we drive on gradually turns into a corridor with

barricades to most of the roads that enter the territory on the right and left sides. Sometimes, as described above, there are guards ensuring that these barriers are respected (Figure 7.5). We drive by rows of abandoned houses and stores that were partially destroyed by the earthquake and the following tsunami. Sometimes cars are still parked in front of them, left behind. In the fields, we see boats positioned strangely next to houses or street signs indicating that schoolchildren would normally be crossing here (Figure 7.6), kilometers away from the water they had once been floating on, now “integrated” into the landscape with plants growing over them. The tsunami had left its traces, and the nuclear disaster had somehow kept these traces of devastation in place—time seemed frozen, even though we were told that this area had already been partially cleaned up. Too many items felt out of place. The order constituting a space as one to be inhabited by people seemed disrupted. Here we can observe that neither spacing nor synthesis (Löw 2016) worked to make sense of the space; the elements we observed seemed neither to be where they should be, nor were they able to be connected in any meaningful way beyond stating that this was a devastated land. Moreover, the corridor upon which we drove had been transformed from being a main road to being just a temporary crossover; a line is drawn through a space but not a road into an inhabited and inhabitable space (Figure 7.5).

The second moment of visible spatial work was when the control points materialized the borders of the evacuation zones, transforming the lines on the map into clear-cut physical demarcations (see Figures 7.7 and 7.8). The moment the Japanese government had decided to identify evacuation zones, people had to leave these areas and nobody would be allowed to enter them without special permission (moreover, if permission was granted,



Figure 7.5 Cross-road closure on the way to the evacuation zone (corridor road).
© Ulrike Felt 2013.



Figure 7.6 A boat in the middle of the plain in front of a sign “schoolchildren crossing”. © Ulrike Felt 2014.

it was only for a limited time). From one moment to the next, people no longer inhabited this space, and it was handed over to “the nuclear” with an unclear perspective of whether it would be possible to recover (Figures 7.7 and 7.8).

Although we had already driven through nuclear borderlands with relatively high radiation, crossing the checkpoint was intended to signal a more profound change in space. We moved into a space formally identified as an evacuation zone. Holding a permit would give us access, and we continued on our road to Namie, a town that had been left by its nearly 20,000 inhabitants on March 12, 2011, because it is located well within the 20-kilometer exclusion radius around the damaged Fukushima Daiichi nuclear power plant. The city stood like an icon for the abrupt change of ontological status of the space of which it is a part, a lively inhabited area at one moment and left to the invisible radiation the next. Walking through the empty streets, it was the silence that was the most striking. People were missing, but there were also no sounds of animals such as birds. Abandoned houses were partly destroyed by



Figure 7.7 Physical demarcations: Border control at the evacuation zone. © Ulrike Felt 2013.



Figure 7.8 Physical demarcations: Radiation control barriers. © Ulrike Felt 2014.

the earthquake; there was a train station where people once commuted that had been taken over by all types of plants; there was a bike rack next to the station filled with bikes that nobody will ever pick up again; a child's toy was lost near one of the bikes; there were abandoned vending machines that contained drinks. (Figures 7.9 and 7.10) An orange blinking streetlight is actually the only sign that would allow the assumption of an inhabited space. Leaving the space exclusively to “the nuclear,” containing it there by monitoring the border, continuously measuring and controlling it, while it would hopefully decay to a degree that would make cohabitation possible was seen as the sole action one could take. In one part of town, the first decontamination efforts started, washing away the nuclear fallout from walls and roofs, collecting the radioactive leaves, taking away the upper level of the earth and putting it into plastic bags in an effort to “evacuate” the nuclear from this space.

Later, we would also have to pass a second type of control point at which radiation was measured on the car, our shoes, and our clothes when moving back to inhabited zones (see Figure 7.8). It felt as though this was about not unknowingly taking the radioactive material across “the border” (Figures 7.9 and 7.10).

The third spatial work was not only performed through the public use of Geiger counters, which were put in place to monitor radiation in the various regions, but also used by the many citizens who had bought radiation counters for private use. These measuring devices became the material



Figure 7.9 Namie City – abandoned to “the nuclear”: One of the empty main streets.
© Ulrike Felt 2014.



Figure 7.10 Namie City – abandoned to “the nuclear”: Abandoned bike rack at the train station. © Ulrike Felt 2014.

incarnation of the regime of counting and accounting that had been established, of producing the data that would be an essential basis for defining the ontological status of the concrete space one inhabited. Whereas the maps would bring together official aggregated measurements to represent “the region”—as represented in the two maps in Figures 7.2 and 7.3—people soon realized that the homogeneity performed across any of the areas was an artifact of how these maps were made. The spatial representations produced on radiation maps did not account for the multiple, local variations of variable radiation and denied the differences between farms or houses within one purportedly homogenous region. Therefore, the claims maps would make often did not reflect the micro-realities that were shown through the local use of Geiger counters. Spaces that had been defined as safe and were located in green zones turned out to show high radiation measures, at least in part. Zones initially defined as dangerous turned out to be less irradiated. Therefore, people were very much left to their own judgment, their own measurements. These observations tie into discussions on self-experimenting societies or living in a risk society because citizens are required to take action and decide whether to accept the maps (Figure 7.11).¹¹



Figure 7.11 Public Geiger counter within an inhabited zone showing 0.52 microSievert/h. © Ulrike Felt 2013.

COLLECTING, DISPLACING AND CONTAINING RADIOACTIVE SOIL

On our way back to Fukushima City, we drive miles through former agricultural territories, often rice paddies—evidently, no longer cultivated. I watch power shovels, busy placing earth into black plastic bags. Workers in overalls with masks over their mouths and noses seem to coordinate these efforts. We drive by pyramids of bagged radioactive earth. It is explained to us that it had been decided to remove the upper layer of the radioactive soil and to cut grass and some trees to decontaminate the land and make it again usable for agricultural purposes. It is unclear yet—the narrative would continue—where this radioactive earth would be taken and whether these measures would actually enable the area to return to “normal.” What an exceptional experiment, I think. Who is orchestrating this large-scale redefinition of space? And for whose good? Who are the people standing there doing the work? Where would the bags go? Questions begat questions.

I am puzzled by the idea that the problem would go away by scratching off the upper layer of the contaminated soil. Nevertheless, this is proposed as the only possible action to return life to this agricultural region. My efforts to capture and somehow order all of the impressions

and information and to understand the spatial choreography that I had experienced throughout the day in the field, away from a world ordered by PowerPoint presentations and expert conversations, started to overwhelm me. I scribbled some notes in my booklet, hoping that they would capture at least parts of what crossed my mind. Then I started simply staring out of the window, silent as the power shovels and piles of plastic bags moved by.

The third type of spatial reordering that is part of this huge containment and recovery experiment is performed through the extensive efforts made to *displace* the radioactive soil that covers the region around the NPP. Through the radioactive contamination of the ground, space that was once inhabited or was valuable agricultural land—the latter being essential for this part of Japan—was now regarded as uninhabitable and not to be cultivated. In this context, radioactivity and contaminated soil could be classified, following Mary Douglas (1984), as “matter out of place.” Connecting to what has been discussed in the section on mapping, we can see how “systematic ordering and classification of matter” performed through radiation maps and related measurements “involves rejecting inappropriate elements,” and as a consequence, soil showing radioactivity beyond a certain level was to be displaced. Radioactive contaminated soil, as the dirt in the case of Douglas, becomes a subject that demands thinking across different scales of evidence. Radioactive earth has thus to be seen from the materiality in everyday life, over the language we use to speak about it (e.g., through numbers that stand for the radiation intensity) to the cultural symbolism developed around it. Although there is no space to develop all of these facets of matter out of place, it is nevertheless essential to keep them in mind. What counts as “being in place” or “out of place” is always linked to an instantiation and disruption of a shared symbolic order. The nuclear has been in the region for many decades, but had always been imagined as confined to and controlled by the power plant—it was “in place” there. Now that the nuclear had left the confinement of the power plant and invaded spaces where it should not be, work would be needed to reinstall order. Reading what has happened through Douglas also makes us aware that what we label as being out of place is by no means a stable and clearly defined category: it is always situated, under negotiation. How much radiation is acceptable to classify a space as agricultural or as inhabitable is therefore closely linked to which type of order seems possible and acceptable within certain limits.

Indeed, after the accident and with the commencement of decontamination efforts, ten million bags have been filled with highly contaminated radioactive soil and vegetation weighing approximately one ton each. Right now, in many cases, the bags sit as icons of the disaster next to the places where they had been collected (Figure 7.12). Others were brought to huge



Figure 7.12 Radioactive soil collected in plastic bags. © Ulrike Felt 2014.

collection sites, which are considered as a temporary solution—it has remained unclear where this “matter” could have its place in Japanese society. When Tropical Storm Etou hit Japan in September 2015, several news outlets reported that the floods had taken the bags with radioactive soil and grass with them, showing that the idea of containment was still tentative and fragile.

This spatial reordering is perceived with a high level of ambivalence by the region’s citizens. On the one hand, it creates the hope of redefining contaminated territories as agricultural land, as land ready to be re-inhabited or as courtyards where children can play. Critical voices, however, note that the government has pushed these efforts hastily to start moving people back to the region in early 2017. As an interviewee from the region was quoted in the *New York Times* in August 2015, “The government just wants to proclaim that the nuclear accident is over” (NYT, 8/8/2015). This effort thus could be read as not only an effort of decontamination, but also an effort to signal that the sociotechnical imaginary of containment and control of the nuclear had been reenacted through the massive displacement of irradiated material. The question remains: To where? (Figure 7.12)

Discussion and conclusion

The chapter began by examining the efforts after the Fukushima Daiichi nuclear accident as a real-world experiment. The aim was to observe the efforts to uphold and to restore the disrupted sociotechnical imaginary of control and containment of the nuclear by working on the continuous rearrangement of space. I showed a laboratorization of the region around the

power plant, with spatial rearrangements being the key practice to install a controllable inside. The analysis has been gravitating around three spatial practices that were key in the aftermath of the Daiichi NPP accident. Maps begin to link space to radiation, people, and potential actions in continually new ways. These linkages were achieved “by bringing together onto a common presentational plane [rather diverse] propositions about territory” (Wood 2010, 1–2). The propositions that normally are important in making maps have changed in the aftermath of the nuclear accident: it was no longer the roads or the leisure activities that were essential, but rather the radioactivity measures. Maps thus attempted to exercise the power to (re)define the ontological status of a space: as inhabitable, as a temporally problematic site (to varying degrees), or as safe. Second, we witnessed how these maps were materialized through control posts and entry permits. However, these material demarcations were also challenged by localizing the measurements with private Geiger counters, thus materializing variations within seemingly homogenous spaces, i.e., drawing new borders within. Finally, the displacement of radioactive soil was a third type of intervention in the nuclear order that had emerged. This intervention appeared more complex than imagined and involved the question of where to displace this matter that was ontologically perceived as out of place.

What can we learn from the analysis offered? What challenges are we confronted with in experimental societies?

First, investigating what happened in the aftermath of the meltdown at the Daiichi power plant as a real-world experiment sensitizes us to a more fine-grained analysis of how spatial politics matters. Understanding the notion of the laboratory as the outcome of processes of successfully separating the controlled inside (and in our case, an outside that does not need to be controlled), we observed a continuous making, unmaking, and remaking of the boundaries of the laboratory. This continuous shifting of the laboratory’s boundary, and with it, the shift in the possible experimental set-ups and the different understandings of space attached to it, then defined the various experiences of experimentation that were co-present and the various types of learning that were occurring simultaneously but were not necessarily coordinated. This moves us away from simply seeing what happened as a rational way to act in the light of a nuclear disaster, to focusing our attention to the efforts of maintaining power and control over the “site of experimentation” in such a situation, and to understanding the degree to which this is relying on the capacity to shape the production and distribution of space. Therefore, living in a self-experimental society calls for closer consideration of the distribution and classification of space, for posing the question of who holds the power to redefine space, and for considering the values and agencies tied to spatial orders.

Second, investigating the post-Fukushima disaster from this perspective also enabled a better understanding of the ontological politics at work in this experiment. Ontological politics (Mol 2002) helps us not only to account for

the sociomaterial enactment of reality (i.e., how spaces of a specific type are brought into being) but also to make us aware that any enactment of reality is always entangled with political choices. Therefore, any reality always builds only a specific set of sociomaterial relations while ignoring or disregarding others. Acknowledging that ontologies are always plural and that our practices shape them in important ways makes us aware that we must better understand how institutional actors in the aftermath of nuclear accidents and other complex disasters embrace some ontologies and not others. Being attentive to the ontological politics around the classification of space thus becomes an important element to not only describe what happened after the nuclear accident as a relatively well performed crisis management, but also to examine the options and choices made when performing specific spatial practices. This is particularly visible at the time this chapter was written, as we witness citizen groups' battles against the government attempting to force people to return to areas that have been relabeled as safe (enough). Acceptance of this reclassification of space would enable the government to declare closure to the disaster, and would be another example of how space matters in the exercise of power in technoscientific worlds.

Third, the study allowed the witnessing of the creation, rehearsal, and continuous adaptation of a new sociotechnical imaginary of containment and control (Felt 2015). In that sense, containment was no longer a fixed category, as imagined in most of the classical texts on nuclear risk assessment. Instead, the continuous redefinition of space enabled a more fluid handling of the notions of containment and control. Once the nuclear had moved out of the confinements of the reactor, new types of spaces were ceded to it, evacuating people and hoping that the nuclear could be temporarily contained in these new spaces and controlled through continuous measurement. These continuous spatial adaptations thus played an essential role in attempting to restabilize the nuclear imaginary—therefore moving to the core of the real-world experiment. This focus on redefining the territory where people had their houses and spent their lives has, however, rendered less visible the fact that the seawater is still continually contaminated with radioactive material that is gradually expanding around the globe. There is no imaginary of containment available here, no fixed territories that could help remedy the lost imagination of control.

Finally, as with every experiment, questions of responsibility and the distribution of power need to be asked. We need to go back and see that scientific and technical advances made promises to the region: beneficial development for all. However, these advances obviously also brought about new types of uncertainties and failures. We could therefore ask whether in a self-experimental world we should leave questions of risk management solely to technical experts and their predictive tools—i.e., to “technologies of hubris” (Jasanoff 2003) that promise command and control over technology. The case of Fukushima reminds us that predictions of potential risk are not only about the nuclear, but also about people's lives and how they can be

lived in a situation of disrupted trust and profound uncertainty. We could continue with Jasanoff (2003, 225) and ask,

Has our ability to innovate in some areas run unacceptably ahead of our powers of control? Will some of our most revolutionary technologies increase inequality, promote violence, threaten cultures, or harm the environment? And are our institutions, whether national or supranational, up to the task of governing our dizzying technological capabilities?

These are all questions of core importance for my argument. Upholding the sociotechnical imaginary of nuclear containment, control, progress, and development came at the price of displacing people, of excluding them from the spaces to which they had been attached and that are now occupied by the nuclear, of bagging radioactive earth and grass that is buried somewhere else. Next, the question might be about developing “technologies of humility” that complement “the predictive approaches: to make apparent the possibility of unforeseen consequences; to make explicit the normative that lurks within the technical and to acknowledge from the start the need for plural viewpoints and collective learning.” (ibid.) Living in an experimental society then would mean calling for more collective forms of experimentation, for including different types of expertise in implementing the spatial choreography we witnessed; thus, to open up debate on whether we want to continue living in an economy of technoscientific promise (Felt et al. 2007) of which the crossroad sign in Futaba was a silent witness: “Nuclear Power: Energy for a brilliant future.”

Notes

- 1 For a collection of five stories from evacuated citizens capturing their situation, see: <http://www.greenpeace.org/international/en/campaigns/nuclear/safety/accidents/Fukushima-nuclear-disaster/fukushima-dont-forget/>.
- 2 For webspaces collecting information on Fukushima for sharing and discussing the disaster, see <https://fukushimaforum.wordpress.com/online-forum-2/online-forum/making-a-case-for-disaster-science-and-technology-studies/>; or <http://teach311.org/>.
- 3 Different versions of this chapter have been presented at several occasions over the past two years: the workshop “Technological Visions and Revisions,” STS-Harvard, April 4, 2014; the EASST conference in Torun, September 17–19, 2014; and the workshop “New technologies as Social Experiments,” TU Delft, August 20–23, 2015. I always received valuable feedback, which was tremendously helpful in sharpening my argument. I visited Fukushima Medical University in the framework of two international conferences co-organized by IAEA, and had the privilege to participate in workshops at the premises of the IAEA in Vienna. My thanks, therefore, go to all those who shared their experiences during these meetings and give me the possibility to see the complexities at work. For a documentation of these conferences, see: <http://fmu-global.jp/>

- workshop/symposium/fmu-iaea-international-academic-conference/ and <http://fmu-global.jp/workshop/symposium/25-27-july-2014-2nd-fmu-iaea-international-academic-conference-convened-in-fukushima/>. The field trips referred to in this chapter were organized outside the conferences. For the second field trip from Fukushima City to Namie, my special thanks go to Shineha Ryuma, who not only organized it, but also shared his very personal insights into what happened after the nuclear accident.
- 4 http://www.nytimes.com/interactive/2011/03/16/world/asia/japan-nuclear-evacuation-zone.html?_r=0.
 - 5 For a reflection on the different evacuation recommendations, see Kyle Cleveland, *Mobilizing Nuclear Bias: The Fukushima Nuclear Crisis and the Politics of Uncertainty* <https://fukushimaforum.wordpress.com/workshops/sts-forum-on-the-2011-fukushima-east-japan-disaster/manuscripts/session-4a-when-disasters-end-part-i/mobilizing-nuclear-bias-the-fukushima-nuclear-crisis-and-the-politics-of-uncertainty/>.
 - 6 http://www.meti.go.jp/english/earthquake/nuclear/roadmap/pdf/20120723_01.pdf.
 - 7 This map has been modeled after many of the circulating maps. For a concrete case of a quite sophisticated “circular map” published in daily newspapers, see, <http://www.nytimes.com/interactive/2011/03/16/world/asia/japan-nuclear-evacuation-zone.html>.
 - 8 Schematic representation of a typical radiation map circulating after the Fukushima Daiichi accident; for one of the many color versions circulating, see for example, http://www.world-nuclear-news.org/rs_radiation_declines_at_fukushima_0603131.html. Accessed: June 2016.
 - 9 Reworked version of a map published by METI, Japanese Ministry of Economy, Trade and Industry, <http://www.meti.go.jp/english/earthquake/nuclear/roadmap/index.html>. Accessed: June 2016.
 - 10 For an example of such alternative maps see the following webpage <http://safe-cast.org/tilemap/>.
 - 11 <http://www.nytimes.com/2011/10/15/world/asia/radioactive-hot-spots-in-tokyo-point-to-wider-problems.html>.

References

- Anderson, Benedict. 1991. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.
- Callon, Michel, Pierre Lascoumes, and Yannick Barthe. 2009. *Acting in an Uncertain World: An Essay on Technical Democracy*. Cambridge, MA: MIT Press.
- Daston, Lorraine, and Peter Galison. 1992. “The Image of Objectivity.” *Representations* 40 (fall):81–128.
- Douglas, Mary. 1984. *Purity and Danger. An Analysis of Concepts of Pollution and Taboo*. London/New York: Routledge.
- Felt, Ulrike. 2015. “Keeping Technologies Out: Sociotechnical Imaginaries and the Formation of Austria’s Technopolitical Identity.” In *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, edited by Sheila Jasanoff and Sang-Hyun Kim, 103–25. Chicago, IL: Chicago University Press.
- Felt, Ulrike, Brian Wynne, Michel Callon, Maria Eduarda Gonçalves, Sheila Jasanoff, Maria Jepsen, Pierre-Benoît Joly, Zdenek Konopasek, Stefan May, Claudia Neubauer, Arie Rip, Karen Siune, Andy Stirling, and Mariachiara Tallacchini.

2007. *Taking European Knowledge Society Seriously*. Luxembourg: Office for Official Publications of the European Communities.
- Felt, Ulrike, Maximilian Fochler, and Peter Winkler. 2010. "Coming to Terms with Biomedical Technologies in Different Technopolitical Cultures: A Comparative Analysis of Focus Groups on Organ Transplantation and Genetic Testing in Austria, France, and the Netherlands." *Science, Technology & Human Values* 35 (4):525–53. doi:10.1177/0162243909345839.
- Fujigaki, Yuko, ed. 2015. *Lessons from Fukushima: Japanese Case Studies on Science, Technology and Society*. Dordrecht: Springer.
- Gross, Matthias, and Wolfgang Krohn. 2005. "Society as Experiment: Sociological Foundations for a Self-Experimental Society." *History of the Human Sciences* 18 (2):63–86. doi:10.1177/09526951050541823V32.
- Guggenheim, Michael. 2012. "Laboratizing and De-Laboratizing the World: Changing Sociological Concepts for Places of Knowledge Production." *History of the Human Sciences* 25 (1):99–118. doi: 10.1177/0952695111422978.
- Harvey, David. 1990. *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change*. Cambridge, MA: Blackwell Publishers.
- Hecht, Gabrielle. 2009. *The Radiance of France: Nuclear Power and National Identity after World War II*. Cambridge, MA: MIT Press.
- Hecht, Gabrielle. 2012. *Being Nuclear: Africans and the Global Uranium Trade*. Cambridge, MA: MIT Press.
- Hecht, Gabrielle. 2013. "Nuclear Janitors: Contract Workers at the Fukushima Reactors and Beyond." *The Asia-Pacific Journal* 11 (1):1–13.
- Jasanoff, Sheila. 2003. "Technologies of Humility: Citizen Participation in Governing Science." *Minerva* 41 (3):223–44.
- Jasanoff, Sheila. 2004. *States of Knowledge: The Co-Production of Science and Societal Order*. London/New York: Routledge.
- Jasanoff, Sheila, and Sang-Hyun Kim. 2009. "Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea." *Minerva* 47 (2):119–46.
- Jasanoff, Sheila, and Sang-Hyun Kim, eds. 2015. *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*. Chicago, IL: Chicago University Press.
- Jones, Christopher F., Shi-Lin Loh, and Kyoko Sato. 2013. "Narrating Fukushima: Scales of a Nuclear Meltdown." *East Asian Science, Technology and Society* 7 (4):601–623. doi: 10.1215/18752160-2392860.
- Knorr-Cetina, Karin D. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press.
- Krohn, Wolfgang, and Johannes Weyer. 1994. "Real-Life Experiments. Society as a Laboratory: The Social Risks of Experimental Research." *Science and Public Policy* 21 (3):173–83.
- Latour, Bruno. 1993. *We Have Never Been Modern*. Cambridge, MA: Harvard University Press.
- Law, John. 2011. "Collateral Realities." In *The Politics of Knowledge*, edited by Fernando Dominguez Rubio and Patrick Baert, 156–178. London: Routledge.
- Lefèbvre, Henri. 1991. *The Production of Space*. Oxford: Blackwell.
- Löw, Martina. 2016. *The Sociology of Space. Materiality, Social Structures, and Action*. New York: Palgrave Macmillan.

- Mol, Annemarie. 2002. *The Body Multiple. Ontology in Medical Practice*. Durham, NC: Duke University Press.
- November, Valérie, Eduardo Camacho-Hübner, and Bruno Latour. 2010. "Entering a Risky Territory: Space in the Age of Digital Navigation." *Environment and Planning D: Society and Space* 28 (4):581–99. doi: 10.1068/d10409.
- Pfotenhauer, Sebastian M., Christopher F. Jones, Krishanu Saha, and Sheila Jasanoff. 2012. "Learning from Fukushima." *Issues in Science and Technology* 28 (3):79–84.
- Pickering, Andrew, ed. 1992. *Science as Practice and Culture*. Chicago, IL: University of Chicago Press.
- Plantin, Jean-Christophe. 2014. *Participatory Mapping. New Data, New Cartography*. London/Hoboken, NJ: ISTE Ltd./John Wiley & Sons.
- Plantin, Jean-Christophe. 2015. "The Politics of Mapping Platforms: Participatory Radiation Mapping after the Fukushima Daiichi Disaster." *Media, Culture and Society* 37 (6):904–21.
- Saito, Hiro, and Sang-Hyoun Pakh. 2016. "The Realpolitik of Nuclear Risk: When Political Expediency Trumps Technical Democracy." *Science, Technology & Society* 21 (1):5–23. doi:10.1177/0971721815627251.
- van de Poel, Ibo. 2011. "Nuclear Energy as a Social Experiment." *Ethics, Policy & Environment* 14 (3):285–90. doi:10.1080/21550085.2011.605855.
- Wood, Denis. 2010. *Rethinking the Power of Maps*. New York/London: The Guilford Press.
- Wynne, Brian. 1993. "Public Uptake of Science: A Case for Institutional Reflexivity." *Public Understanding of Science* 2 (4):321–337.
- Yamaguchi, Tomiko. 2016. "Introduction: Examining Lay Peoples Responses to Risk: Knowledge Politics and Technosciences in the Asian Context." *Science, Technology & Society* 21 (1):1–4. doi:10.1177/0971721815622730.

8 “Dormant parasites”

Testing beyond the laboratory in Uganda’s malaria control program

René Umlauf

Introduction

Nurse Norah: “I have heard that sometimes some people may take strong drugs that can make the parasites hide, and when you do the test, you are not able to find the malaria when it is actually there! The parasites become dormant but not destroyed. I also don’t know how that happens, but I have heard that it happens!”

Nurse Norah’s answer points to some of the complexities of practicing medicine in lower-level health facilities in Uganda. It specifically refers to the tricky question of if and how a patient’s practice of self-medication might affect the capacities of newly introduced Rapid Diagnostic Tests (RDTs) to detect malaria. In addition, the assumption that parasites can hide serves also to Norah as a justification for prescribing antimalarial medication despite a negative RDT result. From a public health perspective, Norah’s action is part of a widely noted problem of nonadherence (Homedes and Ugalde 2001, Vermeire et al. 2001). If she follows protocol, she should not administer antimalarials, but instead some other drugs, such as antibiotics or antipyretics to treat the presented symptoms. Or, she would refer the patient to the next higher-level facility with better diagnostic equipment to further narrow down the most appropriate treatment. By prescribing antimalarials instead, Norah risks delaying treatment of other potentially harmful diseases.

This ethnographic observation serves as an entry point to explore the relation between testing and experimentation as it increasingly becomes part of a broader framework of evidence production in Global Health interventions. I will focus on the introduction of Rapid Diagnostic Tests (RDTs) in lower-level primary health care facilities in rural Uganda.¹ One of the core objectives of RDTs is to identify malaria more accurately and thereby improve the use of antimalarials. RDTs are expected to base the prescription of antimalarial drugs on parasitological evidence only. Health workers are only authorized to prescribe antimalarials if RDTs prove the presence of plasmodia. If the test result is negative, health workers should not give out antimalarials but instead follow up with additional examinations that

will eventually provide a differential diagnosis. As I will show throughout the paper, being able to follow this protocol is based on various implicit assumptions regarding therapeutic context as well as understanding of the disease itself.

While RDTs are considered an innovative technology for diagnosing malaria, they are not the only way of identifying the disease. Throughout the Ugandan health care system, we find at least two more procedures, namely *presumptive treatment* and *microscopy*.² Before the introduction of RDTs, presumptive treatment of malaria was the most common mode of dealing with fever or symptoms that resemble malaria. Either as self-diagnosis at home or clinical assessment in health centers, the presumptive administration of antimalarials constitutes a vital part of this pragmatic procedure. When symptoms resolve, it can be assumed that malaria was the cause; when symptoms prevail, patients and health workers can at least rule out that malaria is the cause of the ailment.

From a public health perspective, it is assumed that RDTs can substitute the practice of presumptive treatment in all lower-level facilities where no laboratory service is available. Moreover, the testing inscribed in RDTs explicitly excludes and substitutes for any form of experimentation (Beisel et al. 2016). A test is either positive or negative; malaria is either present or absent. Any other onto-epistemic state of the disease cannot be indicated and is thus rendered invisible by the technology. However, as I will argue in this chapter, the explicitness in RDT testing requires and produces novel engagements of health workers as well as patients in order to determine the “real” state of the disease. As Norah indicated, this becomes particularly evident in cases of negative RDT results. Here health workers need to relate the evidence of RDTs to the patients’ own evidentiary practices as these emerge from self-diagnosis/treatment. I will argue that as a side effect of these conflicting claims of evidence regarding the state of the disease, users of RDTs start to test themselves. Health workers’ testing practices not only involve patients and their drug use patterns, but also the technology and its appropriateness for determining the presence and absence of malaria. Testing beyond laboratories and without experimentation reveals what more comprehensive experimentation with the technology in these specific settings could have found in the first place. As such, the testing inscribed in RDTs can be conceived of as a “covert social experiment” (Wynne 1988, 158) in which users of a novel technology are implicitly forced to deal with technicalities for which they are not qualified and which can only be solved by involvement of expert knowledge as well as more sophisticated technologies.

Testing beyond the laboratory

Testing is a crucial aspect of any laboratory setting in which experiments are performed. Testing instantiates all practical components and concrete

enactments of the experimental procedure through which scientists ultimately establish if something becomes measurable or not. While it is by far not the only necessary aspect for enacting experiments, testing is still one of the most viable parts of "science in action" (Latour 1987). Early laboratory studies have pointed to the relation between testing and situated practices in experimental environments (Latour and Woolgar 1979; Knorr-Cetina 1981). In *The pasteurization of France*, Latour talks of "trials of strength," describing a typical mode of how people and experimental materials in laboratories constantly and situatively relate to and test each other (Latour 1988). Messy procedures count as stabilized "facts" when they have passed several trials of strength. When these facts or technologies eventually leave the laboratory setting, it is the relation to their new environment that determines if they can continue to function as a reality test, say for detecting the presence or absence of a disease.

If we look at RDTs against this background, we can perceive of them as a materialized form of formerly situated testing practices as they emerge in laboratory settings. Due to its scientific principles, the testing inscribed in RDTs is assumed to be mobile and ready to be applied as a "stand-alone" technology outside laboratories and experimental spaces. Testing for malaria parasites can thus travel to remote regions in Uganda and become part of the everyday routines in lower-level health facilities. This mobilization, however, also means that RDTs are simultaneously cut off from other experimental, as well as testing, devices. In a laboratory, for instance, determining if and how forms of self-treatment of malaria affect the RDTs capacity to detect the disease could eventually be delegated to other tests and technologies (either to the microscope or polymerase-chain-reaction). While this is almost impossible in the context of Uganda's lower-level health facilities, health workers still have to make decisions regarding the validity and relevance of the test results.

From my interviews with health planners as well as from literature review, it became clear that many experimental trials of RDTs were predominantly concerned with the impact the devices had on the prescription behavior of health workers (Hopkins, Asiimwe, and Bell 2009; Odaga et al. 2014). In these studies, underlying concepts (e.g., "feasibility" and "acceptability") relate the technicalities of handling and performing the tests mainly to cost-effectiveness models (Asiimwe et al. 2012; Kyabayinze et al. 2012). It is not surprising, therefore, that questions regarding the validity of test results in relation to self-treatment play a very minor role. This is also mirrored in a quote from the official *RDT Users Manual*, which is, to a great extent, the product of prior feasibility and impact studies:

Prior treatment: What has been done to treat this illness before coming to your health center today? What other medications have been taken? If medications were taken, was the dose complete, or partial? This

information will help to guide your treatment decisions. For example, if a patient has taken a full course of Septrin [antibiotic, RU] but has not improved, you should not prescribe Septrin again. As another example, a patient may come to the health center after swallowing only part of a dose of Coartem [ACTs, RU]. (Note a complete course of Coartem requires 6 doses over 3 days.)

(MoH 2009)

The statement shows that while the widespread practice of self-treatment prior to attending the health center is recognized as such by Ministry experts, the manual doesn't provide any advice regarding how to proceed or how this might affect the RDT test result.³ In the following section, I show that the indicated lack of explicitness of how users can or should treat the relation between RDT results and prior treatment is part of an under-determination of technology (Feenberg 2010). More explicit attempts to define the potential relations between RDTs and prior treatment would not only exhaust the capacities of formal training protocols but, more fundamentally, would conflict with the overall scripting of the devices. The main justification to prioritize RDTs instead of, for example, extending the already existing laboratory and microscopic infrastructure, is the "simplicity" of the test (Mabey et al. 2004). RDTs are deemed a simple technology because they allow inclusion of even less qualified health personnel to engage in evidence-based diagnosis of malaria. However, I will also argue that excluding these complexities from more elaborated formalizations or experimentation is characteristic for the broader framework of implementation of novel technologies in the context of global health and development cooperation. As has been observed particularly in the field of HIV/AIDS interventions, the inversion of practice producing knowledge, rather than knowledge informing practice, is to be understood as an effect within the experimental application of technologies in emergency contexts (Rottenburg 2009; Nguyen 2010). How this can also be observed in the context of malaria control interventions in Uganda forms the wider political framework of this chapter.

Presumptive treatment of fever/malaria I: Self-diagnosis

While microscopic diagnosis of malaria is still considered the scientific gold standard, it is certainly not the norm in everyday diagnosis of the disease in Uganda. Before the introduction of RDTs, the majority of diagnoses were made based on symptoms of fever at home or in health centers. Indeed, most malaria cases are first treated at home, often with inadequate drugs bought in shops or pharmacies (McCombie 2002). Many patients only access formal medical care late in the course of disease. This might lead to treatment delays with life-threatening consequences, as most malaria deaths occur within 48 hours of onset of symptoms. In the following section, I will

illustrate why people *still* engage in self-diagnosis, and how the involved evidentiary practices need to be linked to the state and status of the Ugandan public health care system.

The fact that the disease was/is endemic in Uganda means that many people, particularly in the rural regions, experience malaria—or fever that is associated with malaria—repeatedly throughout a year (Yeka et al. 2012). Against this background, we can assume that the majority of Ugandans have been infected with malaria at least once, but most likely more often, particularly during their early childhood. Perceiving fever/malaria as part of everyday life is a result of constant individual, as well as collective, exposure in many rural communities. Having experienced several times how it actually *feels* to have malaria (e.g., high fever combined with diarrhea, headache, and nausea) affords individuals with a specific body knowledge regarding the presence or absence of the disease. The taking of, for example, antimalarial drugs on the basis of this knowledge constitutes a vital part of this diagnostic process.⁴ In cases where symptoms resolve over a few days, people assume that the cause was malaria. If the drugs have no impact—or if the condition worsens—people can at least rule out that they are suffering from malaria (Nichter and Vuckovic 1994). Self-assessment and the subsequent taking of antimalarial drugs form the sociomaterial basis for self-diagnosis. Medical anthropologists have pointed out that engaging in self-diagnosis can be perceived as a universal and worldwide practice which, nevertheless, bears some strong affordance with the therapeutic context in which it is performed (Van Der Geest 1987). Thus, on the one hand, many people I have talked to confidently negated being afraid of or perceiving malaria only as a life-threatening condition.⁵ On the other hand, self-diagnosis still reflects a strong awareness of uncertainties and that one can potentially die from malaria. The very prioritization of malaria by patients over other diseases can be read as an expression of this cautiousness, but it is also a technologically induced phenomenon. In contrast to other viral, bacterial, or parasitological diseases—which may also cause serious harm—malaria can be relatively easily treated through cheap and widely accessible pharmaceutical solutions.

In addition to the herein highlighted pragmatism, self-diagnosis is also an instantiation of what can be labeled a notorious infrastructural crisis. The frequently restricted mobility and precariousness of transportation which are characteristic for many rural regions (Foley 2010) hint to some existential difficulties of living in the context of scarcity (Redfield 2013). Long distances to the next health post in combination with severe financial constraints households face reveals the existential implications inscribed in self-diagnosis of malaria. These circumstances are further exacerbated by uncertainties regarding the drug supply and overall functionality of public health centers. From past and personal experiences, people know that they cannot trust that on reaching a health center, they will be sufficiently attended by qualified staff, or if drugs and other emergency services will

be available at all. In various ways, people cannot afford to rely comprehensively on public health infrastructures, but are instead required to take action themselves (Umlauf 2017b).

Presumptive treatment of fever/malaria II: Clinical diagnosis

Most of Uganda's lower-level health facilities are planned as outpatient units. As such, they are each expected to provide basic primary health care for an estimated population of between 5,000 and 10,000 people. In addition to malaria, therapeutic services in these units are limited to the treatment of other infectious diseases, e.g., influenza, pneumonia, urinary tract infections, or typhus. Consequently, the bulk of pharmaceuticals most commonly prescribed are antimalarials, antibiotics, and painkillers. During participant observation of consultations, I learned that the majority of patients visiting a health center are diagnosed and treated for malaria or fevers associated with malaria.⁶

Before RDTs were introduced, health workers used a combination of clinical skills and individual experience to diagnosis malaria. Following the history-taking of the patient, health workers combine a semi-standardized questionnaire (e.g., illness narrative) and physical examination (e.g., measuring of temperature if a thermometer is available) to arrive at a diagnosis. Although the detection of fever plays a central role, a significant aspect of this method requires practitioners to collect and combine various signs and symptoms (vomiting, dizziness, body pain, diarrhea). From a physiological perspective, the primary challenge is that most of the symptoms characteristic for malaria may overlap with other diseases, for example, pneumonia (Källander, Nsungwa-Sabiiti, and Peterson 2004). Subsequently, it is assumed that the difficulties of health workers to discriminate more thoroughly on a clinical basis frequently results in an over-diagnosis of malaria (Chandler et al. 2008). While clinical diagnosis is sensitive in detecting malaria, it is much less accurate in excluding the disease. It is this deficiency that is to be superseded by invasive RDTs, a device that promises to offer quick and simple parasitological diagnosis no matter the setting or qualification of the health care provider.

In the previous section, we have learned how people are engaged in therapeutic micro-experiments in which they situatively combine sickness experiences with anticipated pharmaceutical efficacies. But experiments can also go wrong or fail so that people eventually seek help in public health centers. As a consequence of this "treatment-seeking choreography," I often heard health workers complain during consultations when they realized or revealed that patients had already engaged themselves in forms of self-diagnosis. In these situations, the member of health staff is involuntarily put in the position of a correction authority of failed drug and treatment experiments. The complaint, however, is less directed

against self-diagnosis in itself, but rather to a specific use of e.g., substandard drugs or nonadherence. At this point, it is important to note that before the introduction of RDTs, health workers disclosing that somebody might have already taken antimalarial drugs was not likely to affect significantly the course of action. Most nurses would go on and prescribe antimalarials if signs and symptoms indicated malaria. The crucial question health workers are facing currently is: how does the evidentiary practice of self-diagnosis—spatially and temporally performed outside public health centers—affect the capacities of RDTs to determine reliably the presence or absence of malaria? Who and what is actually tested when these procedures contradict each other?

Rapid diagnostic tests as appropriate global health technology?

As an attempt to standardize malaria diagnosis, novel RDTs have been introduced in most endemic countries, including the east-African Republic of Uganda. As part of heavily subsidized global health programs, RDTs are now used in routine case management of malaria throughout the entire public health care service, including remote health centers with no medical laboratories. A core objective of RDTs is to identify malaria more accurately and subsequently improve the use of antimalarials. As simple and mobile devices, RDTs are expected to identify malaria *independent* of the application context as well as the biomedical expertise of the users (Moody 2002).

But what does “simple” mean exactly? Most RDTs used in the field are self-contained lateral-flow immunoassays whose scientific principles are already in widespread use, for instance, in pregnancy tests. In comparison with microscopy, which allows visualization of the whole parasite, RDTs detect antigens and thus only provide indirect proof of the presence of, for example, malaria falciparum parasites (Murray et al. 2008). As demonstrated in the pictorial user manual (Figure 8.1), the formal application of RDTs consists of 16 subsequent steps, which provide the material basis for rendering the tests a “stand-alone” or “point-of-care” technology. The core tasks consist of the pricking and drawing of a specific amount of blood from the patient’s finger (step 6) and then, using a capillary tube, the blood is subsequently put on the RDT cassette (step 9). An additional buffer solution helps the blood migrate over the immunochromatographic cellulose membrane. Following these procedures, users have to wait at least 15 minutes until they can read the test result, which will either be positive (showing two lines) or negative (showing only one line) (step 14).⁷

As part of the widespread introduction of RDTs, improving the individual case management of fever/malaria is not the only task the technology is expected to perform. On an epidemiological level, RDTs are used by national and international bodies as measuring devices for the assessment of

How To Do the Rapid Test for Malaria

Modified for training in the use of the **Generic Pf Test** for falciparum malaria



Collect:

- NEW unopened** test packet
- NEW unopened** alcohol swab
- NEW unopened** lancet
- NEW** pair of disposable gloves
- Buffer
- Timer



Disposable gloves

Timer

Buffer

Test packet

↓ READ THESE INSTRUCTIONS CAREFULLY BEFORE YOU BEGIN.

- 1. Check the expiry date on the test packet.**



Expiry date

- 2. Put on the gloves. Use new gloves for each patient.**



- 3. Open the packet and remove:**



a. Test



b. Capillary tube

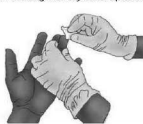


c. Desiccant sachet

- 4. Write the patient's name on the test.**



- 5. Open the alcohol swab. Grasp the 4th finger on the patient's left hand. Clean the finger with the alcohol swab. Allow the finger to dry before pricking.**



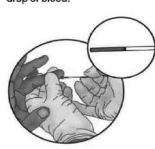
- 6. Open the lancet. Prick patient's finger to get a drop of blood.**



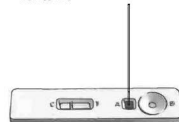
- 7. Discard the lancet in the Sharps Box immediately after pricking finger. Do not set the lancet down before discarding it.**



- 8. Use the capillary tube to collect the drop of blood.**



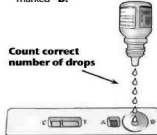
- 9. Use the capillary tube to put the drop of blood into the square hole marked "A."**



- 10. Discard the capillary tube in the Sharps Box.**



- 11. Add buffer into the round hole marked "B."**



Count correct number of drops

- 12. Wait 15 minutes after adding buffer.**



- 13. Read test results.**
(NOTE: Do not read the test sooner than 15 minutes after adding the buffer. You may get FALSE results.)

- 14. How to read the test results:**

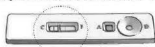
POSITIVE

A line near letter "C" and a line near letter "T" means the patient is POSITIVE for malaria.



P. falciparum

The test is **positive** even if the line near "T" is **faint**.)



P. falciparum (faint +)

NEGATIVE

A line near letter "C" and **NO LINE** near letter "T" means the patient **DOES NOT** have malaria.



Negative

INVALID RESULT

NO LINE near letter "C" and one or no line near letter "T" means the test is **INVALID**.



Repeat the test using a new RDT if no control line appears.

If no line appears near the letter "C," repeat the test using a **NEW unopened** test packet and a **NEW unopened** lancet.

- 15. Dispose of the gloves, alcohol swab, desiccant sachet and packaging in a non-sharps waste container.**



- 16. Record the test results in your CHW register. Dispose of cassette in non-sharps waste container.**



NOTE: Each test can be used ONLY ONE TIME.
Do not try to use the test more than once.



Figure 8.1 RDT training material. Source: www.finddiagnostics.org.

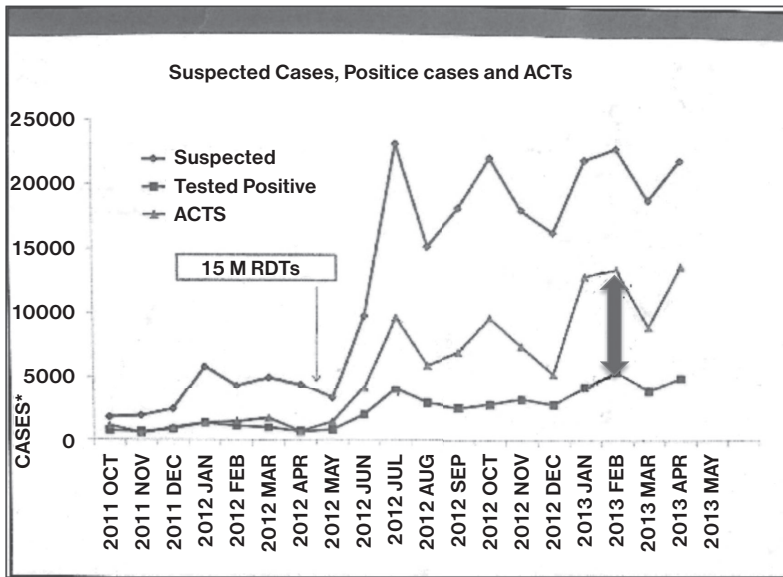


Figure 8.2 Suspected cases, positive cases, and ACTs. Source: PMI/USAID Kampala May, 2013.

prevalence rates as well as the effectiveness of other control interventions (e.g., use of mosquito nets). On this level, testing, however, is no longer limited to the patient's disease status only, but includes users and health care services more generally. As the graph indicates (Figure 8.2), the initial use of RDTs (April 2012) accounted for an increase in the amount of true positive cases (bottom line).

However, as the red arrow indicates, the aggregated data also reveal that the amount of antimalarials prescribed deviates significantly and increasingly from the assumed aim of reducing access to drugs to positive-tested cases only. In the following section, I will use the gap—one could also say the deviation from the anticipated objectives and ideal interplay between the drugs and the tests—as a starting point to explore some of the reasons why users of RDTs adhere to but also overrule negative test results and still prescribe antimalarials.

Test I: positive testing of self-diagnosis

When users of a technology engage themselves in practices of testing, their core motivation is to reduce uncertainties resulting out of conflicting reality claims—in this case, the presence or absence of malaria. To be able to understand fully the relation between self-diagnosis/presumptive treatment and RDTs, I will provide a symmetrical analysis of positive as well as negative test results. My examination starts by highlighting the tests health

workers exercise when positive RDT results are challenged by a patient's self-diagnosis. In the following section, nurse Faridah gives us a first example of how health workers may relate positive test results and self-diagnosis.

Q: Do you sometimes bet against your clinical impressions when doing a test?

FARIDAH: "Yes, many times you can think, even before you do the test, on whether it is positive or negative."

Q: Do you get surprised many times? Like, you get results that you did not expect?

FARDIAH: "Yes, it happens many times."

Q: Why do you think this happens?

FARIDAH: "This one I don't know. Just like you don't know, I also don't know why it happens!"

Q: Could it be that sometimes the patients have taken some medicine before coming here?

FARIDAH: "No, I don't think so, because many times they take the medicine but still, the test turns out to be positive. So, I think that is not the reason."

Faridah's statement points to a generalization practice directed to the relation between positive tests and forms of self-diagnosis. Generalization here is similar to an immunization. By immunization, I refer to a protection practice through which health workers preserve their trust in the accuracy and validity of a positive test result. Put differently, health workers protect themselves against the constant challenges self-diagnosis brings for the treatability of malaria based on parasitological diagnosis. For Faridah, no matter how many or which drugs a patient might have taken before visiting her facility, the tests are "immune" to this potential confusion and still manage to indicate the disease. A first specification of the RDTs capacity to properly detect malaria in the blood of an individual is indicated by nurse Alex:

Q: Do you think self-treatment is likely to affect the test results?

ALEX: "RDTs detect the malaria even after one has taken drugs. Yes, with RDTs, even if someone has taken the medicine—for as long as they have not completed it yet, it will still show you whether the malaria is there or not."

In contrast to Faridah, Alex relates the capacity of the RDTs to the quantity of medication taken by the patient. From his perspective, the tests can only detect malaria⁸ if the patient has not taken a full course of medication. Alex uses a positive test result to put the quality of prior medication use on trial. For this, he draws on experience of local medication use patterns, knowing that many patients do not adhere to pharmacological standards. During consultations, a patient might confess, for example,

that s/he has taken only a few tablets over the previous couple of days.⁹ After probing further, Alex eventually finds out that these tablets were the remainders from a previous prescription that was not finished by the patient. Similar to self-diagnosis, nonadherence can also be perceived as a universal practice, and exposes the conflicts between the perception of experts and laypersons of what it means to be cured. People's sense of *feeling* symptom-free, and subsequent refraining from adherence to a prescribed course of medication, does not necessarily coincide with the pharmacological understanding of cured as being, for example, free from pathogens.¹⁰ On a more technical level, Alex believes that an ideal interplay between the two technologies exists. However, he also trusts RDTs to account for the local deviations from pharmacological standards in that they are able to detect malaria. Whilst Alex refers mainly to the quantity of medications taken, one also finds an extension of this perspective to that of the quality of drugs taken. The following statement by Esther suggests that RDTs can be used as an indicator that (critically) tests the efficacy of various types of antimalarial drugs.

Q: Does it in any way affect the results of the test if someone has done self-diagnosis? How?

ESTHER: "[S]ometimes it depends on *which* drugs they have taken. If they took Quinine for some days, it will be negative. But even if they take a full dose of Coartem [ACTs], if the malaria is still there, it will show. But for patients that have taken Quinine, it shows negative."

Ok, Nurse Gad, you look like you are not agreeing with this view. What do you think?

GAD: "Well, I just hadn't realized that yet! [Referring to the fact that if someone took quinine, it affects the test results]."

Esther's statement is a reference to local classifications of western pharmaceuticals. As part of a broader belief system, Esther distinguishes between categories of "strong" and "weak" drugs, which more or less coincide with their pharmacological efficacy. But these local classification forms are not merely an expression of local belief systems. They are a manifestation of a complex combination of collective experiences of crisis-ridden health systems, local perceptions of side effects (as a result of specific dosage forms), as well as different forms of risk tolerance in dealing with uncertainties. Furthermore, the example Esther gives relates to the national malaria policy guidelines in which antimalarial regimes are also classified into first-line treatment (Coartem) and second-line treatment (Quinine) for malaria. In these guidelines, the intravenous administering of Quinine is recommended, especially for severe cases of malaria. Not only owing to the risky practice of intravenous drug administration, but also because of its painful side effects, achieving the acceptance and widespread usage

of Quinine, the oldest antimalarial drug, is difficult (Achan et al. 2011). Nevertheless, during the consultations I witnessed patients who repeatedly admitted to buying and taking Quinine in both oral and intravenous forms. The apparent astonishment of Gad also shows that links between the performance of RDTs and the efficacy of different medications are not part of formal guidelines or a standard procedure. Rather, it involves situated testing practices by which health workers interpret RDT results alongside local pharmaceutical efficacies.

Before moving to the next section, I would like to contrast briefly the examples provided of the health workers' use of positive RDT results to test the quality of self-diagnosis with the scientific principles inscribed in the devices. The antibody used currently in the immunoassays of many RDTs can detect traces of malaria antigens (HRP-2) for up to 21 days after someone has completed a full course of antimalarials and fully recovered.¹¹ In clinical practice, the efficacy of this antibody/antigen reaction can prove to be a limitation of RDTs, as it is associated with false positive results (Murray et al. 2008). A false-positive refers to cases in which the tests indicate malaria although malaria isn't the cause of the symptoms, which increases the risk of mismanagement of fever. From my experience, many health workers weren't aware of this technical limitation, or at least refrained from making this complex issue more explicit. On the one hand, even if health workers were more aware of this issue, very little could be done in actuality to prove that a positive result was a false positive. On the other hand, the tendency of health workers to want to "protect" positive RDT results from the confusion of self-diagnosis implies that malaria and antimalarials have important roles in this setting. For both health workers and patients, malaria is a rather desirable diagnosis because the course of action (and cure) is quite clear. By prescribing antimalarial drugs, health workers can secure their local expertise to a great extent; in turn, patients receive antimalarials, which they consider a crucial step to reducing uncertainty over the cause of their symptoms. As mentioned above, if somebody is actually suffering from malaria and takes prescribed antimalarial drugs, he or she is likely to recover very quickly and without additional complications. In the following section, I will discuss how to understand this peculiar configuration between the everydayness of the disease and its therapeutic context as a sociotechnical background against which nonadherence to negative RDT results needs to be analyzed.

Test II: Negative testing of self-diagnosis

In this section, I am interested in practices of testing through which health workers justify their nonadherence to negative RDT results. While in the former section, RDTs were used to test the quality of self-diagnosis, I will

now turn to examples in which RDTs are tested in their capacity to determine the absence of the disease. By drawing an analogy to food that can be mistakenly assumed to be ready, in the following statement, Halima highlights the possibility that self-diagnosis can affect, and even hamper, RDTs capacity to detect malaria.

Q: Under what circumstances, then, would the RDT show negative results when you actually think that the patient has malaria?

HALIMA: "When someone has taken some drugs, maybe."

Q: So, do you think that the concept of hiding parasites is actually real? Can parasites hide and the test cannot show any positive results, even when someone has malaria?

HALIMA: "Yes, it happens. Just like when you are cooking food, and you cover it well and put firewood and fire below the saucepan. But the fire in this case is actually less than what the food requires for it to get ready. So, when you see smoke getting out of the food, you are tempted to think that the food is getting ready, but it is not, because the fire isn't enough for it. So it is with self-diagnosis. Sometimes patients take drugs that are either not strong enough or not appropriate for their illnesses. Then we do the test and it is negative. That doesn't mean that the musujja [fever associated with malaria] is cured. And so, such patients should be given malaria drugs again if the symptoms suggest so."

With the analogy between food and self-diagnosis, Halima links this insight to different forms of visibility of the disease. In this scenario, whilst RDTs won't be able to detect malaria in the blood of an individual, health workers are still able to recognize persisting symptoms—either as bodily manifestations or in the patient's illness narratives. Regardless of whether these explanations are actually in line with expert accounts of antigen-antibody reactions, it still tells us about the "un-black boxing" in which most health workers are embroiled when using RDTs. The design of RDTs strongly builds on making invisible the actual scientific working principles; the very fact that the tests are mobile and simple to use means that their scientific complexity is black-boxed.

The idiom of the "black-box" or "black-boxing" is an attempt to describe functional processes of closure through which a specific use should be prioritized and secured over others. Bruno Latour defines black boxing as:

the way scientific and technical work is made invisible by its own success. When a machine runs efficiently, when a matter of fact is settled, one needs to focus only on its inputs and outputs and not on its internal complexity. Thus, paradoxically, the more science and technology succeed, the more opaque and obscure they become.

(Latour 1999, 304)

The (counter-)practice of un-black-boxing can be understood as a series of tests through which the implicit agreement between manufacturers and users is challenged exactly at this point where users are experiencing conflicting claims of reality. In most situations, this can be resolved by consulting experts or colleagues, or by calling helplines. In cases and contexts, however, where additional knowledge is not readily available, users are forced to engage in testing practices. Put differently, in an attempt to negotiate conflicting claims of evidence over the presence and absence of malaria, health workers seek to rationalize and justify deviations from test results. Part of this justification involves the formation of a new knowledge object as already indicated in the introductory statement of nurse Norah.

Norah's mentioning of "dormant parasites" constitutes a specific knowledge object in that it reveals the observation of a particular skill or behavior of parasites, one of the actors involved in the process of diagnosing malaria. This kind of animation suggests that parasites are themselves an adaptable species capable of reacting to changes and conditions in their environment. This epistemic position is quite in line with biomedical expert thinking in which parasites and other pathogens are also perceived as highly adaptable species (e.g., for the formation of drug resistance) (Bloland 2001). In accordance with her colleagues, Norah links dormant parasites to the uncertainties about which drugs people may have already taken outside the public health care system. Under-dosing, as well as the use of substandard drugs, may have the effect that not all parasites are cleared from the bloodstream.

In addition, dormant parasites can also be seen as a translation of biomedical knowledge concerning the life cycle of parasites. In this model, a crucial etiological stage is reached when malaria parasites are released from the liver into the bloodstream of an individual. This moment marks a grey area both for the development of symptoms as well as the ability to diagnose malaria. During this time, it may be that RDTs as well as microscopes will not detect antigens or parasites, although patients are already suffering from malaria-like symptoms. For health workers, this phenomenon manifests itself repeatedly as a confusing episode in the perception of the performance of the tests as the following statement indicates.

Q: When do RDTs fail to detect malaria?

EDDIE: "I just know that sometimes it fails to detect malaria, but I don't know why."

Q: Could it be that maybe someone took medicine before coming to do the test?

EDDIE: "No, I don't think so. Sometimes they fail to show even when someone has not taken any drugs, and the person really feels sick, but the test is negative. This has happened to me about two times. I send the patients to another health center to go and test in the laboratory. I then

heard they came back with positive results, yet they tested negative here. But I have also heard that parasites can hide for some time until we see them."

The confusion Eddie experienced, and which he ultimately links to the capacity of parasites to hide, implicitly refers to two characteristics of malaria. First, malaria, like any other disease, has an incubation period.¹² As already mentioned, symptoms only become noticeable and visible when parasites are released from the liver into the bloodstream, the moment from which the lethal destruction of red blood cells by the parasites begins. The subsequent struggle of the body against this destruction triggers symptoms such as fever. Nevertheless, there are also forms of coinfection in which symptoms indeed refer to malaria, but are triggered by influenza virus or other pathogens. In such cases, the subsequent weakening of the body's resistance can lead to an "awakening" of formerly dormant parasites, at which point an accelerated malaria infection takes place. But the confusion Eddie referred to also relates to some infrastructural constraints. Sending patients to other higher-level facilities exposes them to issues related to mobility and transportation. The time involved in, for example, raising money results in delays of patients in pursuing the referral advice. However, this also means that the disease might have changed and entered another state of visibility. Eddie's reference to a different technology, such as microscopy, is unjustified in this case, because the disease would have changed into a state that was detectable by RDT by this point.¹³

Secondly, it becomes clear that knowledge pertaining to "hiding" or "dormant" parasites constitutes a rather cautious articulation of the limitations of RDTs for detecting malaria. The animation of parasites prevents health workers from crude generalization, assuming, for example, that *all* negative RDT results are always the consequence of dormant parasites and therefore invalid. This also indicates that health workers apply situated testing practices and mainly in cases in which discrepancy between the clinical impression, patient illness narratives, and a negative RDT result becomes most obvious. Situated testing enables health workers to synchronize different places and temporalities of evidence production. Testing can, thus, be understood as an *investment in forms* (Thévenot 1984), for instance, the formation of dormant parasites that prevent health workers from completely losing confidence in the technology. To prevent this from happening, they engage in testing practices that involve production of knowledge objects, which are stable enough to circulate between colleagues as well as different levels of health care services.

The routine use of RDTs in primary care health facilities in Uganda makes the devices part of the everyday contradictions and improvisations needed for filling specific knowledge gaps. Lastly, two aspects can be highlighted:

On the one hand, RDTs are used to test the quality of self-diagnosis. The form in which RDTs make malaria visible is related to the less visible forms and effects of people's self-diagnosis experiments. On the other hand, health workers also use forms of self-diagnosis to test RDTs capacity to identify (or exclude) malaria. We can say that the test's inscribed objectivist ideal of making malaria diagnosable independent of any local circumstances is translated by health workers into multiple forms of drug use patterns and disease manifestation in rural Uganda.

Concluding remarks

Throughout the chapter, I have analyzed how a former laboratory procedure has been outsourced as a stand-alone technology. My aim was to show that the testing inscribed in RDTs still affords features and requirements of the testing carried out in laboratories. As an effect of the specific experimental origin, users of RDTs struggle to integrate properly the provided evidence-based diagnostic knowledge into their everyday working routines of treating malaria and fever associated with malaria. As a consequence, aligning different claims of evidence with acceptable treatment decisions requires health workers to situatively deviate from protocols and prescribe antimalarials. The underlying failure of accounting for different and conflicting claims of evidence in field studies that had been carried out prior to the widespread introduction of RDTs constitutes a characteristic feature of the broader context of global health interventions. On the one hand, more and/or long-term experimentation is increasingly restricted by the projectification and fragmentation of most of the global health interventions. On the other hand, more comprehensive experiments might also reveal that the appropriateness of the technology in question is strongly hampered by the very context in which it was assumed to fit. Current attempts to answer questions related to the high nonadherence rates to negative RDTs tend to frame the issue predominantly as a knowledge and/or technical problem. As a result of these framings and put as a learning process, new and more complex diagnostic devices are under development which are expected to screen for a variety of diseases prevalent in a specific region. Consequently, improvement of drug use patterns through RDTs also requires improvement of the ability to diagnose other diseases. The initial straightforward objective of improving the (cost-) effectiveness of anti-malaria drug use results in a transformation of the primary health care facilities in Uganda. Put differently, the introduction of RDTs leads to a questioning of the overall therapeutic capacities of these frontline facilities and whether they are at all able to deal with diseases whose main symptom is fever. The appropriateness of the technology is not questioned, but the context in which they have been introduced is transformed. Testing beyond the laboratory becomes a social experiment

exactly at this point where "learning-while-implementing" rests on a transformation of the context that makes shifts of problems look like new technical problems.

Notes

- 1 The here used ethnographic observation and interviews are part of my empirical case studies I collected during a ten-month field research project (between 2011 and 2014) in the Mukono district (Uganda) which became part of my PhD thesis (Umlauf 2017a). In that, I explored the broader political economy that has emerged around the implementation of Global Health technologies.
- 2 In this chapter, I will mainly focus on the relation between presumptive treatment and RDTs. It should be pointed out that the use of microscopy in medical laboratories is a widely practiced procedure in Uganda and is still considered the "gold standard" for diagnosing malaria. However, their reliance on expert knowledge to handle specimens and detect parasites, and also their dependence on infrastructural requirements such as electricity and water supplies, make the use of microscopy a comparatively costly and laborious procedure.
- 3 It is also striking that this is the only passage in the entire document where the problem of self-diagnosis of patients is alluded to.
- 4 Putting this point differently, one can say self-diagnosis converts drug treatment into a temporarily and spatially extended diagnostic procedure.
- 5 This kind of everyday perception strongly contrasts Eurocentric representations of malaria as "killer-disease" (see e.g., "Malaria: a major global killer" www.bbc.com/news/10520289 accessed 20/03/17).
- 6 However, the growing availability and distribution of RDTs for other infectious diseases (e.g., HIV, syphilis, hepatitis) ideally permits the extension and taking over of diagnostic services that were formerly restricted to the laboratories of higher-level facilities.
- 7 As indicated in the training manual, there is also the possibility of RDTs showing an invalid result. This might happen either as a result of user error, or due to technical failures. The latter can have multiple origins, such as production errors, logistical constraints, or climatic issues (e.g., exposure to increased heat and humidity). From my field experience and participant observation, it is more likely that invalid results occur because of user error. Particularly, the pricking and drawing of the right amount of blood requires some skill. This is further exacerbated as many patients in endemic areas suffer from anemia, making the collection of blood a difficult task (Sserunjogi, Scheutz, and Whyte 2003).
- 8 Instead of the malarial parasite, one could use malaria-antigens, as it is these that RDTs detect. However, for the sake of the argument, I will only refer to malaria and will refrain from differentiating between, e.g., antigens or parasites.
- 9 In the case of ACTs, full adherence for adults assumes a full course of 24 tablets has been taken over three days. The dosing regimen requires the patient to take four tablets in the morning and four tablets in the evening.
- 10 Assuming that nonadherence is a practice that can be found around the world—and regardless of the state of the respective public health service—in the case of rural Uganda, it still expresses a strong relation between the everydayness of a potentially lethal disease and a context shaped by scarcity. People anticipate that they themselves or somebody within the family might soon suffer from fever associated with malaria, so the saving of some tablets can be later used as a form of first aid.

- 11 The implicit question regarding the appropriateness of the antigen/antibody approach for endemic regions can be further complicated when put against the background of so called *acquired-* or *partial-immunity* (Doolan, Dobaño, and Baird 2009). In brief, in 2010, 42% of the Ugandan population was estimated to have been exposed to malaria (MoH 2010). This means that were the entire population of Uganda to be tested for malaria using RDTs, 42% would test positive, although most of these people would not actively suffer from the disease.
- 12 For the most common form of malaria tropica, the incubation period on average is 12 days (Glynn and Bradley 1995).
- 13 Nevertheless, it is clear at this point that RDTs and microscopes are two very different methods used to identify malaria. I could, however, not find out which technique can detect the disease earlier.

References

- Achan, Jane, Ambrose O. Talisuna, Annette Erhart, Adoke Yeka, James K. Tibenderana, Frederick N. Baliraine, Philip J. Rosenthal, and Umberto D'Alessandro. 2011. "Quinine, an Old Antimalarial Drug in a Modern World: Role in the Treatment of Malaria." *Malaria Journal* 10:144.
- Asiimwe, Caroline, Daniel J. Kyabayinze, Zephaniah Kyalisiima, Jane Nabakooza, Moses Bajabaite, Helen Counihan, and James K. Tibenderana. 2012. "Early Experiences on the Feasibility, Acceptability, and Use of Malaria Rapid Diagnostic Tests at Peripheral Health Centres in Uganda." *Implementation Science* 7:5.
- Beisel, Uli, René Umlauf, Eleanor Hutchinson, and Clare Chandler. 2016. "The Complexities of Simple Technologies: Re-Imagining the Role of Rapid Diagnostic Tests in Malaria Control Efforts." *Malaria Journal* 15.
- Boland, Peter B. 2001. "Drug Resistance in Malaria." Geneva: WHO. Accessed May 5, 2016. www.who.int/iris/handle/10665/66847.
- Chandler, C. I., Mwangi, R., Mbakilwa, H., Olomi, R., Whitty, C. J., & Reyburn, H. (2008). "Malaria overdiagnosis: is patient pressure the problem?" *Health policy and planning*, 23:170–178.
- Doolan, Denise, Carlota Dobaño, and Kevin Baird. 2009. "Acquired Immunity to Malaria." *Clinical Microbiology Reviews* 22:13–36.
- Feenberg, Andrew. 2010. "Marxism and the Critique of Social Rationality: From Surplus Value to the Politics of Technology." *Cambridge Journal of Economics* 34:37–49.
- Foley, Ellen E. 2010. *Your Pocket Is What Cures You the Politics of Health in Senegal*. New Brunswick, NJ: Rutgers University Press.
- Glynn, Judith, and David Bradley. 1995. "Inoculum Size, Incubation Period and Severity of Malaria. Analysis of Data from Malaria Therapy Records." *Parasitology* 110:7–19.
- Homedes, Nuria, and Antonio Ugalde. 2001. "Improving the Use of Pharmaceuticals through Patient and Community Level Interventions." *Social Science & Medicine*, 52:99–134.
- Hopkins, Heidi, Caroline Asiimwe, and David Bell. 2009. "Access to Antimalarial Therapy: Accurate Diagnosis Is Essential to Achieving Long Term Goals." *BMJ* 339:b2606.
- Källander, Karin, Jesca Nsungwa-Sabiiti, and Stefan Peterson. 2004. "Symptom Overlap for Malaria and Pneumonia-Policy Implications for Home Management Strategies." *Acta Tropica* 90:211–14.

- Knorr-Cetina, Karin. 1981. *The Manufacture of Knowledge: Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon.
- Kyabayinze, Daniel, Caroline Asiimwe, Damalie Nakanjako, Jane Nabakooza, Moses Bajabaite, Clare Strachan, James K. Tibenderana, and Jean Pierre Van Geetruyden. 2012. "Programme Level Implementation of Malaria Rapid Diagnostic Tests (RDTs) Use: Outcomes and Cost of Training Health Workers at Lower Level Health Care Facilities in Uganda." *BMC Public Health* 12:291.
- Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.
- Latour, Bruno. 1988. *The Pasteurization of France*. Cambridge, MA: Harvard University Press.
- Latour, Bruno. 1999. *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.
- Latour, Bruno, and Steve Woolgar. 1979. *Laboratory Life: The Social Construction of Scientific Facts*. Beverly Hills, CA: Sage Publications.
- Mabey, David, Rosanna W. Peeling, Andrew Ustianowski, and Mark D. Perkins. 2004. "Diagnostics for the Developing World." *Microbiology* 2:231–40.
- McCombie, Judith. 2002. "Self-Treatment for Malaria: The Evidence and Methodological Issues." *Health Policy and Planning* 17:33–44.
- Ministry of Health. 2009. "User Manual RDTs." Ministry of Health, Kampala (Uganda).
- Ministry of Health. 2010. "Uganda Malaria Indicator Survey 2009 (MIS)." Ministry of Health, USAID, WHO, Kampala (Uganda).
- Moody, Anthony. 2002. "Rapid Diagnostic Tests for Malaria Parasites." *Clinical Microbiology Reviews* 15:66–78.
- Murray, Clinton, Robert Gasser, Alan Magill, and Scott Miller. 2008. "Update on Rapid Diagnostic Testing for Malaria." *Clinical Microbiology Reviews* 21.
- Nguyen, Vinh-Kim. 2010. "Government-by-Exception: Enrolment and Experimentality in Mass HIV Treatment Programmes in Africa." *Social Theory & Health* 7:196–217.
- Nichter, Mark, and Nancy Vuckovic. 1994. "Agenda for an Anthropology of Pharmaceutical Practice." *Social Science & Medicine* 39:1509–25.
- Odaga, John, David Sinclair, Joseph A Lokong, Sarah Donegan, Heidi Hopkins, and Paul Garner. 2014. "Rapid Diagnostic Tests versus Clinical Diagnosis for Managing People with Fever in Malaria Endemic Settings." *The Cochrane Database of Systematic Reviews*, 4:1–49.
- Redfield, Peter. 2013. *Life in Crisis the Ethical Journey of Doctors without Borders*. Berkeley: University of California Press.
- Rottenburg, Richard. 2009. "Social and Public Experiments and New Figurations of Science and Politics in Postcolonial Africa." *Postcolonial Studies* 12:423–40.
- Sserunjogi, Louise, Flemming Scheutz, and Susan R. Whyte. 2003. "Postnatal Anaemia: Neglected Problems and Missed Opportunities in Uganda." *Health Policy and Planning* 18:225–31.
- Thévenot, Laurent. 1984. "Rules and Implements: Investment in Forms." *Social Science Information* 23:1–45.
- Umlauf, René. 2017a. "Mobile Labore. Zur Diagnose und Organisation von Malaria in Uganda." Bielefeld, Transcript.
- Umlauf, René. 2017b. "Precarity and Preparedness: Nonadherence as Institutional Work in Diagnosing and Treating Malaria in Uganda". *Medical Anthropology* 36:5.

- Van Der Geest, Sjaak. 1987. "Self-Care and the Informal Sale of Drugs in South Cameroon." *Social Science & Medicine* 25:293–305.
- Vermeire, Edward, Heleen Hearnshaw, Peter Van Royen, and Johann Denekens. 2001. "Patient Adherence to Treatment: Three Decades of Research. A Comprehensive Review." *Journal of Clinical Pharmacy and Therapeutics* 26:331–42.
- Wynne, Brian. 1988. "Unruly Technology: Practical Rules, Impractical Discourses and Public Understanding." *Social Studies of Science* 18:147–67.
- Yeka, Adoke, Anne Gasasira, Arthur Mpimbaza, Jane Achan, Joaniter Nankabirwa, Sam Nsobya, and Sarah Staedke 2012. "Malaria in Uganda: Challenges to Control on the Long Road to Elimination: I. Epidemiology and Current Control Efforts." *Acta Tropica* 121:184–95.

9 Experimenting with ICT technologies in youth care

Jeugdzorg in the Netherlands

Ben Kokkeler and Bertil Brandse

Introduction

The Dutch children and family social services sector, youth care,¹ is in transition. Where professionals in this sector once relied on their disciplinary training and were employed for a lifetime, they now enter an era of fundamental uncertainty: no longer are their jobs guaranteed, nor is the effectiveness of each of their care methods taken for granted. Three major developments challenge the youth care organizations. The Dutch national government recently transferred its administrative responsibilities and funding to almost 400² local municipalities who cooperate in 41 regions and severely cut budgets (Transitiecommissie Sociaal Domein 2016). The uncertainty is further increased by the fact that Dutch children and their parents are becoming more critical about social services, claim more autonomy and responsibility, and urge care professionals to supply their respective services in a coordinated way. Furthermore, technology comes into play: Information and Communication Technology (ICT), apps, and a range of social media disrupt services. The main actors in this transition are care providers who increasingly work in networked teams of about ten professionals. They focus on a specific neighborhood, or on a specific theme (e.g., sexual abuse or violence in the home or neighborhood) and related care methods that require an interdisciplinary approach (e.g., methods to pick up weak signals of abuse or health problems). As in other domains of the Dutch health and social services system, youth care professionals have to live up to high moral standards, and ethical reflection is regularly required in critical situations. Where it concerns young clients, many situations, both controversial and noncontroversial, challenge care workers. New technologies sharpen controversial situations, as ownership of data becomes an issue. Parents may threaten foster parents via social media. Young clients cling to new technologies as part of their lifestyle, or, conversely, refuse to engage professional caregivers in their “second life” of digital worlds and tools.

Youth care organizations and their new employers, the local councils or municipalities, are confronted with a new situation of complexity and uncertainty. Their first reaction has often been to follow strictly planned transitional procedures. These rationalist planned approaches aim at transferring

responsibilities from central and regional governments to local municipalities while providing continuity for clients and minimizing financial losses for councils. These transitions are combined with substantial budget cuts, resulting in fundamental uncertainties for clients, public servants at local councils, and youth care workers alike.

In many places, this complexity and uncertainty has sparked experiments. Expectations about ICT and the Internet were an important driver. These experiments coincided with a development stimulated by the then Minister of Youth and Family³ to enhance the decision-making power of patients over their health planning. At first glance, many of these experiments were aimed at testing new apps and social media, but they also challenged existing codes of conduct and stimulated health and social workers and their clients to reflect on norms and values, i.e., they amounted to what may be called moral experiments.

This chapter will delve deeper into these emerging experiments in youth care. First, we sketch the developments in youth care and how they lead to the opening up of spaces for experimentation. Second, we distinguish four different activity domains in youth care, and we ask which experiments occur, and how these experiments interact with, the activities of each domain. Third, we explore the emergent characteristics of experiments in youth care; in particular, those experiments that cross the boundaries of the four activity domains. We will argue that on some occasions, cross-domain experiments amount to moral experiments.

Developments in youth care

Complex transformations are occurring in the Dutch health and social services sector, and technology is one of the drivers of change. In particular, the youth care shows complex dynamics. Whereas in the medical sector (hospitals, general practitioners), institutional approaches are accompanied by large-scale investments in ICT infrastructure, often pushed by technological developments, in the family health and social services sector, a scattered pattern is visible. In 41 regions, hundreds of health and welfare organizations developed their own approaches. In each of these regions, due to open procurement procedures, up to 30 organizations are active. This regionalization and marketization is the result of a new policy scheme that, after years of preparation, was implemented in January 2015 (Rijksoverheid.nl 2016). While responsibilities and finances for youth care were decentralized and transferred to municipalities, they, in turn, tended to cooperate in new regional structures. This decentralization is a major operation that includes other fields in social welfare as well, and in national policy it is labeled as “the transition of the social domain” (Transitiecommissie Sociaal Domein, TSD 2016). The political expectation is that this transition will offer professionals and citizens opportunities to raise

the quality of health and social services, even while budgets are shrinking. They can, it is assumed, do so by improving the local coordination of services and by enhancing the engagement of citizens and volunteers. Interestingly, this situation offers opportunities for co-creation of more or less tailored services between care professionals and young citizens. These opportunities are at the same time limited by a diverse range of constraints, such as a variety of ethical codes for professional caregivers with different disciplinary backgrounds and limited financial space for experimentation (TSD 2016), as well as the lack of suitable technologies. On the one hand, youth care relies on commercial software platforms mainly aimed at control and governance. On the other hand, youth and professionals make use of social media. Both platforms have their own dynamics that cannot be influenced by the organizations and individuals concerned (Kokkeler et al. 2008; Kokkeler and Van 't Zand 2013). This results in a situation wherein both technologies are used simultaneously. While the commercial control software is crucial for the accountability of care services, social media platforms allow open and spontaneous communication and knowledge-sharing between young clients and between health professionals, and sometimes even between professionals and young clients. As will be illustrated in this chapter, one could speak of a clash of technology platforms.

As in other domains of social life, technology is innovating rapidly and becoming pervasive. About ten years ago, technology and ICT applications were almost nonexistent in the primary processes of youth care. The main focus of ICT suppliers and administrators was to digitize the support processes, resulting in electronic client dossiers as the dominant factor of the health care organization's information policy. This ICT development became the backbone of Dutch policy in 2008 when the Parliament urged the government and health organizations to establish a seamless exchange of patient data in the *Electronic Patient Dossier* (ECD). This policy focused primarily on medical records of hospitals and medical doctors. Youth care showed another path of ICT application and innovation. Digitalization of ECDs in social services evolved at a slow pace, as information exchange in a professional network working with a family took place only as personal contact among care workers. A market-driven reason was that the scattered landscape of youth care was not easy to access. Another feature was that the primary process in youth care required ICT applications that were by design different from those in hospitals. Whereas professionals in hospital care use systems that are protocol-driven and therefore closely connected to management information systems, in youth care, many of the processes are practice-based. Moreover, their services aim at reassuring the resilience of a social system (e.g., a family). Thus, processes and content of data exchange comprise mainly data that are only partially medical in a strict sense, and largely of a behavioral and social nature.

The technologies introduced to youth care were existing technologies in use in the private sector, such as social media. First, mobile devices swiftly evolved into smart phones, becoming common in professional use within a few years. This was planned: devices were supplied by employers to their employees (youth care workers) to enable them to use closed systems controlled by the health or social service organization. Additionally, health professionals and young clients brought their own devices, and even insisted on using their own devices and apps. This resulted in parallel technological worlds in the youth care workplace. This situation continued until social media platforms started to support communication and share functions that their users, professionals and clients, appreciated as engaging, easy to use, and, as they saw it, under their control and ownership. Iconic examples that attracted much attention were open source platforms that enabled young clients to manage, tailor, and own the content of their own health and social services portfolios. The introduction of smartphones supporting these applications caused disruptive situations, as health workers and their organizations were no longer fully in control what, when, and in particular, with whom young clients were sharing information. This resulted in situations that were not only disruptive for managers of closed and fully secured ICT systems, but also for clients and professionals. The 24/7 access to, and ownership of, a portfolio was a fundamental break with the past when access was limited to office hours for clients and workers and was only possible when authorized by the organization. The next stage in this disruptive development began when open-source content management platforms, such as Drupal, became available, enabling multi-user blogs, forums, and community websites, including the necessary functionalities of user registration and control.

The clash of these technology platforms became visible around 2010. At that time, ICT suppliers did not anticipate new requirements from youth care organizations and they, therefore, continued the promotion of their ECD solutions as the ultimate application. Building on their ECDs, both ICT developers and administrators in youth care expected portals to be the next step in giving clients and professionals access to dossiers. However, driven by the rapidly growing availability of social media platforms and beliefs about their enabling impact, health and social workers and their young clients started to experiment with other technologies. They often did so in their own time. In some cases, they were backed by team leaders and administrators who saw these approaches as support for their vision of self-organization by clients. Due to these bottom-up initiatives, a varied pattern of emerging experiments materialized in a few years' time.

Emerging experiments

The transition described above offered ample opportunities for experimentation with new forms of youth care because “protected spaces” could be created (Kokkeler 2014) wherein experiments could occur. Sometimes

activities in these protected spaces were deliberately organized as experiments, for example, as pilot studies. In other cases, de facto experimentation was recognized and organized in living labs. Furthermore, some developments may be analyzed as experiments, even if the actors themselves would not use this term.

As a consequence, in the last ten years, a wide range of experiments emerged in youth care: primarily bottom-up, local, and temporary. Many experiments started as technical innovations. The overall understanding among administrators, professionals, and clients alike was that experiments had to be well organized and purposeful and thus they were often named pilots or tests. Although a lot of pilots took place, successful adoption of new practices was scarce. Re-use and transfer of best practices was limited, as can be seen in the somewhat technical and mechanistic approach, pinpointed by words as “roll out” and “implementation.” The health care organizations’ reaction was to adopt only new methods that were evidence-based. In addition, new approaches enabled by technology and developed in practice and in close co-operation with young clients were welcomed, but often to lack follow-up.

Experimentation in four activity domains

To come to grips with experimentation in youth care, we distinguish between four activity domains. As we will show, the room for experimentation and the organization of experiments are different in each of the domains. We describe our underlying analytical model and the experimentation in each of the four domains by illustrating them with brief case studies based on multi-annual empirical studies (see also Kokkeler et al. 2011–2015). These case studies give insight into the challenges that new ways of working and organizing pose for health and social services professionals, improvisation, reflection, and learning in practice.

The four activity domains that we distinguish in youth care are:⁴

- **Administration and quality management.** Here, we consider activities conducted by professionals with management functions in their organizations. Dominant values in this domain are accountability and quality control. The focus is on centralization.
- **Chain cooperation.** In this domain, we encounter professionals from different disciplines working together on one dossier, around one client or his/her social system, but not necessarily working in a team. Dominant values are results, efficiency, reliability, and predictability.
- **Inter-organizational teams of professionals.** In this domain, joint activities are initiated by professionals in networked inter-organizational teams. Dominant values are effectiveness (rather than efficiency) and serious consideration of the wishes of the client.
- **Open networks.** In this domain, a variety of activities ranging from professionals working in open networks (often temporary and thematic) to

explorations of new ways of working and co-created projects wherein clients, as (envisaged) owners of their health and social services, have a prominent role. Dominant values are the autonomy of the client and openness.

The model allows us to describe and understand different dynamics of each field of activity. In our empirical cases, we recognize differences that we regard as “contrasting dynamics”.

Experimentation in administration and quality management

Due to the municipal organization of the decentralization of youth care, service organizations that had before relied on five-year contracts with provinces and insurance companies were now contracted for a period of one to three years. At the same time, budgets and tariffs were under severe pressure, often leading organizations to focus solely on productivity. As a result, there was limited room and money for experimentation. Managers felt and continue to feel the need to focus on optimal use of scarce resources, such as human capital and money, in the short to medium term.

Nevertheless, openings for experimentation occur. The case study in the box below illustrates how a youth care organization reflects the discussions among administrators, managers, and team leaders in youth care organizations that took place in many regions.

Case: Regional youth care organizations squeezed between two technology regimes

In a youth care organization, discussions come to a climax when the care manager proposes to invest in ICT-enabled innovations and the business manager considers the evaluation of the next-generation of ECD software. The organization embarks on an exploration. The first step is an in-depth analysis of current information and IT use.

The investigation by an external consultant reveals many detailed observations that were discussed and validated during workshops with professionals of all disciplines, managers, and with foster parents. These workshops were highly appreciated by the participants, as they facilitated open dialogue about motives, moral considerations and uncertainties, and possible future directions. Basically, the research showed that professionals saw the existing IT system as a black box in which data had to be entered while nothing useful came out. Furthermore, professionals were not using the central ECD for cooperation or knowledge sharing. They did not do so within the organization: academically-trained assessors who diagnosed patients and imposed treatment plans did not share dossiers with the social service workers who coached clients. Clients had no access and the system was unknown to them. There was also no

digital data exchange in chain cooperation with professionals from other organizations.

The business manager appreciated the clarity of this analysis, as she was convinced that data-supported health management and service needed improvement. The care manager saw this problem and the lack of cooperation as an opportunity for radical change. Instead of reorganizing internal procedures with a new ECD protocol, she decided to shift as much ECD content as was legally feasible into digital portfolios that would be managed by the clients.

Workshops were organized with professionals, envisaged chain cooperation partners, and foster parents to establish a digital platform for the client-owned portfolio. This process generated much enthusiasm among clients who were invited to participate in small-scale pilot projects. Clients could improve their self-esteem in a fundamental way as they became involved in their health recovery planning. Professionals were very pleased with new possibilities to share information and cooperate with clients, immediate colleagues, and chain cooperation partners (in this case, municipalities and schools). Meanwhile, the business manager talked with the ECD provider to engage it in this new development and in preparatory pilots, but the provider refused to cooperate. As a result, pilots were limited to experimenting outside the IT infrastructure and thus the full impact of client ownership could not become visible. After about a year, while an uptake to an integral implementation approach and a range of chain cooperation pilots were being prepared, a turning point came. The care manager driving the innovative operation left for a position in another region. The new manager reconsidered the development as the ECD provider signaled that its product would support caregivers in chain cooperation, and the new management thus stopped the experiment.

In this case, care managers had high expectations for the development of client-managed portfolios and a belief that such new technologies would encourage self-management of social service plans by clients. Business managers were open to these new developments as well, but struggled with the organizational transition, control of the risks involved, and the discontinuity that might occur if the cornerstone of their accountability, their ECDs, would be reconstructed.

Even though administrators and managers themselves are often professionals who understand the moral drivers of professionals working in networked and loosely coupled teams, their main focus is on exclusion of risks and experimentation due to governance codes that focus on financial control. Specialization is one of the resulting strategic choices, resulting in task divisions with other organizations. When it comes to investment decisions regarding ICT, a preference for control and for closed systems is the result. This approach

limits space, also in terms of ICT platforms and tools, for experimenting in conjunction with clients or with professionals from other organizations. As a consequence, planned experiments have to be successful, or, if this cannot be ex ante guaranteed, conducted as a pilot outside of the organization (e.g., on an external ICT platform).

Experimentation in chain cooperation

As the Dutch government pursues the decentralization of responsibility for public health services, youth care, and elderly care, health and social welfare organizations are increasingly pushed to organize services in chains. To control quality and costs, clusters of activities are organized according to strict task divisions, as packages in a chain. As clear as chain cooperation might look in theory, in practice, it is a challenge for professionals, in particular, for those who take up project management tasks to get chains in place and to secure their quality.

The case study in the box below illustrates this challenge. It describes a two-year process in which organizations embarking on chain cooperation engage various members of their teams who must adjust to this new practice and manage the uncertainties and dependencies of chain cooperation. The envisaged chain partners were not only other health or social service organizations but also professionals who worked for municipalities in their region.

Case: Decentralization and ambiguous expectations causing uncertainty within a youth care organization

In a two-year process, a youth care organization entered the phase in which the transition in the social domain started to materialize. The management of the organization acknowledged that new challenges were ahead. The organization had to make major efforts to engage in marketing, relationship management with local councils, and a range of other new tasks. After some negotiation, they were given the opportunity by the regional government to submit a transition plan. The administrators seized this opportunity and quickly developed a plan that included broad introduction of digital portfolios that would also serve communication and learning with authorized municipal public servants. In its efforts to implement a client-owned portfolio, the youth care organization had already invested in technology development, implementation, and training for years. The portfolio approach had become part of the care process; trust and ownership of private data and sharing of it with care workers as a very precious process had become an acknowledged and respected issue of skill development among care workers.

The plan was granted, but during its deployment, ICT-enabled chain cooperation was removed from the agenda. Business management and care managers responsible for marketing felt that being in control of care results was the most important condition to meet requirements from the local municipalities that would not appreciate experimentation. It was decided to transform the digital client-owned portfolio into an instrument for ehealth, owned by the organization to maintain control over and to secure quality.

When that the management decided to cancel the client-owned dossier, young clients quickly retreated. They were very sensitive to issues of ownership and trust, and state-of-art facilities such as watching educational videos via the ICT infrastructure were not allowed. Clients got annoyed and, in turn, health professionals voiced complaints about the lacking quality of digital media and portfolios. Ambitions were lowered, which in turn caused uncertainties among social service workers and clients who were in favor of self-managed portfolios.

In order to secure proper positioning in the transition towards local municipalities, management decided to focus fully on non-digital services, pilots, and training sessions with representatives and professionals from these local governments. The digital portfolio program was put on hold, and the youth care organization decided to wait for initiatives from the local councils, partner organizations, and ECD suppliers.

As the case study illustrates, expectations and ambitions to achieve innovative results are high. Nevertheless, chain social services managers tend to organize activities as closed operations. Their prime mission is to keep the operation on track. This happens in the starting phase but is prolonged if the context remains ambiguous. Moreover, reflection and learning in action by chain managers is mainly focused on their own part of the chain. In this activity domain, professionals are expected to be result-oriented and efficient, while room for improvisation and experimentation is limited as they must adhere to organizational procedures and protocols. The overarching values of social service workers and their managers are predictability and reliability in their chain cooperation partners.

Nevertheless, they still seek ways to organize reflection on action beyond the service activity packages in the chain. Aside from their own professional motives to promote overall effectiveness, there are other drivers as well. Professionals intentionally organize, formally or informally, their own learning communities to understand new developments. Chain and project managers use these gatherings as an opportunity for, what in higher education is labeled as, “intelligent accountability” (Fullan 2005). First, they manage result-oriented operations and feed performance data into reflecting-in-action

learning sessions. These interventions are often of a formal nature as part of a plan-do-check-act cycle, resulting in top-down pushed understanding that care services should be improved by adopting new technology. Second, they stimulate practice-based informal learning while showing leadership towards professionals, leading and inspiring them in thinking and acting beyond their own prime tasks, and feed data into reflection-on-action sessions. These are data that go beyond immediate performance of their part of the chain and inform and inspire professionals to understand chain spanning patterns, reflect on them and, where possible, anticipate possible improvements for chain co-operation by taking innovation steps that include several health activity packages in a chain. In this way, new openings for experimentation are created.

Experimentation in inter-organizational teams of professionals

In the Netherlands, many health and social service professionals currently work in loosely coupled teams. Professionals are sent to join multidisciplinary teams that cover thematic actions or specific neighborhoods. There are different drivers for the creation of these inter-organizational teams that have implications for the kind of experimentation that occurs. The national policy of decentralization and the transition of youth care funding in the Netherlands is one driver. Local councils initiated the formation of thematic or neighborhood teams. In many cases, the explicit aim of these teams was to guarantee better health and social services at lower costs. This situation created ample space for practice-based experimentation. Another driver was the response of caregiving organizations that, in many regions, formalized informal cooperative networks. This is illustrated in the case that we present below.

Case: Living laboratories in a regional consortium of 23 youth care organizations that heads for a joint client-centered platform

This living lab lasted about three years. In a preparatory stage, consultative meetings took place for board members within existing teams and within newly formed thematic groups between the organizations. These dedicated sessions turned into arenas of ethical debate in their own right. What started as ICT and information-centered informational sessions led to sharp controversies. Fierce debates took place in an open atmosphere, representing diverse opinions on the possible role of new technologies in health service organization and innovation.

One of the outcomes was that about ten organizations took the lead and committed themselves to invest jointly in the development of a

common underlying ICT-platform that eventually would support a range of specific client-centered applications and interoperable data exchange between the organizations concerned. The overall idea and ambition at the time was that this diverse vanguard and complementary group of organizations would experiment together and learn systematically. Over time, they would engage the other members of the consortium to create a standardized platform with a range of client-oriented services and a sound financial base, thus becoming integrated in the going concern of contracted care services.

Two living labs were organized. The first experiment focused on a new inter-organizational team that had the responsibility to implement a new multidisciplinary youth care service approach. This pilot was expected to transfer the lessons learned to other networked teams in their start-up phase. Expectations were high that this team, comprising professionals from five organizations, would be the right co-creative setting for a structured experimentation with new ICT tools. An existing inter-organizational team with a proven chain approach took up the second experiment. This team was in a process of collaborating with professionals from three other organizations. One specific challenge was the maintenance of privacy for exchange of data, as it concerned organizations working with clients under judicial supervision.

The second experiment began swiftly, driven by the enthusiasm and ambitions of the initial team. Nevertheless, enlisting other professionals and their young clients in their experiment took more time than expected. These young clients were very hesitant to join because they felt they were not the main actors in the experiment, but rather were invited to participate in an experiment that only served the interest of the cooperating organizations.

The first team struggled as well. In addition to start-up problems, it encountered difficulties in convincing clients to participate. In this living lab, the care workers postponed inviting their clients to participate until after the new ICT environment was functional. Clients felt they were being used as guinea pigs. Moreover, the living lab was meant to be integrated in the social workers' daily workloads, yet only limited time had been allotted for this work by their managers. By the time clients were invited to participate in the experiment, little time was available to coach them.

The experiments were evaluated and stopped without a follow-up in a period of turbulent reorganization. The overall conclusion was that co-creation was simply too much effort for the caregivers involved. Coproduction of existing modules was sometimes feasible, co-design only applicable where experts from outside were leading the process.

The professionals in this activity domain are driven by an ethically motivated desire to be effective. As a consequence, they favor the wishes of their clients, no matter the cost. Nevertheless, these professionals are often risk-averse and therefore prefer to act and learn in the comfort zones of closed systems that create mandated spaces for experimentation supported by secure and functioning information technology. They do so for a number of reasons: one, they do not want to put their clients at risk; two, they lacked experience on how to organize experimentation securely; and three, experience with experimentation did not serve their career perspectives. They were trained in a specific discipline, and the continuing development of their disciplinary knowledge is more important for their certification than experimentation. These limitations are fairly structural.

Despite these structural limitations, opportunities for organizational learning emerge. These teams work in challenging contexts that require multidisciplinary approaches. Loosely coupled network teams work in clusters and broader learning communities. As a consequence, trading zones might occur wherein common reflection on competing values evolves (Kokkeler 2014). The basis for these learning spaces are shared multidisciplinary approaches and protocols, more or less integral approaches towards thematic and neighborhood programs.

A next level of learning will occur once new actors enter this arena of competing ethical values. Patients and former patients increasingly claim an autonomous role, as is illustrated by the following case about a living lab organized with mentally disabled young adults.

Case: Clients contest health workers while claiming their own digital space

While organizing a living lab in an organization for mentally disabled people, the initial selection of participants and testing environments caused some discussion and showed that different perspectives on innovation with new technologies were at stake. Client safety and quality control of care was a central theme of the organization. Therefore, the initial experiment design was that clients would only work and test in completely secured and supervised environments. The environments included secured desktops in supervised rooms, the use of only organization-owned mobile devices and, according to house rules, they could only be used during daytime hours. This caused fierce reactions from young clients who insisted on using their own mobile devices. They also suggested that if the organization had budget available, it would be better to use it to engage more clients in the experiments. The clients' self-esteem grew in the course of the pilots. They showed remarkable abilities when it came to working with trusted environments

and, for instance, distinguishing what to post on open social media and what private stories to preserve for their secured digital portfolios in the pilots.

The innovation manager and care managers were pleased with this result. They invited clients to assist them in presentations, to tell and share their own experiences elsewhere in the region and in national presentations. Interestingly, parents reported the positive impact. They were very content that their children created their own digital environments, stories, and other personal expressions.

Some professionals met this growing client self-esteem and self-management of devices and portfolios with doubt. Of course they were positive about the stimulating effect of digital means for their clients, but the way that clients contested the experimental design and claimed their own digital space caused uncertainty and ethical discussions among professionals.

Experimentation in open networks

The fourth and final experimentation domain focuses on activities that are initiated or driven by clients.

Young people often act as de facto agents of change. In particular, when it concerns ICT and social media, they learn quickly from each other and often take their (foster) parents and grandparents along. This creates opportunities for experimentation and learning that are often overlooked in youth care organizations that focus on client security and control of new developments. This intrinsic tension becomes productive in well-organized living labs and related experimental learning environments where youth care professionals develop new learning-by-doing styles. While in conventional professional work they apply only evidence-based methods, now their learning space is widened as they work together with young clients in experimental spaces, partially adopting the activity-based learning style of young clients.

The impact of non-users on effective innovation of user-initiated experimentation with new technologies is relevant for this domain (Hyysalo, Jensen, and Oudshoorn 2016). Although pilots and living labs are often facilitated by youth care workers, the effective execution depends on the clients who have the crucial roles to manage, update, and share content in their portfolio. In particular, active sharing and giving feedback with digital means was an aspect of the living labs. This is a valuable and sensitive activity, as we will see below. The case summarized in the box below illustrates the impact of content ownership on effective use of new applications in related health and social services. It concerns a youth care organization that embarked on a programmatic approach of about three years in which living labs were organized.

Case: Clients threaten to withdraw from a pilot if their requests for trust relations are not met

The management of a youth care organization embarked on an innovation program that gave the clients control of their own digital portfolio. Driven by the ambition to initiate radical innovations, it organized a living lab in which professionals had no access to client portfolios unless invited by the young clients. Clients received the freedom to express their own interests and feelings by adding to their portfolio texts, pictures, sound recordings, poems, and other materials. A range of pilots were planned for the first phase of about six months, in which clients would add content to their portfolio without any engagement or interference from their professional coaches.

It turned out that participating clients, young adolescents between 14 and 18 years of age, were stressing the necessity to work in small groups in response to peer pressure and negative group dynamics. This preference was accepted by the health organization, although it raised concerns that such innovations would be too time-consuming and expensive.

Another unexpected and controversial outcome was that clients insisted on special trust relations. They were very reluctant to share the management and ownership of their stories and other content with their coaches and threatened to withdraw from the pilot unless an external gatekeeper coached them in their daily decisions about what to share with other clients and professionals. Clients wanted full control of the content they published; they were reluctant to invite their regular coaches because they worried that the experiment would interfere damage their relations with their coaches. This led to deliberations in several meetings of caregivers and managers, as it touched upon fundamental elements of the ethical code of conduct. An important consideration for health and social service workers and managers was that the new digital expression enhanced the participation of young clients in this living lab and intensified and enriched communication between clients and professional coaches.

We see from this case that trust is important in living lab experimentation. In the stages that followed this half year, youth care workers claimed to have their own pilot environments as they were triggered by and keen to keep pace with the experiences of their clients. In doing so, they sustained trust. Nevertheless, the living lab approach caused, unexpectedly, a new challenge to trust relations in the regular services programs. A valuable tipping point turned out to be the clients' willingness to engage in systems that they could not control.

Nevertheless, clients expressing reluctance gave credence to youth care workers who believed that the experiment would fail. Conversely, clients

embracing a new tool and claiming sole ownership of their portfolios and excluding social service workers made the youth care workers feel a loss of trust necessary to help their clients. Professional motivation to help young clients can also be challenged if, enabled by her digital portfolio, a client unintentionally breaks with the protocol as the following case summary illustrates.

Case: Personal storytelling challenges, existing protocols, and technological infrastructures in a living lab

The case concerns the unintended violation by a young client in a shielded youth care environment, a foster home. One evening, a client was working in her digital portfolio on a desktop PC in the living room where, at the time, she was alone and felt free to write. She wrote a story about an event in her life that she had neither shared with her coach nor reported during the initial phases of youth care evaluation. It concerned a traumatic event that, in her writing, related to her relationship with her current foster parent. The client left the desktop open when going to the toilet, at which point one of her foster parents entered the living room and read her text. The foster parent, following the organization's protocol, was required to turn off the PC. The management was informed immediately and within hours the client was placed in another foster home to secure her safety and to enable investigations into the life event she had described and the possible relationship with her current foster home inhabitants. Foster parents and health workers were shocked because they feared her story had been shared with other clients via her portfolio. A truly ethical dilemma resulted. For a time, the organization considered denying this client and other participating clients in the experiment access to computers and the ability to share their digital stories and portfolios. Nevertheless, social service managers who supported the pilot were very pleased with the result: it had shown that the ownership of a safe digital portfolio could enable a young client to tell a life event in her own words, thereby leading to improved future care. This event caused turmoil and fueled other debates among youth care about clients' free access to available technology and portfolios. Young clients were fond of writing short stories, even poems; preferably shaped, colored, and illustrated by stills and short videos. Their requests for the freedom to write in a trusted portfolio that only they could access put pressure on the protocol that required care workers to have access to these portfolios at any time and share them with other care workers. Moreover, the available IT system caused unforeseen limitations to the living lab, limiting the young clients in expressing their feelings via short videos. This experiment led to the adjustment of the protocol and, though with much delay, the IT system was sufficiently upgraded to accommodate the clients' needs to express themselves.

The activities in this experimentation domain comprise a range of perspectives because the client participates and is responsible in part for the experiment. A proper set-up requires socially specific contextual analysis of the clients; the roles played by the clients in co-designing, producing, or even creating new technologically enhanced approaches may depend on the kind or phase of experimentation. To avoid risks for clients and to avoid setbacks, potential risks (for instance, privacy leaks) must be anticipated and proactive communication encouraged. In one case, an experiment was stopped when clients lost trust in pilots. The technology used in a living lab also is associated with risks. A diversity of apps become available that does not connect or interlink, resulting in a rich but not necessarily usable abundance of data that may form a risk, not only for clients' privacy, but also for the methodological robustness of a living lab.

Despite these risks, there are new possibilities and advantages as well. Broader debates in society about privacy enable citizen initiatives that motivate youth care organizations to reflect on ownership of personal data. Educational institutions and public awareness organizations stimulate young people to become streetwise on the digital highway.

Summary of empirical observations

We see different moral dilemmas emerging in the four experimentation activity domains: (1) administration and quality management, (2) chain cooperation, (3) inter-organizational teams of professionals, and (4) open networks. In the business and quality management domain, steered by administrators and managers, we see focused experiments. These experiments are aimed at successful implementation of new technologies, keeping risk low and time efficiency high. Moral dilemmas that occur concern the use of digital data. Youth care workers do not apply digital data in their primary social services procedure; figures and other data are perceived of as part of the secondary process, enforced by the management: reporting and accounting. Feelings of alienation from their clients are voiced by health workers, once they are required by the management of a care organization to participate in experiments.

In the chain cooperation domain, we see more problematic situations occur than in the business and quality management domain. Experiments are planned by managers of the organizations in a youth care service chain, and are organized as closed spaces to minimize risks of data leakage. Often youth care workers are not authorized to enter into the back-office systems, the ECDs. Moral issues become quite apparent, as a diversity of—often tacit—organizational cultures collide.

The domain of informal professional learning communities shows contrasting developments. On the one hand, care organizations implement formal e-learning environments that are often closed and not connected

to client data for privacy reasons. In that case, technologies are fully entangled with internal procedures and only limited, or no, space exists for experimentation. On the other hand, youth care professionals initiate informal group learning, often supported by social media such as Yammer, in which they share information about daily practice and their reflections. This offers many informal opportunities to share insights about small-scale, emergent experiments. Moral dilemmas occur when youth care professionals want to generalize lessons learned and must anonymize cases, thereby losing details of their examples.

The domain of client self-management and client portfolio development shows a completely different picture. Experiments, often short-lived, are scattered, develop bottom-up, attract much attention, and are influenced by the availability of new social media technologies including apps and other social media services. In this domain, moral issues are numerous and manifold but not embraced or pursued by youth care workers or their organizations.

Cross-domain and moral experimentation

In addition to specific characteristics in the four activity domains, we increasingly see developments that transcend the domains. Two examples include professionals who search for new forms of inter-organizational cooperation, and young clients who are extremely critical towards changes in health and social services while they simultaneously are early adopters of new technologies in their private spheres. As a result, new temporary constellations emerge, crossing boundaries between different organizations and domains. These constellations can be described as “trading zones” (Kellog, Orlikowski, and Yates 2006) or “third spaces” (Nooteboom and Stam 2008; Whitchurch 2008; Kokkeler 2014).

Governance of cross-domain experiments with new technologies is unknown terrain in youth care organizations. Administrators and managers have limited experience with experimentation and no practice-based policies that would support such experiments. Due to decentralization and related budget cuts, the near future promises to be turbulent and, thus, difficult for youth care organizations to find proper partners for innovation endeavors. Their boards of governors avoid risks, their contractual customers (the municipalities) lack interest in innovations, and potential partner youth care organizations are committed to immediate (short-term) improvements.

Cross-domain experiments are therefore still scarce, and when they occur, are emergent and unplanned by an organization’s management. While experiments in each of the domains lead to moral dilemmas, the expectation is that experiments bridging two or more domains will amplify moral issues. This is to be expected because each activity domain is characterized by different dominant values. It may be expected that when experiments cross the

boundaries of the existing domains, values will collide. The experiments may be seen as moral experiments in which not only technologies or procedures are put to the test, but also conflicting values important to youth care. We illustrate this with two examples of such experiments.

In the first example, the cultural collision occurred between the domains of administration and quality management and of chain cooperation. Once care workers engage in chain cooperation, they have to overcome their discomfort with sharing data about their clients with colleagues from other organizations. They succeed in doing so, where trust is created and data in the primary process are shared that are of direct importance for improved joint care interventions. Once these collisions about data sharing within the chain cooperation get extended towards activities in the administration domain, for instance, sharing performance data, youth care workers voice reluctance, as they feel alienated from their primary tasks. Instead of enriched interactions with their colleagues and clients and support by ICT applications, they feel distracted by new cross-organizational administrative responsibilities, and they feel discomfort as they don't know what other organizations are doing with the performance data that they deliver. Both effects tend to narrow the space for experimentation.

The second example concerns the emergent spaces that evolve when experiments bridge the administration and quality management domain and the open network domain. Here, the values clash becomes most apparent in technologies used: the ECD and the digital portfolio. ECDs are owned by the youth care organization. For the administration, they are the pillars of accountability for the quality of services, management, and finances. With few exceptions, clients do not have access to their own ECD. In contrast, the client-owned portfolios, in which content is created by the clients, are based on the value that content is owned by a citizen (i.e., the client⁵) who controls whether or not a youth care worker may see it. However, moral aspects of these coexisting file types and the values they represent are rarely reflected upon, and little structured dialogue about moral aspects takes place.

Youth care professionals and their managers, as presented in the cases, do not yet see experiments as moral experiments. Their clients, however, may not use the vocabulary of ethics but do recognize the moral aspects in these experiments, such as ownership, access rights, trust, and the difference between trust relations in care relations versus those in an experimental situation. In this sense, clients and professionals who engage in living lab experiments with new technologies act as agents of change. They articulate underlying tacit assumptions and principles and make the contrasting values of different perspectives on innovations with technology visible. The cases show that professionals, while being challenged by competing values, often articulated by rethinking and repositioning in action and, while being engaged in structured reflection, *on action* (Schön 1983).

Experimental spaces that are more or less orchestrated by managers of youth care organizations, or of chain cooperation between organizations, limit moral experiments. In particular, in chain cooperation, technologically-oriented experimentation exposes moral dilemmas, but youth care professionals have little opportunity to experiment or reflect on ethical dilemmas. Despite this limited space for experimentation, it is important to note that technological innovations are often imposed top-down. Intended or not, ethical issues arise in technological innovation projects in the youth care and are articulated by professionals and concerned clients alike.

Conclusion

Experimentation with new technologies in youth care shows an emergent pattern. Bottom-up initiatives taken by professionals and clients are manifold, resulting in spontaneous unstructured initiatives. Spaces for experimentation are narrow and closely entangled with technology-driven experiments limited to a specific activity domain. Professionals in the youth care sector are motivated to learn and improve their client care. Nevertheless, innovation and experimentation with new technologies is still a new notion and raises little enthusiasm. The role of new technologies such as mobile devices and related ICT platforms raises concerns among youth care professionals, causing moral dilemmas about shifting trust relationships with their clients and privacy issues of sharing data with colleagues from other organizations in chain cooperation.

The practice of experimentation with new technologies in youth care shows two patterns: one, where administrative activity domain is dominant, managers take the lead in organizing experimentation as no risk technological innovations. These initiatives are often technology driven. Two, bottom-up informal initiatives are embraced by young clients and youth care professionals because they are seen as improving self-esteem and aiding clients to regain responsibility. These two patterns suggest the existence of two separate worlds. Innovative technologies—new ICT applications and social media—are not moderating between contrasting experimentation activity domains nor stimulating transformative developments (Kiran, Oudshoorn, and Verbeek, 2015). They do, though, expose moral dilemmas that likely will contribute to more fundamental transformations of youth care as we know it.

Two dominant vectors become visible in the scattered pattern of bottom-up initiatives and more directed top-down innovation projects; in particular, when professionals have to bridge different values of the four activity domains. First, the dominant vector of innovation can be observed in experiments in chain cooperation. Professionals are challenged to cooperate with both new chain colleagues and clients beyond the borders of their organizations where they are confronted with physical, procedural, and/or

technological barriers to data exchange. Where some professionals strive for maximum data sharing with clients and experimentation, others are reluctant or simply forbidden to do so. An extra dimension of moral learning emerges where activities and values from the administration activity domain coincide with chain cooperation.

The second vector is efficiency. The main ethical driver for professionals in chain cooperation is increased effectiveness in client-centered care. The administrative domain, in the context of decentralization and budget cuts, expects chain cooperation to result in increased efficiency. This focus on efficiency influences the dialogue among professionals in the chain. Some will plea for innovative radical steps to give clients more autonomy and invite them as partners in experimentation leading to long-term changes if successful. Others will focus on the short term, encouraging improved efficiency in their own organization only.

In addition to these two vectors, a new vector is now emerging that places more emphasis on the autonomy of clients. In the coming years, given pressure by both insurance companies who value knowledgeable consumers and by citizen groups who stress self-management, the dominant vector might move towards bridging the open network domain and the administration domain. Compared to the precious vector of moral dilemmas in bridging chain cooperation with administration values, the dilemmas and openings for moral learning in this new vector might be more fundamental. Values of open, client-owned and -steered experimentation in partnership with care professionals may clash with the closed organization and non-risk oriented approaches supported by actors in the administration domain.

Learning from good practices in other sectors, new modes of governance for experimenting within and between youth care organizations can be developed. An important aspect of this renewed governance approach would be quality management for re-use of best practices; at present, practice-based experiments of new technologies with clients do not fit in knowledge reproduction mechanisms of evidence-based protocols (Kokkeler et al. 2012). As is the case in other sectors during the general decentralization of national government services, many argue that services will improve when control is returned to local municipalities, citizens, and professionals. Unfortunately, this underestimates the time, space, and expertise for proper experiments with new technologies and subsequent reflection necessary to implement the decentralization. An opportunity for co-creative experiments is at hand, in particular in the youth care sector, once basic requirements for structured living labs (Kokkeler 2017) and action-oriented group learning with new ICT can be fulfilled (Bondarouk 2006).

Triggered by the promise of new technologies, expectations about innovation in health care are high (Ministerie VWS 2016). The digitization of primary and secondary processes in youth care poses challenges to managers and administrators who are not accustomed to organizing experiments. Clients and their insurers demand new services and business structures

(Kokkeler et al. 2008). Care service providers are challenged to improve effectiveness or they experience the decline of job security as they are drawn into open, networked organizations that emerge. Complexity and changing demands require multi-perspective and flexible approaches to experimentation (Pettigrew 1990; Rip and Schot 2002; Kokkeler 2014) to contribute to practice-based organizational learning (Schön 1983) and bottom-up sense-making in ambiguous contexts (Weick and Quinn 2004). Conceptual models to understand the current situation and to anticipate accordingly seem to be lacking. Traditional organizational concepts are still dominant, building on accountability, professional codes of conduct and protocols. Training in the co-design and co-introduction of new technologies is still a minor, if not absent, part of the curriculum in educational programs for youth care professionals. There are clear opportunities for responsible experimentation in youth care.

Notes

- 1 Youth care (*Jeugdzorg*) in the Dutch system covers a wide range of services including, depending on the definitions applied, all services, welfare, and health, for families, children, and youth aging from 0 to 18 years. In this chapter, the focus is on services for youths from 8 to 18 years of age. Organizations that were engaged in the empirical studies ranged from care organizations that exclusively focus on youth up to care organizations for disabled youth and for alcohol and drug addiction treatment.
- 2 To be precise: 393 local councils on January 1, 2015 (www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/overig/gemeentelijke-indelingen-per-jaar/indeling%20per%20jaar/gemeentelijke-indeling-op-1-januari-2015).
- 3 At the time, the Dutch government had a minister with an integral responsibility for youth and family: Minister Rouvoet, Minister voor Jeugd en Gezin.
- 4 The authors made an attempt to adopt best practices and theoretical notions from a wider spectrum of innovation studies, STS, and change management literature: distributed leadership in innovation (Kokkeler 2014); competing values in organizations (Cameron et al. 2006); organizational learning (Schön 1983); sense making (Weick and Quinn 2004); ambidexterity (Birkinshaw and Gibson 2004); structuration (Giddens 1984); and multi-level processual innovation (Pettigrew 1990).
- 5 Much discussion has focused on this labelling. We see it as follows: everyone is a citizen who owns his own data; when a citizen is involved in a (mental) health care program, he or she becomes a client and still stays a citizen. It is a role with a limited duration.

References

- Birkinshaw, J., and C. Gibson. 2004. "Building Ambidexterity into an Organization." *MIT Sloan Management Review* 45(4):47–55.
- Bondarouk, T. V. 2006. "Action-oriented Group Learning in the Implementation of Information Technologies: Results from Three Case Studies." *European Journal of Information Systems* 15:42–53. doi:10.1057/palgrave.ejis.3000608.

- Cameron, K., R. Quinn, J. DeGraff, and A. Thakor. 2006. *Competing Values Leadership: Creating Value in Organizations*. Cheltenham: Edward Elgar Publishing.
- Fullan, M. 2005. "Professional Learning Communities Writ Large." In *On Common Ground: The Power of Professional Learning Communities*, edited by R. DuFour and R. Eaker, 209–23. Bloomington, IN: National Education Service.
- Giddens, A. 1984. *The Constitution of Society. Outline of the Theory of Structuration*. Cambridge: Polity Press.
- Hyysalo, S., Jensen, T. E., and N. Oudshoorn. 2016. *The New Production of Users: Changing Innovation Collectives and Involvement Strategies*. London: Routledge.
- Kellog, K. C., Orlikowski, J. W., and J. Yates. 2006. "Life in the Trading Zone: Structuring Coordination across Boundaries in Post Bureaucratic Organisations." *Organisation Science* 17 (1):22–44.
- Kiran, A., N. Oudshoorn, and P. P. Verbeek. 2015. "Beyond Checklists: Toward an Ethical Constructive Technology Assessment." *Journal of Responsible Innovation* 2:1, 5–19.
- Kokkeler, B. J. M. 2017. "Living Labs Revisited: Enabling Tool or Policy Instrument for Social Innovation?" *Avans University of Applied Science*, Den Bosch.
- Kokkeler, B. J. M. 2014. "Distributed Academic Leadership in Emergent Research Organizations." PhD Thesis, University of Twente. doi:10.3990/1.9789036535854.
- Kokkeler, B. J. M., P. J. J. Bremmer, M. Glas, H. Kerkdijk, B. Hulsebosch, and F. Ebeling. 2008. Kernrapport haalbaarheid (ketenbrede) informatie-uitwisseling binnen de jeugdsector. Kamerstuk 31001-55-b1.
- Kokkeler, B. J. M., B. Lokerse, and W. Hesselink. 2012. *Kwaliteitsaanpak CJG Twente, een ontwerp op hoofdlijnen*. Amersfoort: BMC Advies.
- Kokkeler, B. J. M., and R. Van 't Zand. 2013. "De vierde man in het nieuwe speelveld van lokale informatievoorziening." In essaybundel transitie sociaal domein. Amersfoort: BMC Advies.
- Kokkeler, B. J. M., and B. J. Brandse et al. 2011–2015. Confidential Case Study Reports in Collaboration with Youth Health Organisations Combinatie Jeugdzorg, Rubicon, Jarabee, Ambiq, Carint, Tactus, Jeugdpartners Twente.
- Ministerie van Volksgezondheid, Welzijn en Sport. 2016. Voortgangsrapportage e-health en zorgvernieuwing (Progress report to Dutch Parliament). The Hague.
- Nooteboom, B., and E. Stam. 2008. *Micro-foundations for Innovation Policy*. Amsterdam: Amsterdam University Press, Dutch Scientific Council for Government Policy.
- Pettigrew, A. M. 1990. "Longitudinal Field Research on Change: Theory and Practice." *Organization Science*, 1 (3):267–92.
- Rijksoverheid.nl. 2016. "Mededeling decentralisatie van overheidstaken naar gemeenten." Retrieved from www.rijksoverheid.nl/onderwerpen/gemeenten/inhoud/decentralisatie-van-overheidstaken-naar-gemeenten.
- Rip, A., and J. W. Schot, 2002. "Identifying Loci for Influencing the Dynamics of Technological Development." In *Shaping Technology, Guiding Policy: Concepts, Spaces and Tools*, edited by K. Sørensen and R. Williams, 158–76. Cheltenham: Edward Elgar.
- Schön, D. 1983. *The Reflective Practitioner. How Professionals Think in Action*. London: Basic Books.

- Transitiecommissie Sociaal Domein. 2016. *Vierde rapportage transitie sociaal domein (TSD): Een sociaal domein*.
- Weick, K., and R. Quinn. 2004. "Organizational Change and Development: Episodic and Continuous Changing." In *Dynamics of Organizational Change and Learning*, edited by J.J. Boonstra, 177–96. Chichester: Wiley & Sons.
- Whitchurch, C. 2008. *Professional Managers in UK higher Education: Preparing for Complex Futures. Final Report for the Leadership Foundation for Higher Education*. London: Leadership Foundation for Higher Education.

10 Adversarial risks in social experiments with new technologies

Wolter Pieters and Francien Dechesne

Introduction

Ideally, all potential problems with new technologies are resolved in the design stage. By embedding values such as safety, security, and privacy in technological design, technologies could become less controversial and more acceptable to a wider range of stakeholders. Rather than being naysayers, ethicists and risk experts could work constructively with the developers to “make things better.” At the same time, there is justified skepticism about whether all potential problems can be conceived in such an early stage, and whether we can decide before deployment which technologies are acceptable and which are not.

Within this development, the literature that explores new technologies as social experiments points specifically to the responsibilities associated with deploying technologies in society (Van de Poel 2011, 2016). Rather than demanding full coverage of issues in the design stage, or a general laissez-faire with respect to technological developments, this approach focuses on conditions for responsible experimentation with technologies in the real world. By learning from small-scale experiments, the Collingridge dilemma—which states that at the time when effects become known, the technology can no longer be steered or redesigned (Collingridge 1980)—could (at least partly) be avoided. Thus, the deployment of new technologies in society is conceived as an experiment—albeit not a controlled experiment in the scientific sense—with associated responsibilities and acceptability conditions. Proposed conditions include monitoring, containment of risk, and conscious scaling up (Van de Poel 2011, 2016).

Studies that explore new technologies as social experiments have focused primarily on unintended effects in the deployment of new technologies (Pieters, Hadžiosmanović, and Dechesne 2016). By design flaws, nature, chance, and/or human error, threats associated with health, environment, and/or safety may materialize and unintentionally cause harm. For example, genetically modified organisms could escape into existing ecosystems and wipe out native variants, or a nuclear power plant could suffer a meltdown and radioactive material could cause health problems, pollution, and

even death. What is typically not explicitly considered in these studies is intentional misuse of new technologies.¹ In these cases, the new technologies enable new vectors for other actors to achieve their own goals, which may conflict with societal values. While nuclear technology has been described as social experimentation (Van de Poel 2011), its potential for adversarial use, such as the risk of nuclear material being acquired by terrorists to produce dirty bombs, has not yet been addressed. The origin of this risk arises not from probabilistic natural or accidental events but from the strategic behavior of actors with conflicting goals. Such actors are called “adversaries,” and the associated risks are called “adversarial risks” (Rios Insua, Rios, and Banks 2009).

In the cyber security research domain, understanding the adversarial aspects of new technologies as social experiments is essential to account for the societal damage associated with cyber attackers and attacks enabled by new technologies. In this chapter, we extend earlier work on this issue (Pieters, Hadžiosmanović, and Dechesne 2014, 2016) by developing a typology of adversarial risks in social experiments. To this end, we re-interpret adversarial risks using actor-network theory, thereby offering a new perspective on the difference between safety and security risks in social experiments in terms of (unintentional) events versus actions by actor-networks. We use the Bitcoin distributed digital currency, which enables money transfers without central control (i.e., without a bank) by distributed processing of transactions, as inspiration and case study to illustrate the difference between accidental and adversarial risks and their implications for social experiments.²

This chapter aims to point out the necessity of including adversarial risks in risk analyses of social experiments with new technologies. Our typology contributes to improved identification of such risks, to identification of methods for learning about these risks in pilots, and ways to manage adversarial risks.

Definition of adversarial risk

Risks of new technologies are often framed in terms of *safety* and *security*, commonly used in combination. Everyday use and dictionary definitions of those terms do not demonstrate a clear difference: both refer to protection against harm, hazard, and threats. In risk literature, however, a clear distinction is made. *Safety* is reserved for protection against *accidents*, i.e., unintentionally harmful events, for example, due to the breakdown of a worn-out part. *Security* is defined as protection against *adversarial threats*, i.e., deliberately initiated actions with harmful consequences, such as terrorist activities involving technology (see e.g., Aven 2007). In other words, safety refers to risks caused primarily by something that *happens* (an *event*), whereas security refers to risks caused primarily by someone who *acts*

intentionally (an *action*).³ The risks associated with safety and security are thus of a different nature: the first type of risks, which we will call “accidental risks,” is probabilistic, whereas the intentionality within the second type implies an *adversary* actor who employs a *strategy* to reach his or her goal.

Although most risks can be assigned to one of the two categories, the causal chain leading to harm can involve elements of both safety and security risks. For example, an accident can be caused by someone deliberately bypassing a safety procedure, and sabotage may be enabled by an accidental failure of an access control system. However, when someone intentionally bypasses a safety procedure, this is most likely not with the intent to cause a catastrophe. In this contribution on security, we restrict ourselves to risks in which the outcome was intended by the adversary actor, and the associated action was meant to cause this outcome.

The strategic nature of adversaries requires a different approach to mitigation than in the case of safety risks. If one is concerned about accidents involving a new technology, protection against one type of accident does not lead nature or those working with the technology to avoid protection mechanisms and try to cause another type of accident. Deliberate attempts to cause damage are exactly what we should be worried about in the case of adversarial risks: if we protect one part of the technological system against adversaries, they may direct their efforts towards another. For example, if we make it harder for adversaries to obtain nuclear material from a plant, they may try to take it during transport. Adversarial risks require *thinking thief*: reasoning from the point of view of the adversary (Ekblom 2014).

In this contribution, we are interested in ways to approach adversarial risks if we conceive of the introduction of new technologies into society as a social experiment. In particular, we are interested in cases in which the adversarial risks are enabled by the technology that is deployed. We use the term “adversarial risks” for the negative effects of the deployment of new technologies that are due to intentional adverse behavior with respect to that technology.

There are different interpretations possible of what constitutes the negative effects that characterize adversarial risks. According to one interpretation, a new technology comes with appropriate goals that make the technology (morally or otherwise) desirable. In this interpretation, *adversarial* refers to any intentional use of the technology for morally undesirable purposes, for example, in ways that cause harm to others, such as crime (illegally gaining advantages) or destabilization of society (war, terrorism). This is related to discussions on the ethics of dual use (cf. Miller and Selgelid 2007).

In a second interpretation, there are different possible expectations (Van Lente 1993; Borup et al. 2006) of how a new technology will and should be used both in the experimental stage and when deployed extensively. These expectations are often related to the goals or interests of the stakeholders

involved. From the perspective of stakeholders, the use of the technology that counteracts their expectations can be considered adversarial if it makes it difficult to achieve the associated goals. This means that there are always two points of view: that of those who deploy the technology, and that of potential adversaries who behave strategically to reach a goal that may conflict with the intentions of the initiators. If a stakeholder develops nuclear technology for peaceful purposes and it is then used by others in warfare, this conflicts with the goals of the stakeholder. This is a subjectivist interpretation in which the starting point is not a common good but rather the *goals* of a particular stakeholder or set of stakeholders.

Although readers with a background in ethics may be more inclined to support the first interpretation, the field of cyber security risk analysis is often not interested in whether the adversarial behavior is legal/illegal or moral/immoral, but in the fact that the behavior of the adversary, strategically directed at some adversary goal, harms the goals of other stakeholders, including those initiating the deployment of the technology. This is, for example, systematized in the form of misuse cases, describing possible sequences of actions with the system by misusers of a system that lead to an undesirable outcome for a stakeholder (Sindre and Opdahl 2005).

Both interpretations can be used to illustrate adversarial risks such as nuclear terrorism or cyber-attacks. These risks can be interpreted (a) as use of the technology that goes against proper, reasonable, or morally acceptable use, or (b) as use of the technology that counteracts the goals of the stakeholders that initiated the experimentation or deployment.

Adversarial risks in social experiments

Given this definition of adversarial risk, how can we leverage the understanding of this type of risk in studies that conceive of the introduction of new technology into society as a social experiment? In the cyber security domain, understanding the adversarial aspects of new technologies as social experiments is key (Dechesne, Hadžiosmanović, and Pieters 2014; Pieters, Hadžiosmanović, and Dechesne 2014, 2016). Dechesne, Hadžiosmanović, and Pieters (2014) discuss this for the case of smart electricity pilots in the Netherlands. In smart electricity networks, adversarial use may involve intentional remote disconnection of parts of the network in order to cause societal damage, or an illegitimate collection of privacy-sensitive data in order to advance one's own business interests. Adversarial risks may not become visible in small-scale pilots unless they receive specific attention. Pilots are considered successful if the technology works as expected, but possibilities for actors to misuse the technology—and associated security controls—are often not considered. This underscores the importance of pilot design for security-sensitive technologies in addition to the design of security controls. Earlier, a similar lack of attention for security created the electronic voting fiasco in the Netherlands where widely used electronic voting machines were

abolished after the possibility for adversaries to manipulate election results became clear (Jacobs and Pieters 2009).

A reason for these failures may be that much of the discussion in the cyber security field revolves around “security-by-design,” and focuses on embedding security (and privacy) in technologies *before* deployment (Cavoukian and Chanliau 2013). How to learn more about security *after* or *during* deployment, such as in the smart grid pilots, is often not thought through. This seems to suggest that the target is to get security completely right before experimenting with a new technology in society, which may not be realistic. Discerning exactly which adversarial risks may materialize around an emerging technology and which alternative uses are possible—via technical vulnerabilities or otherwise—may be too ambitious at the design stage. Pieters, Hadžiosmanović, and Dechesne (2014) therefore argue that we need security-by-experiment in addition to security-by-design: proper procedures for learning about security issues of a technology after deployment.

This work also outlined the implications of the approach of new technologies as social experiments for security, discussing electronic voting systems, smart grids, and public transport chipcards as examples. It has been shown how different conditions for responsible experimentation (Van de Poel 2011) could be used in the context of adversarial risk and security, but need adaptation or extension. The addition of responsible adversaries pointed out the necessity of adversarial roles in social experiments to enable feedback and learning about adversarial risk. Such adversarial roles could be played, for example, by ethical hackers: those who find security weaknesses in software and report them to those responsible instead of misusing them.

New techniques for responsible deployment, already in use in the cyber security community, may improve existing pilots with security-sensitive technology inside and outside the cyber-domain. Pieters, Hadžiosmanović, and Dechesne (2016) provide an overview of existing attempts at responsible deployment in cyber security, for example, in the form of responsible disclosure of security vulnerabilities discovered in software by hackers or scientists including incentives such as bug bounties (bonuses for reporting security weaknesses), which can serve as inspiration for dealing responsibly with adversarial risks in other technological contexts.

What is still missing in the literature is an analysis of what kind of adversarial actions can materialize as risks in social experiments with new technologies. Below we will distinguish different types of adversarial risks, using the Bitcoin distributed currency as an example.

Bitcoin

The Bitcoin distributed currency (Nakamoto 2008; Grinberg 2012) makes it possible to store and transfer digital assets without a central authority, such as a bank and/or a government. Instead, transactions are processed and verified by a network of computers with a complete transaction history being

stored as a distributive database: the so-called blockchain. In this section, we outline the basic features and implications of Bitcoin and how they led to adversarial risks associated with the deployment of the technology.

Bitcoin is a so-called *cryptocurrency*, i.e., a medium of exchange based on cryptographic mechanisms that serve to secure the transactions and control the total number of units (*bitcoins*) available in the system. The fact that the total number of bitcoins is designed to be limited gives Bitcoin its capacity to represent value in the way in which gold is used in gold standard currencies.

The central innovative technology underlying Bitcoin is the blockchain. New transactions are added to the chain on the basis of a consensus mechanism which involves increasingly difficult computations. New bitcoins are released as a reward for successful computations, but the reward size is halved by a rate inscribed in the protocol, thus limiting the total supply of bitcoins to 21 million, all of which will be in circulation around the year 2140. Given this “scarcity-by-design,” and in an analogy with the delving for gold, participating in these computations is called *bitcoin mining*.

The blockchain mechanism allows for a transaction history that is complete, accessible, censorship-resistant (the cryptography also provides anonymity to the owners), and tamper-proof, and all of this without centralized control. This means that blockchain-based technologies could be applied to replace any transaction-based service that currently requires trusted parties in finance (banks, credit card companies), land ownership (governments), and contracts (notaries), for example. By bypassing the need for centralized institutions for ensuring trust, blockchain-based technologies are more cost-effective as well.⁴ This potential for disruptive institutional change makes Bitcoin a key example of a social experiment with a new technology.

To exchange bitcoins for regular currencies and products or services, exchange platforms have emerged, and a small but increasing number of merchants have started accepting bitcoins. The first successful exchange service, Tokyo-based Mt. Gox, which served as exchange between bitcoins and dollars, went out of business in 2014 after hackers exploited their flawed execution of security procedures. The first marketplace to adopt Bitcoin fully as currency was the online black market Silk Road. It was shut down by the FBI in 2013, but successors have been launched since (Popper 2015). The Bitcoin social experiment thus entails adversarial risks that involve violations of moral goals and harm to stakeholder goals caused by intentional activities.

The Bitcoin case also illustrates that the adversarial aspects in social experiments may have a more complex structure than technical security weaknesses (“vulnerabilities”) such as the ones we addressed in our earlier work, and that further distinctions can be applied. The most obvious form of adversarial risks in Bitcoin is that, if it turns out that the Bitcoin protocol has technological flaws, adversaries may be able to obtain large amounts of money or even disrupt the currency altogether. A problem in the transaction

approval system caused a disruption in 2013,⁵ leading to a drop in exchange rates. Several security weaknesses have also been reported, most of them related to the exchanges: the interfaces between bitcoin and the traditional financial sector. Because of the effects on exchange rates, accidents, as well as intentional exploitation of security weaknesses, may affect participants not directly involved. So, adversarial activity within the Bitcoin system itself may lead to risks.

However, there is a second adversarial aspect here, namely that Bitcoin is used to enable anonymous money transfer for illegal activities (Moser, Bohme, and Breuker 2013). Because of the anonymity of the currency owners, criminals can use it for money laundering and, in general, to hide the link between money and criminal activity. The first type of risk (hacking Bitcoin) is induced by unknown flaws in the technology; the second (money laundering) by the possibilities of use even if the technology functions correctly.⁶ The second type involves use of Bitcoin by adversaries as a tool to gain advantage within a different sociotechnical system.

Thirdly, Bitcoin may be affected by external adversarial risks, such as botnets (networks of infected computers mining bitcoins) or malware (malicious software trying to steal Bitcoin wallets from users). When adversaries have access to other cyber-technologies, they may use them to take advantage of the Bitcoin system as well. In this type of adversarial risk, other sociotechnical systems (existing, deployed simultaneously, or deployed later) can be turned against the Bitcoin system by the activity of adversaries.

These distinctions show that adversarial risks in social experiments have to be understood within a broader context of the surrounding sociotechnical ecosystem. Of course, this is also the case with safety risks, but it is even more important in an adversarial context because adversaries may turn technologies and systems against each other *strategically*. If there is a possibility to gain advantage in the context of one system by exploiting another, this is not just possible but even likely in an adversarial context. The assumption of independence of harmful events no longer applies in such cases, precisely because adversaries act strategically.

Note that in the second subjectivist stakeholder interpretation of adversarial risks, Bitcoin could itself be seen as adversarial from the perspective of traditional institutions (banks in this case), as it basically makes them unnecessary. If, through continued experimentation with Bitcoin as a currency, it actually sustainably corrects known flaws in the banking system, the traditional banks—if they intentionally resist the development to preserve their position—would also become adversaries in the first moral interpretation of what is defined as adversarial. In the second interpretation, those with stakes in Bitcoin and the traditional banks would be each other's adversaries in the experiment because of their conflicting goals. Thus, experimentation itself may lead to shifts in what counts as adversarial risk. However, this interaction between experimentation and capturing adversarial risks is beyond the scope of the current chapter.

The Bitcoin case thus provides pointers towards distinguishing different types of adversarial risks. In the following section, we will strengthen the theoretical foundations of this distinction by reinterpreting adversarial risks within the framework of actor-network theory.

Actor-network theory interpretation

The Bitcoin case demonstrates that social actors can use new technologies for their own purposes unintended by the developers. New technologies give adversaries new opportunities. A technology may have unforeseen usage possibilities (such as in the case of security vulnerabilities), but even if it functions as planned, it may be used for unexpected purposes. The technology itself may also become a target of attacks executed with other technologies. In all these cases, technological features align with human intentions and capabilities in order to create uncertain future loss in the form of risks.

As discussed, such risks differ fundamentally from the safety risks usually considered because they stem from strategic behavior rather than from probabilistic events. Criminals may align with networks of infected computers to execute cyber-attacks, and terrorists may align with vulnerable critical infrastructures to cause disruption. In such cases, both the technological affordances and the human capabilities and intentions contribute to the actions. In order to study the processes by which such risks emerge in social experiments, we will look deeper into the configurations of humans and things that create such risks. To this end, we rely on actor-network theory (see e.g., Latour 2005) as a key theory on hybrid human-technology networks. Actor-network theory (ANT) makes it possible to consider adversarial risks in social experiments from the perspective of new alliances between actors (human and technological) made possible by the deployment of the technology, thereby changing action possibilities. Latour calls such generalized actors *actants*.

A key idea of ANT is that behavioral possibilities can be changed by technologies. In this contribution, we rely on Verbeek's (2005, 148–61) representation of Latour, as this provides a systematic summary of the key features needed for the interpretation of adversarial risk.⁷ Following Verbeek, we distinguish four different aspects of technologically mediated behavior: translation, composition, reversible black-boxing, and delegation. We highlight how these dimensions are relevant for understanding adversarial risks of new technologies. In illustrating our ANT interpretation, we will not restrict ourselves to the Bitcoin example. Later, we will return to the Bitcoin case.

First, new technologies enable new or modified possibilities for action for stakeholders. In ANT terms, their programs of action are translated. In Latour's example, a man with a gun has different action possibilities than the man or the gun by itself (Latour 1999, 176–77). Terrorists with access to nuclear materials may have different programs of action than those with

access to traditional weapons only. Similarly, a cybercriminal network has a different program of action than a traditional criminal network, with a new portfolio of criminal activities, targets, and mechanisms to prevent discovery.

Latour refers to this transformative process in terms of the distinction between *intermediaries* and *mediators* (Latour 2005, 39). An intermediary “transports meaning or force without transformation,” whereas mediators “transform, translate, [and] distort.” In adversarial risks, we have to think of possible implications in terms of mediators, where the effects of a new technology occur only via complex network effects and feedback mechanisms. Rather than composition of causes and effects, as in non-adversarial risks that induce failure probabilities, the alliances of actors in adversarial risks are more complex and they *transform* the effects. For example, cybercriminal networks are not an immediate consequence of the development of the Internet, but they have emerged gradually after a certain critical mass was achieved in, for example, Internet banking activity, which was then transformed into an opportunity for criminal business.

At the same time, translations of programs of action are only possible via the formation of new alliances via *composition*. Van der Wagen and Pieters (2015) discuss how hybrid cybercriminal networks are formed from the combination of criminal intent and technological possibilities. They introduced the notion of “cyborg crime” to refer to the actions made possible by such relations. For example, botnets (digital networks of infected computers) and their actions involve a complex combination of infection methods, victims, malware, control software, operators, and clients. The possibility for adversaries to form new networks with newly deployed technologies is thus another feature of security risks in social experiments as opposed to safety risks.

The newly enabled programs of action via composition of hybrid agency may oppose existing programs of action: they may form *antiprograms*. Latour (1991) uses the example of a hotel manager trying to make guests return their keys, whereas the guests have an antiprogram (not returning their keys). The hotel manager may invoke different types of technology (signs, heavy key chains) to translate and support the program. (For the guests, this is an antiprogram against their own program.) Adversarial risks are typically antiprograms working against an intended program of action, either for the actor-network around the new technology or for other actor-networks. For example, smart (connected) electricity infrastructures can be disrupted remotely to cause chaos, or nuclear waste can be used against existing societal programs in terrorist attacks. This approaches our second interpretation of adversarial risks, in which adversaries have goals that conflict with those of the champions of the technology. The analysis of the dynamics of programs and antiprograms is a

relevant instrument to understand adversarial risks because it allows us to describe how adversarial actions can emerge as antiprograms of networks composed around a new technology.

In *reversible black-boxing*, the composition of new human-technology configurations may become invisible: the complexity of the network becomes hidden behind an interface. We do not generally observe the complicated sociotechnical network that supports the Internet and associated connectivity, but simply use our browser to navigate it. When something breaks down, the black box may need to be reopened to find and resolve the problem, thus *reversing* the black-boxing. Similarly, adversarial activities around new technologies are subject to black-boxing. The antiprogram of cybercrime owes part of its success to black-boxing possibilities. In particular, black-boxing cybercriminal networks and activities in service offerings on the black market is key here. This enables a division of labor between different fields of expertise and different technological infrastructures, thereby enabling more complex actor-network formation in turn. For example, different forms of cybercrime, such as distributed denial-of-service (DDoS) attacks, are offered as a service to be purchased on demand, based on botnet infrastructures (Manky 2013; Santanna and Sperotto 2014). The programs of action become business models—marketable programs of action—in which results can be sold to other adversaries rather than monetized directly. New criminal interfaces and markets thus emerge around a new technology, black-boxing part of the complexity. To stop the associated crime, the black-boxing needs to be reversed in order to disable the networked actions. In probabilistic risks, by contrast, there is no intentional hiding of complexity or formation of service packages and business models. There are no adversaries that have interest in these forms of invisibility (although there may, of course, be actors that have incentives to hide the safety risks themselves, which can be seen as a special type of security risk).

Finally, the concept of *delegation* signifies that actors may inscribe certain features into technologies to make them support their program of action better. For example, speed bumps support a program of action that aims at speed reduction by making drivers reduce speed to avoid damage. The designers intentionally inscribe this support, but adversaries and their antiprograms may employ similar delegations. In the space of cybercrime, cybercriminals may delegate certain tasks to botnets by simply sending them commands. Part of the success of botnets (and of computers in general) lies in the fact that they can be instructed to support different programs of action depending on the needs. A more specific approach is visible in so-called “ransomware:” software that makes files on a computer inaccessible by encrypting them and then proposes to decrypt them for a ransom. Here, the blackmail mode of operation is delegated to a piece of software.

ANT thus provides a framework for analyzing adversarial opportunities created by currently deployed technologies as well as initial ideas on how to imagine (predict is probably too ambitious) possible adversarial risks in future social experiments. This can be done in terms of:

- Investigating how the deployment of new technologies could *translate* programs of action, notably those involving known adversaries such as criminals and terrorists;
- Investigating which new alliances could be *composed* through the deployed technology;
- Investigating the extent to which the deployment of new technologies offers adversarial opportunities that can be *black-boxed* in the form of adversarial (criminal) services offered to others;
- Investigating the extent to which adversarial programs of action can be *delegated* to the deployed technology.

More generally, various stakeholders will aim to inscribe certain features into a new technology to enable it to serve their own purposes. Therefore, technical as well as political lobbying (standards, laws) also plays a role in the adversarial risk picture. Here, the adversaries may not be only criminals and terrorists, for example, but also organizations that try to make new technologies “work” for themselves. We have seen certain Internet-focused companies become immensely powerful by forming alliances with the new technological possibilities. This raises concerns about moral values and acceptability, notably privacy and autonomy of users, as well as compatibility of the goals of these companies with goals of other stakeholders. The net neutrality debate illustrates the point of stakeholders trying to inscribe features beneficial to themselves. The issue at stake is whether Internet service providers may discriminate between different types of Internet traffic (thereby inscribing certain programs of action into the technology), and the debate is significant due to the different interests.

Typology for adversarial risks

We have seen earlier that the deployment of Bitcoin brings several adversarial risks, including exploitation of weaknesses in the system, using Bitcoin for dubious activities such as money laundering, and leveraging other technologies such as botnets for use against the Bitcoin system. Based on our ANT interpretation, we revisit the adversarial risks distinguished. This provides a typology of adversarial risks, as follows:

- 1 In case of attacking Bitcoin with, for example, botnets or malware, a new alliance (*composition*) is formed with other technology *against* the Bitcoin program of action (by means of an antiprogram). The program

of action of the other technology is *translated* in order to make it support the antiprogram against Bitcoin. Bitcoin is used by adversaries as a *target*, in this case made possible by the exploitation of technological flaws. We call this *adversarial targeting*.

- 2 In case of money laundering, an alliance with Bitcoin is formed against the program of action of other technological systems. Bitcoin is used as a *weapon or tool* (as part of the translated antiprogram), and the money laundering program is being (partly) *delegated* to the Bitcoin technology. Money laundering services can be offered as a *black box* to other criminals. The new affordances provided by the new technology are exploited by adversaries as a new tool. We call this *adversarial tooling*.
- 3 Finally, in the case of exploiting weaknesses in the Bitcoin system itself, a new type of alliance with the Bitcoin technology is formed in which the technological system is turned against itself (against its commonly known program of action), thereby *translating* the intended program of action. This is achieved via a new type of malicious use of Bitcoin.⁸ In this case, Bitcoin is used as *both* weapon and target (it is part of the program as well as the antiprogram). We call this *adversarial hijacking*.

In a more general sense, the complexity of adversarial risks is due, in part, to the fact that weapons and targets become hybrid as well. That one technology is part of both the program and antiprogram (Bitcoin being turned against itself through the exploitation of a vulnerability) is only one side of this hybridity. Another is when the target in one adversarial action becomes the weapon in the next. For example, by infecting computers with malware to make them part of a botnet (target), they can be used to execute DDoS attacks later (weapon). Black-boxing contributes to this complexity, as common attacks may be offered as a service such that they can easily be reused in other adversarial programs.

These examples focus on the use of the technology once deployed, i.e., risks related to the use of the technology for adverse purposes such as crime or destabilization of society. In addition, adversaries may have reasons to influence the development of the technology in order to make the technology support their own goals. To include both aspects of use and aspects of design, we propose a fourth type for adversarial risks of a new technology:

- 4 Exploitation of the evolutionary character of the embedding and deployment of the new technology within the actor-networks of institutions and governments. We call this *adversarial lobbying*.

For Bitcoin, examples of adversarial lobbying may lie in attempts to influence institutions to favor or to oppose Bitcoin adoption or distributed currencies in general. Long-term goals may include disruption of the existing financial system or the use of Bitcoin for adversarial targeting, tooling, or hijacking.

Adversarial lobbying operates on a meta-level because it comprises intentional attempts to blur potential undesirable or adverse effects.⁹ One could think of activities that (re)define the intended/unintended consequences of a technology thereby transforming adverse use into intended use. In addition to embedding Bitcoin in the financial system, imagine groups who lobby governments in favor extended use (so-called “function creep”), for example, to influence availability and confidentiality of personal medical records. Medical records can provide valuable information for both health care providers and commercial companies. Such use is currently hampered by laws protecting medical confidentiality as well as by concerns that corporations would abuse the information for financial gain, but lobbying may contribute to redefining intended use.

This typology of adversarial risks can serve as a basis for refining conditions for responsible experimentation (Van de Poel 2011, 2016). We highlight three conditions here, namely sufficient monitoring of the experiment, reasonable containment of risks in the experiment, and consciously scaling up the deployment. For adversarial targeting and tooling, a system of *monitoring* is important to detect unexpected exploitation of the flaws and adversarial affordances of the technology as soon as possible. Such monitoring aims at identifying *translations* of programs of action of the new technology and the associated *compositions* of actants in which the technology serves as a target or tool, for example, if Bitcoin is used for money laundering. The information gained from exploitation of the institutional embedding of the technology should provide information for the *containment of risks*, especially regarding the adaptation of the institutional environment of the technology (adversarial lobbying). This would, for example, illuminate the implications of Bitcoin institutionalization for traditional banks. Adversarial hijacking is most relevant for *conscious scaling up*, the gradual expansion of the experiment. Scaling up may increase the likelihood of adversarial risks as the value at stake for adversaries increases with scale (Herley 2014). For example, Bitcoin becomes increasingly attractive for cybercriminals as the amount of currency in the system rises. Scaling up may also alter the balance between hazards and benefits by shifting the definition of the program and antiprogram. Technologies may be used for non-adversarial purposes different than intended when scaled up, such as the use of Bitcoin in general webshops rather than only niche markets.

Discussion

Some aspects of adversarial risk that arise from our study merit further discussion. Firstly, the answer to the question “who is the bad guy?” is not always as clear, as in our examples. As new technologies can often support different societal interests, it may depend on the context which use can be considered adversarial. The brief and recent history of Bitcoin has demonstrated this (Popper 2016). Different views on the intended or desirable use

compete: is Bitcoin technology a solution to the inefficiencies and unreliability of the traditional banking system (by its decentralized character), or is it a means to avoid central oversight on transactions (by its anonymity)?

Similarly, conflicting views exist on the further development of encryption on communication channels and databases. On the one hand, there is the societal interest of national security, which requires that certain information be shielded from potential enemies (states or individuals). Encryption also enables individuals to meaningfully oppose their governments in oppressive regimes. On the other hand, criminals and terrorists may use encryption to hide their communication from authorities, supporting their adversarial program of action. Thus, both those trying to crack encryption (oppressive regimes) and those trying to use it (terrorists) may be adversaries. This demonstrates the dilemma of determining which behavior is considered a threat and who should decide this, making the identification of adversaries an inherently political endeavor (Fichtner, Pieters, and Teixeira 2016). A recent debate revolves around whether or not critical information technology providers can be required by law to provide backdoors (access) for government intelligence services when they deem it necessary. While this can be seen as a safeguard for certain adversarial risks, it also creates new adversarial risks: access to information through backdoors will be the prime targets for adversaries such as oppressive regimes and may afford them to use the technology as a weapon.¹⁰

Secondly, one could argue that our use of ANT is incompatible with our focus on human adversaries because it foregrounds human agency. This is not our intention. Rather, we point to the possibility of adversarial behavior by human-technology networks. Adversaries are necessarily hybrids, or “cyborgs,” precisely because both their human and their technological constituents enable their programs of action. Still, the distinction we make between intentional and unintentional events may not be fully compatible with a strict ANT interpretation of the system as a whole.

In essence, all adversarial risks, whether related to technical weaknesses or to lobbying for specific features, are based on shifts in power balances. Technical weaknesses may upset existing power balances between cyber-criminals and the police; designs with centralized data storage may upset power balances between citizens and companies or the government. In both cases, programs of action that previously controlled each other’s power are translated into forms in which this may no longer be possible, at least not without involving additional technological or human actors.

An important question is how our insights on responsible deployment in adversarial contexts can be used in practice. Embedding security aspects in pilots with new technologies can clearly be improved, but where do we draw the line between responsible piloting and irresponsible trial and error? The conditions for responsible deployment need to be applied in real-life case studies with security-sensitive technologies to define more specific criteria. In particular, simply waiting for real adversarial actions may not be

effective, as the adversaries themselves may wait until the technology is deployed more extensively. Therefore, adversarial roles (such as ethical hackers) need to be included in the pilots. The level of containment of effects is crucial here: how many security controls need to be put in place in the pilots to make them acceptable? There is a direct link here with the scale of the pilots and the associated possible impact.

Finally, introducing more security controls to prevent adversarial risks is not the Holy Grail. Security controls implemented in new technologies may have side effects themselves. The implications of security controls, particularly monitoring, on privacy are well known. They may also lead to non-adversarial risks: in the case of Bitcoin, the intentionally hard mining work causes high power consumption with associated environmental effects. The fact that an adversary might act is not always a reason to invest in controls (Herley and Pieters 2015). Thus, security controls themselves also require careful consideration of the advantages and disadvantages of their deployment.

Conclusion

We have shown that types of adversarial risks of the Bitcoin case can be described and understood in terms of ANT. The typology we developed can be used to identify adversarial risks, and the terminology of translation, composition, black-boxing, and delegation can be used to describe these risks. This provides a more thorough theoretical underpinning of the inclusion of adversarial risks in responsible social experimentation by focusing on actor-networks, their programs of action, and the role of the new technologies in these programs, rather than on accidental or probabilistic risks only. Instead of asking *what could go wrong* when deploying a new technology, we should also ask *which new programs of action* could be formed in relation to the new technology? How can it be used as a weapon, target, or both? How can the development be influenced strategically to make it work for particular stakeholders? We should pay attention to how such risks may affect moral values and/or how they may inhibit the supporters of the new technology in reaching their goals.

Acknowledgments

The research of the first author has received funding from the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement ICT-318003 (TREsPASS). This publication reflects only the authors' views and the Union is not liable for any use that may be made of the information contained herein.

The research of the second author is part of the SCALES project in the research programme Responsible Innovation (MVI) with project number 313-99-315. This programme is (partly) financed by the Netherlands Organisation for Scientific Research (NWO).

Notes

- 1 We substantiate this argument in more detail in Pieters, Hadžiosmanović, and Dechesne (2016).
- 2 That Bitcoin is in fact a true social experiment is explicitly recognized by initiator Satoshi Nakamoto, according to the account of the turbulent history of the currency by Nathaniel Popper (2016).
- 3 Note that errors by operators of technology count as accidents, as they are not performed intentionally.
- 4 For an optimistic view on applications of the blockchain, see e.g., <http://techonomy.com/2016/03/blockchain-global-growth/> (consulted May 8, 2016); for a more skeptical one, see http://www.slate.com/articles/technology/future_tense/2016/02/bitcoin_s_blockchain_technology_won_t_change_everything.html (consulted May 8, 2016).
- 5 To be more specific: two incompatible transaction histories circulated, a so-called *hard fork*—a detrimental problem for the blockchain. This was caused by an update of the protocol running in parallel with older versions. It was solved when a large conglomerate of mining power voluntarily downgraded to the older version to put a halt to the steep drop of value of the currency (and it did).
- 6 Wikipedia calls the former “Security” and the latter “Criminal activity”, <http://en.wikipedia.org/wiki/Bitcoin>, consulted February 25, 2016.
- 7 Note that we do not follow Verbeek in his moral interpretation of the mediation of action here; we just use his concise representation of what ANT is after. We believe that this is more helpful for explicating our points than reconstructing the intricacies of Latour’s (2005) own introduction, although we will cite Latour himself where needed.
- 8 A similar case is hacking public transport chipcards to enable free travel.
- 9 In fact, as the social experiment with the new technology evolves, the goals of the technology and what is considered to be desirable will shift, making it very difficult to distinguish which of these shifts are intended—and how to determine whether these intentions are adverse, and from whose point of view, given that power structures may shift as well.
- 10 A vivid discussion on this topic took place recently at the presentation of the PrivaTegrity system by David Chaum at the *Real World Crypto* conference in January 2016, see www.wired.com/2016/01/david-chaum-father-of-online-anonymity-plan-to-end-the-crypto-wars/ and <http://uk.businessinsider.com/david-chaum-privategrity-proposal-furious-debate-privacy-cryptography-privacy-cmix-2016-1>, consulted January 26, 2016. Another instance of this tension is the legal case in the United States of the FBI against Apple, when the company refused to create software that would allow the FBI to break into a terrorist’s iPhone basically by introducing a backdoor that could work for all iPhones. See e.g., the interview with Apple CEO Tim Cook in Time Magazine, March 28, 2016 (<http://time.com/4262480/tim-cook-apple-fbi-2/>, consulted May 11, 2016).

References

- Aven, Terje. 2007. “A Unified Framework for Risk and Vulnerability Analysis Covering Both Safety and Security.” *Reliability Engineering & System Safety* 92 (6):745–54.
- Borup, Mads, Nik Brown, Kornelia Konrad, and Harro Van Lente. 2006. “The Sociology of Expectations in Science and Technology.” *Technology Analysis & Strategic Management* 18 (3–4):285–98.

- Cavoukian, Ann, and Marc Chanliau. 2013. *Privacy and Security by Design: A Convergence of Paradigms*. Ontario: Office of the Privacy Commissioner (Ontario).
- Collingridge, David. 1980. *The Social Control of Technology*. New York: St. Martin.
- Dechesne, Francien, Dina Hadžiosmanović, and Wolter Pieters. 2014. "Experimenting with Incentives: Security in Pilots for Future Grids." *IEEE Security & Privacy* 12 (6):59–66.
- Eckblom, Paul. 2014. "Designing Products against Crime." In *Encyclopedia of Criminology and Criminal Justice*, edited by Gerben Bruinsma and Davis Weisburd, 948–57. New York: Springer.
- Fichtner, Laura, Wolter Pieters, and André Teixeira. 2016. "Cybersecurity as a Politikum: Implications of Security Discourses for Infrastructures." In *Proceedings of the 2016 New Security Paradigms Workshop*, 36–48. New York: ACM.
- Grinberg, Reuben. 2012. "Bitcoin: An Innovative Alternative Digital Currency." *Hastings Science & Technology Law Journal* 4:159–207.
- Herley, Cormac. 2014. "Security, Cybercrime, and Scale." *Communications of the ACM* 57 (9):64–71.
- Herley, Cormac, and Wolter Pieters. 2015. "If You Were Attacked, You'd Be Sorry: Counterfactuals as Security Arguments." In *Proceedings of the 2015 New Security Paradigms Workshop*, 8–11 September 2015, 112–23. New York: ACM.
- Jacobs, Bart, and Wolter Pieters. 2009. "Electronic Voting in the Netherlands: From Early Adoption to Early Abolishment." In *Foundations of Security Analysis and Design V*, edited by Alessandro Aldini, Gilles Barthe, & Roberto Gorrieri, 121–44. Berlin, Heidelberg: Springer.
- Latour, Bruno. 1991. "Technology is Society Made Durable." In *A Sociology of Monsters: Essays on Power, Technology and Domination*, edited by John Law, 103–31. London: Routledge.
- Latour, Bruno. 1999. *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, Ma.: Harvard University Press.
- Latour, Bruno. 2005. *Reassembling the Social - An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.
- Manky, Derek. 2013. "Cybercrime as a Service: A Very Modern Business." *Computer Fraud & Security* 6:9–13.
- Miller, Seumas, and Michael J. Selgelid. 2007. "Ethical and Philosophical Consideration of the Dual-Use Dilemma in the Biological Sciences." *Science and Engineering Ethics* 13 (4):523–80.
- Moser, Malte, Rainer Bohme, and Dominic Breuker. 2013. "An Inquiry into Money Laundering Tools in the Bitcoin Ecosystem." In *eCrime Researchers Summit (eCRS), 2013*, 17–18 September 2013, 1–14. IEEE.
- Nakamoto, Satoshi. 2008. "Bitcoin: A Peer-to-Peer Electronic Cash System." Accessed March 29, 2017. www.bitcoin.org/bitcoin.pdf.
- Pieters, W., Dina Hadžiosmanović, and Francien Dechesne. 2014. "Cyber Security as Social Experiment." In *Proceedings of the 2014 New Security Paradigms Workshop*, 15–18 September 2014, 15–24. New York: ACM.
- Pieters, Wolter, Dina Hadžiosmanović, and Francien Dechesne. 2016. "Security-by-Experiment: Lessons from Responsible Deployment in Cyberspace." *Science and Engineering Ethics* 22 (3):831–50.
- Popper, Nathaniel. 2015. *Digital Gold – The Untold Story of Bitcoin*, Penguin Books UK.

- Rios Insua, David, Jesus Rios, and David Banks. 2009. "Adversarial Risk Analysis." *Journal of the American Statistical Association* 104 (486):841–54.
- Santanna, José Jair, and Anna Sperotto. 2014. "Characterizing and Mitigating the DDoS-as-a-Service Phenomenon." In *IFIP International Conference on Autonomous Infrastructure, Management and Security*, 30 June – 3 July 2014, 74–78. Berlin, Heidelberg: Springer.
- Sindre, Guttorm, and Andreas L. Opdahl. 2005. "Eliciting Security Requirements with Misuse Cases." *Requirements Engineering* 10 (1):34–44.
- Van der Wagen, Wytse, and Wolter Pieters. 2015. "From Cybercrime to Cyborg Crime: Botnets as Hybrid Criminal Actor-Networks." *British Journal of Criminology* 55 (3):578–95.
- Van de Poel, Ibo. 2011. "Nuclear Energy as a Social Experiment." *Ethics, Policy & Environment* 14 (3):285–90.
- Van de Poel, Ibo. 2016. "An Ethical Framework for Evaluating Experimental Technology." *Science and Engineering Ethics* 22 (3):667–86.
- Van Lente, Harro. 1993. "Promising Technology: The Dynamics of Expectations in Technological Developments." PhD diss., Universiteit Twente.
- Verbeek, Peter-Paul. 2005. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park: Penn State Press.

This page intentionally left blank

Index

- action-guiding experiment 17–18; and control 19–24; and control paradigm 26–30
- actor-network theory 11, 229–32
- adaptability 3, 104, 117–19
- Adderal 127, 131
- adversarial risk 11, 222–5; in social experiments 225; actor-network interpretation of 229–32; typology for 232–4
- Alzheimer's disease 127
- antimalarials 179, 185
- antiprogram 230–2
- attention deficit and hyperactivity disorder 127
- bioeconomy 9, 103, 105, 120–1
- bioethics 134, 143
- Bitcoin 11, 226–9; adversarial risks of 232–3
- Boal, Augusto 88–9, 95
- brain 126–7, 143
- Campbell, Donald 7, 36–7, 47
- challenges: ethical 47–8; political 49, 51, 141; to evaluation 135–8; to goal-setting 134, to prediction 135–8; to the experimenting society 47–51, 173
- clash of technology platforms 201
- client-owned portfolio 216
- co-creation 201, 209, 218
- coercion 130, 139–40
- cognitive enhancement (CE) 10, 67, 125–7, 128–9, 132, 141
- collective experimentation 4, 50, 175
- collective interpretation 94–6, 99
- competition 92, 130, 133–4
- contrasting dynamics 204
- controlled experiment 4–7, 26, 72; *see also* controlled experimentation
- controlled experimentation 4, 37, 41–3, 49–52, 69–70; and evolutionary experimentation 43–4; and challenges to the experimenting society 47–50
- creative democracy 100
- cyber security 223, 225–6
- deliberate experimentation 3, 11–12, 115
- deliberation 63–5, 86–8, 99, 115; democratic 81; moral 63, 87, 99; public 82, 85
- dementia 127
- design experiment 5–6, 30–1, 36, 49, 70
- Dewey, John 9, 19, 63–4, 69–70, 83–6, 97–8
- diagnosis 11, 91, 180, 182–5
- dramatic rehearsal 9, 63–4, 81, 85–7; in theatrical debate methodology 97–9
- Ecover 9, 103, 107, 110, 119
- effectiveness 48, 126, 133–4, 194, 203
- electroceuticals 125, 127
- emerging experiment 202
- emerging technologies 80, 130, 140–1, 226; and ethical assessment 81–3
- epistemic experiment 17–18
- equality 107–8, 119, 126, 133–4
- ethical: assessment 81–3, 93; consequences 9, 74; hacker 226, 236; issues 2, 47–8, 60, 73–4; reflection 9, 80–1, 199; *see also* challenges
- ethical dilemma *see* moral dilemma
- ethics: Dewey's reconstruction of 84–5; experimental 96–7, 99; of moral experimentation 74–5; of technology 11, 141; pragmatist 87–8, 99

- evaluation 39, 71–4, 87, 135–6, 142
- evolutionary experiment 72–73;
 - see also* evolutionary experimentation
- evolutionary experimentation 9, 43–6, 47–9, 51–2, 70, 119
- experience machine 62–3
- experiment 2, 16–19, 36–41, 151–4;
 - control over 19–24, 27–30; definition of 18, 38, 40–1; *see also* action-guiding experiment; controlled experiment; design experiment; emerging experiment; epistemic experiment; evolutionary experiment; experiment in living; explorative experiment; field experiment; generative experiment; laboratory experiment; moral experiment; policy experiment; practical experiment; real-life experiment, real-world experiment; scientific experiment; simulation experiment; social experiment; thought experiment
- experimental approach 80, 103–5, 115
- experimenting society 36–7; challenges to 47–51
- experiment in living 60, 65–9; learning from 73; ethics of 74
- explorative experiment 72–3
- field experiment 6, 25–6, 42
- flourishing 82, 141; *see also* good life
- framework 11, 30, 84, 104; and social learning 108–10, 113–15
- free market competition 133–4; *see also* competition
- freedom 132, 138–40; *see also* liberty
- Fukushima disaster 1, 3, 149; and maps 156
- generative experiment 9, 44–6, 70, 72; *see also* generative experimentation
- generative experimentation 4, 9, 45–7, 70; and challenges to the experimenting society 47–9; learning from 72–3
- Global Health 179, 185
- good life 128, 135; and experiments in living 65–8, 74
- goods: individual 128, 132; social 128, 130; universally valued 130
- Google Glass 59, 67–8
- hybrid 27, 152–3; agency 230; network 229, 235
- ICT 10, 199, 201
- imagination 20, 136, 155, 174; in Dewey 64, 85–6; in theatrical debate methodology 98
- impacts 2, 59, 80; hard 130, 145n9; soft 129–30, 133, 145n9
- increased motivation pill 131
- informed consent 5, 74, 137–8, 153
- innovation trajectories 97, 103–4
- inquiry 19, 40, 70, 80, 84–5; and pragmatist ethics 84–8, and theatrical debate methodology 94–6
- laboratory 4, 17, 37, 115, 151–5, 180–2
- laboratory experiment 4, 20–2, 42, 45–6, 153
- learning 38–41, 43–4, 47–9, 59, 72–4; in practical experiments 29; in scientific experiments 29; institutional 2, 59; moral 2, 59–61, 72–4, 105–6, 120, 218; normative 2, 59; social 9, 104–5, 108–9, 113–15, 120; technical 2
- liberty 126, 140
- living lab 6, 71, 208–14
- malaria 10, 179–180; diagnosis of 182–5
- materiality 171
- matter out of place 171
- medical devices 134–5
- medications 125, 127–8, 135
- military drones 59
- Mill, John Stuart 65–6, 69–70
- modafinil 131, 135
- moral ambiguity 104–5, 115; and frameworks 108–10; and social learning 113–15
- moral dilemma 11, 94, 213–15
- moral experiment 59, 60–2, 200; types of 62–72; learning in 72–4; *see also* moral experimentation
- moral experimentation 3, 7–8, 59–60; ethics of 74–5; in youth care 215–17
- moral issues 2, 28, 59, 60–1, 214–15; in moral experiments 72–4
- moral learning *see* learning
- moral philosophy 9, 60–62, 69
- neuroethics 134
- nuclear disaster 150, 165

- ontological politics 153–4, 173
- participatory theatre 88
- perspectives 84, 109, 114, 120, 216
- pharmaceuticals 125, 134, 184, 189
- pilot projects 36–7, 46, 49, 205
- pluralism 83–4, 94
- policy experiment 36–7, 42, 48–9; *see also* policy experimentation
- policy experimentation 37
- political challenges of experimentation *see* challenges
- political neutrality 126
- political philosophy 143
- positional advantage 131
- practical experiment 8–9, 16–18; and control 27–30
- prediction 26, 135, 174; *see also* challenges
- prisoner's dilemma 139
- productive ambiguity 92, 99
- program of action 229, 232–3; *see also* actor-network theory; antiprogram; hybrid
- prototyping 36, 45–6, 51
- Provigil 127; *see also* modafinil
- public engagement 82, 93, 141
- radiation 150, 156, 163
- randomized controlled trials (RCTs) 6–7, 42
- Rapid Diagnostic Tests (RDTs) 179–81, 185, 194
- real-life experiment 4, 17, 98
- real-world experiment 4, 151–3, 174; *see also* real-world experimentation
- real-world experimentation 1, 46
- regulation 70, 105, 133, 152; and freedom 138–40; methodology for 140–2
- responsible experimentation 11, 219; conditions for 74, 222, 234
- Ritalin 127, 131
- safety 82, 126, 133, 223–4
- scaling up 115–17, 234
- scenarios 60, 63–5, 89–90, 97–98
- Schön, Donald 38, 41, 45
- scientific experiment 16, 51; and control 19–24; control paradigm for 24–6
- security 223–4, 235–6; -by-design 226; *see also* cyber security
- self-experimental society 152, 173; *see also* experimenting society
- side effects 129, 133–6, 189, 236; medical 129, 135–6; social 129, 134–8, 141; *see also* impacts
- simulation experiment 19, 80, 97
- social engineering 137
- social experiment 37, 59, 69–76, 126, 194, 222–3; and adversarial risk 225, 232; learning from 73–4; new technologies as 2–4, 11–12, 30, 59, 80, 222–3; *see also* social experimentation
- social experimentation 2–3, 59, 126, 142–3, 153
- social media 67, 199, 201–2, 215
- societal experimentation 1; *see also* social experiment; social experimentation
- societally desirable innovations 103
- sociomaterial 153, 174, 183
- sociotechnical imaginary 150–1, 172
- sociotechnical system 27–30, 228
- spaces for experimentation 210, 216–17
- spatial arrangements 152, 154
- spatial reordering 150, 156, 171–2
- sustainability 44, 103–5, 108; and worldviews 109
- tacit experimentation 1–3, 11–12, 75
- technopolitical culture 149–50
- technopolitical identity 149
- Tesla 5–6
- testing 21, 85, 179–80, 187, 190, 200; beyond the laboratory 180–1; of hypothesis 5, 20, 61–2, 72, 98
- the Netherlands 3, 9
- theatrical debate 9, 65, 80–1, 89, 94–5; methodology 87, 96–100
- thought experiment 7–8, 20, 62–5, 72–3, 81, 130; ethics of 50, 74; learning from 20, 73
- torture 63
- transcranial direct current stimulation 125
- transcranial magnetic stimulation 127
- trolley problem 62, 64
- Uganda 3, 179, 193–4; and Rapid Diagnostic Tests 181–2
- uncertainty 2, 38–9, 97, 104–8, 120–1, 161, 199–200
- user-initiated experimentation 211

Value Sensitive Design (VSD) 64

values 11, 82–3, 94–5, 120–1, 135, 141,
222–3; and frameworks 106–9, 114;
and moral learning 59–60, 71–2;
and worldviews 108–9; in youth care
203–4, 216

vector of innovation 217

work faster pill 130–1

work longer pill 131

worldview 87–8, 108–10, 120–1; in the
Ecover case 110–13

youth care 199–202; experiments in
203–14