THE FRAGILITY OF RESPONSIBILITY

NORWAY'S TRANSFORMATIVE AGENDA FOR RESEARCH, INNOVATION AND BUSINESS

*Edited by Giovanni De Grandis and
Anne Blanchard*

The Fragility of Responsibility

De Gruyter Studies in Innovation and Entrepreneurship

Volume 9

The Fragility of Responsibility

Norway's Transformative Agenda for Research,
Innovation and Business

Edited by
Giovanni De Grandis and Anne Blanchard

DE GRUYTER

Foreword

In April 2024 the OECD arranged a ministerial meeting on science, technology, and innovation in Paris, where the attending science ministers presented a "Declaration on Transformative Science, Technology and Innovation Policies for a Sustainable and Inclusive Future".[1] Behind this mouthful of terms hides a partial paradigm shift in contemporary research and innovation policies.

The ministers committed to:

> develop and implement transformative science, technology, and innovation agendas, as appropriate, that are more inclusive, agile, anticipatory, allow for policy experimentation and reflect socially relevant directions, to help achieve the Sustainable Development Goals, including for climate and the sustainable use of the Ocean . . .

Furthermore, the ministers acknowledged the need

> to balance the transformative potential of emerging technologies for providing novel solutions to global challenges and opportunities for sustainable economic growth against the ethical, safety and security risks arising from possible misuse or unintended consequences.

This is a clear reference to the need for a policy approach that includes responsible research and innovation (RRI), which is the topic of this book. In other words, RRI, in some form or another, is to become an integrated part of the overall policy approach to the role of research and innovation in social transformation.

This book presents many of the lessons learned within the framework of the Norwegian AFINO Research Center for responsible research and innovation, funded by the Research Council of Norway.[2]

Transformative innovation policy

The term "transformative innovation policy" has been around for some time now and has – for instance – had an effect on the EU development of "missions" for change.[3] Researchers have contributed with reflections on what the term may entail and the policy consequences it brings.[4]

---

1 *OECD, Declaration on Transformative Science, Technology and Innovation Policies for a Sustainable and Inclusive Future, OECD/LEGAL/0501.*

2 https://www.ntnu.edu/afino/

3 https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe_en

4 See, for instance, Schot, J., & Steinmueller, W.E. (2018). Three frames for innovation policy: R&D, systems of innovation and transformative change. *Research Policy* 47(9); and Diercks, G., Larsen, H., & Steward, F. (2019). Transformative innovation policy: Addressing variety in an emerging policy paradigm. *Research Policy* 48(4).

Basically, the concept refers to a shift in policy thinking from a linear reservoir model of research and innovation towards a challenge- and needs-driven model. We are gradually moving from a technology push narrative to a societal pull narrative.

Many will argue that a transformative innovation policy cannot be reduced to an innovation policy alone, given that many of the solutions to the grand challenges will have to be found in other policy areas. So far, however, this kind of thinking has been primarily owned by ministries of research and/or innovation. In the OECD, the Directorate for Science, Technology and Innovation has the lead.

The freedom of silos

In his chapter of this book, Matthias Kaiser gives a helpful analysis of how the idea of "value-free science" came to dominate Western science and science policy. According to this approach, science takes care of the facts, Kaiser argues, while the state takes care of the values.

This understanding has made it possible for scientist to argue that it is not their job to consider the consequences of their research. Even if the Second World War, with its Nazi "science experiments" and the American atom bomb, made it much harder to uphold this idea, vague concepts of "freedom" and "objectivity" continue to be used to protect scientists against this co-responsibility.

The advantage of black boxing research and innovation is that no one needs to take responsibility for the negative ways in which R&D may transform society.

Policy-makers can fund such activities without worrying too much about unintended negative consequences. Many of the important decisions are delegated to civil servants, researchers, and innovators.

Researchers can go on doing their thing, arguing that whatever problems their research may cause, they are caused by others, like users, companies, or politicians.

Companies can argue that the economy will, in the end, produce the results the society needs. They are often referring to the myth of the invisible hand of perfect markets.

There are many here who may have an interest in upholding the status quo, as it makes life easier for them.

An end to black-boxing of research and innovation

The traditional research and innovation policy paradigm has for a long time been dominated by this kind of black-boxing of science and technology.

Economists have seen technological change as an exogenous factor, outside the economy.

Science activists have argued for a social contract where scientists are given funds to follow their own interests, which – according to these activists – will lead to helpful knowledge and useful technologies in the end.

It is true that systemic innovation policies have, since the early 1990s, looked at the interaction between innovating companies and organizations on the one hand and the surrounding system of institutions, regulations, and funding mechanisms on the other. However, the relevant policy-makers and stakeholders have normally not questioned the premise that if you fund innovation, the result will be economic growth, helpful solutions, more jobs, and a better quality of life for all.

Mind you, this reservoir approach is still included in the new OECD declaration and its accompanying report.[5] This is in accordance with the nature of paradigm shifts and the compromises needed to produce consensus-based declarations like this one.

But these documents also reflect a new way of thinking, where the current onslaught of crises – as in climate change, the pandemic, pollution, increasing inequality, political instability, and the rise of fascism – require a political reorientation towards more immediate social and environmental needs.

OECD no longer believes that the science and innovation system will bring the needed solutions by themselves. Indeed, the OECD documents clearly recognize that practically all the crises we are now facing are at least partly caused by science and technology.

Research and innovation can no longer be placed in a box outside the economy or outside society. These activities are now seen as an integrated part of the social, economic, and cultural systems that create the problems we are facing, and if we are to *transform* society and make it more sustainable and responsible, we have to look at how research and innovation has been transforming society up till now and discuss how it may shape the society of the future.

The need for new skills

OECD recognizes that policy-makers, to a much larger extent than now, will have to develop an understanding of the social, cultural, economic, and technological processes that create our future.

---

5 OECD. (2024). *Agenda for Transformative Science, Technology and Innovation Policies*, OECD Science, Technology and Industry Policy Papers, OECD Publishing, Paris. Another document that mixes up the old reservoir paradigm with the new transformative one is the EC's Rome Declaration on Responsible Research and Innovation in Europe. (2014, November 25). https://digital-strategy.ec.europa.eu/en/library/rome-declaration-responsible-research-and-innovation-europe, where the goal seems to be to use RRI to make technology push policies work better.

This also means that the policy-makers will have to develop some kind of "futures literacy", i.e., skills that make it possible for them to understand how their preconceptions and prejudices shape the way they see the future. It is these preconceptions that, after all, lead them into repeating the errors of the past. They are locked into old mental maps and narratives.[6]

Given the complexity of the social, cultural, and ecological systems involved and the many actors affecting the transformation of these systems, it becomes clear that no single group can tell the rest what to do. There is not one pool of "experts" that have all the facts.

Moreover, as Kaiser points out, the uncertainties and risks of "post-normal science" and the existence of "wicked problems" mean that we all have to make decisions without having all the facts.

Indeed, given that these transformation processes require that we question the very ideas and "truths" that have brought us into the current situation, it is also clear that the overall goals for a transformative policy must be up for debate. Such processes require involvement of a wide variety of actors, not limited to researchers, industrialists, and policy-makers.

This means that the original way of addressing possible negative consequences of research, namely research ethics, will not solve this problem. Research ethics remains important, but it is normally seen as a way of making researchers follow specific norms and best practices as far as the immediate effects their research may have, for instance, on research subjects, vulnerable groups, and the environment. Research ethics does not cover the wider social and cultural interaction between research and society.

Researchers and industrialists will have to develop an insight into the complexity of science/innovation/society interaction, and to do that, they will often have to learn from actors that are normally not involved in such processes, including policy-makers, stakeholders, and citizens. They must develop common learning arenas with and for people who may, directly or indirectly, be affected by their research and innovation. And they must discuss the potential transformative power of this new knowledge or technology with them on a macro level.

The chapter of Harald Throne-Holst on including stakeholders and citizens in R&D provides some important insight into how you may establish RRI learning arenas that go beyond the dialogue between RRI "experts" on the one hand and researchers and innovators on the other.

Such processes also require an awareness of the values of those taking part, as Elin M. Oftedal, Tatiana Iakovleva, and Matthias Kaiser point out in their chapter on responsible innovation. They underline the importance of understanding and integrating the multifaceted values of users, municipalities, NGOs, and companies.

---

6 Miller, R. (2018). *Transforming the future: anticipation in the 21st century.* Routledge/UNESCO.

Moreover, as Siri Granum Carson points out in this book, we must make sure that such processes are more than just "legitimizing processes", i.e., processes aimed at gaining, sustaining, or regaining social legitimacy. "Greenwashing" or "Sustainable Development Goal Washing" should not be allowed to replace a genuine exploration of the effects of an institution's practices.

But even if researchers are given the skills needed to initiate such processes, it would be unreasonable to give them full responsibility for foreseeing and handling the possible negative consequences of their own activities. They are, after all, part of a larger system, which includes their own teams and units, institutions or companies, users and customers, funders and stakeholders, and policy planners and politicians, all of which have to be engaged.

This is in no way an easy task, as the main actors of the system operate in institutional settings where the main objectives and "key performance indicators" are publications, funding, policy goals, or company profits.

In his chapter, Giovanni De Grandis argues that the transformative ideals of RRI require a transformation of the research and innovation ecosystem and that the will to carry out such a transformation seems to be missing both in Norway and in the EU. OECD does not answer how the transformations of the research and business systems are going to be achieved either.

Can systemic resistance be overcome?

It took years for the previous paradigm shift promoted by the OECD, the national innovation systems approach, to get a foothold in many countries, and it never fully replaced the old linear paradigm.[7]

The more recent transformative innovation policy approach is just as complex as the national innovation system one. It will take time for the OECD to get its members to truly embrace it, because of institutional resistance and mental lock-ins, although the current realization that the systems are not producing what society needs may make it easier.

In the same way as there is a need to reestablish peoples' trust in democracy, there is also a need to reestablish their trust in science and innovation. Indeed, the two challenges are connected. This cannot be achieved through pro-science and innovation propaganda and greenwashing.

---

7 The OECD popularized the national innovation system approached by way of their Technology Economy Programme, which presented its main reports in 1991 and 1992. *Technology and the economy: the key relationships,* OECD 1992.

The current backlash against democracy and the attacks on an open and inclusive society make this even harder. As Atle Midttun points out in his chapter on companies squeezed between autocratic and democratic regimes, we see a totalitarian form of capitalism where the liberal-democratic values of Western-style corporate social responsibility are seen as some kind of cultural and political imperialism.

RRI may be seen as part of a so-called "woke" agenda that opposes profits and growth. Donald Trump's attacks against policies aimed at reducing CO_2 emissions can serve as an example of this.

This denial of the problems we are facing may also have consequences for the way researchers, investigative journalists, and whistleblowers can expose violations of ethics and of laws aimed at protecting, for instance, workers and marginalized groups. The chapters by Kristian Alm and Heidi Karlsen on whistleblowers and by Caroline D. Ditlev-Simonsen on the Norwegian transparency act may provide some important input in this context.

In its declaration, the OECD repeatedly underlines the role of common values, referring, I believe, to democratic values like human rights and the freedom of speech. OECD clearly sees the new transformation policy as part of a defence of Western democracy.

In other words, we have to look at RRI as part of larger transformative processes that help democracies evolve, which encourage them to handle their systemic dysfunctions in meaningful ways and force them to face threats against the progress we have seen so far.

Because of this, we need more research on how RRI has contributed to such processes up till now and what can be done to make RRI an even more effective approach in the time to come. There is a need for research on how responsible research and innovation should be carried out, research that can provide input to future RRI processes and to RRI researchers who are to assist scientists, innovators, and policymakers in their work.

Because of this, there is also a need for more insight into how we can communicate RRI to a wider audience. In their chapter, Anne Blanchard and Erik Bjørnerud present some interesting lessons from the AFINO Research School PHD courses. They underline the importance of nurturing a caring and supportive RRI community, "based on slow spaces for building relationships and reflections." They see RRI as an iterative learning process with deeply transformative aims.

As Christian Wittrock and his colleagues point out in their chapter, the Research Council of Norway has played an important role in the funding and coordination of efforts in this area in Norway. Norway is, despite its small size, one of the biggest contributors to RRI research, and there is a lot to learn from what has already been taking place when preparing for RRI as part of social transformation. This is knowledge that researchers and policy-makers from other countries can make use of.

This book can play an important role in this respect, as it provides valuable insight into relevant challenges and opportunities, for RRI in particular and for social transformation in general. And to the extent RRI research has failed to provide answers or to establish new practices that work (as Giovanni De Grandis discusses in his chapter), those failures may be made visible in learning processes that lead to more effective ways of using RRI in social transformation.

Per M. Koch, Special Adviser, NIFU, and editor of Forskningspolitikk

Acknowledgements

The first authors of the following chapters have discussed together the central themes of the book and the content of each chapter in a series of meetings from April 2023 to January 2024. All the authors are therefore grateful to their colleagues for very fruitful exchanges and mutual comments.

Some authors have received especially valuable help from other people as well.

Christian Wittrock and his co-authors would like to thank Ellen-Marie Forsberg and Harald Throne-Holst for their insightful comments on improving the chapter. In addition, they thank Karl Ingar Kittelsen Røberg for crucial aid in identifying available comparative measures for the bibliometric part and Dag Øivind Madsen for providing the impetus for their bibliometric study.

Kristian Alm and Heidi Karlsen would like to thank their research assistant, Maiken Bergheim, for her invaluable aid in gathering information during the initial research phase for their chapter.

The work of AFINO and therefore also the writing of this book was supported by the Research Council of Norway.

Contents

Section 3: **Practices: fragile or robust?**

The authors

Professor Siri Granum Carson
Norwegian University of Science and Technology
Department of Philosophy and Religious Studies
Trondheim
Dragvoll campus 7491
Norway

Professor emeritus Matthias Kaiser
University of Bergen
Centre for the Study of the Sciences and the
Humanities
Bergen
Parkveien 9
5007
Norway

Christian Wittrock
Researcher
OsloMet
Centre for Welfare and Labour Research
The Work Research Institute
Oslo
Stensberggata 26
0170
Norway

Dr Giovanni De Grandis
Senior advisor
Norwegian University of Science and Technology
Department of Philosophy and Religious Studies
Trondheim
Dragvoll campus 7491
Norway

Dr Anne Blanchard
Researcher
University of Bergen
Centre for the Study of the Sciences and the
Humanities
Bergen
Parkveien 9
5007
Norway

Professor Atle Midttun
Norwegian Business School
the Department of Innovation and Economic
Organisation
Oslo
Nydalsveien 37
0484
Norway

Dr Harald Throne-Holst
Oslo Metropolitan University
Centre for Welfare and Labour Research
Consumption Research Norway
Oslo
P.O.Box 4
St. Olavs plass
0130
Norway

Professor Elin Merethe Oftedal
University of Stavanger
Faculty of Social Sciences Department of Media
and Social Sciences
Stavanger
Elise Ottesen-Jensen's House 3rd floor
Kjell Arholms gate 37
4021
Norway

Associate professor Kristian Alm
Norwegian Business School
Department of Communication and Culture
Oslo
Nydalsveien 37
0484
Norway

Dr Heidi Karlsen
Senior Academic Librarian
Norwegian Business School
Library
Oslo
Nydalsveien 37
0484, Norway
Senior Lecturer
University of Oslo
Department of Linguistic and Nordic Studies
Oslo
Henrik Wergelands hus
Niels Henrik Abels vei 36
0371
Norway

Professor Caroline Dale Ditlev-Simonsen
Norwegian Business School
Department of Law and Governance
Oslo
Nydalsveien 37
0484
Norway

List of abbreviations

AFINO	Ansvarlig Forskning og Innovasjon i NOrge (Responsible Research and Innovation in Norway)
AI	Artificial Intelligence
AIRR (framework)	Anticipation, Inclusion, Reflexivity, and Responsiveness
BRICS (countries)	Brazil, Russia, India, China, and South Africa, recently joined by Iran, Ethiopia, Saudi Arabia, United Arab Emirates, and Egypt
CR	Corporate Responsibility
CSR	Corporate Social Responsibility
CSOs	Civil Society Organizations
CSPs	Communication Service Providers
CUDOS (norms of science)	Communism, Universalism, Disinterestedness, Organised Skepticism
DLN	Centre for Digital Life Norway
ESG	Environmental, Social and Governance–oriented investment
FDPR (entity list)	Foreign Direct Product Rule
GONGOs	Government-Organized Non-Governmental Organizations
ICSU	International Council for Science
ICT	Information and Communication Technology
IPA	Interpretative Phenomenological Analysis
ISC	International Science Council
NAV	Norwegian Labour and Welfare Administration
NSCC	Norwegian Smart Care Cluster
NSCL	Norwegian Smart Care Lab
NUSAP (uncertainty communication)	Numeral, Unit, Spread, Assessment, Pedigree
OECD	Organisation for Economic Cooperation and Development
R&D	Research and Development
RCN	Research Council of Norway
RI	Responsible Innovation
RRI	Responsible Research and Innovation
SDGs	Sustainable Development Goals
SMEs	Small- and Medium-sized Enterprises
STS	Science and Technology Studies

Giovanni De Grandis and Anne Blanchard

Introduction

The spirit and point of this anthology

This anthology aims to explore the current Norwegian context of implementation of Responsible Research and Innovation (RRI) and Corporate Social Responsibility (CSR), as well as the challenges and fragilities associated with it. It is grounded in the experience of a networking and learning centre called AFINO (acronym for 'Ansvarlig Forskning og Innovasjon i NOrge', or Responsible Research and Innovation in Norway), to which most of the authors of this book are affiliated. The AFINO Centre was established in 2019 for a 5-year time period and is funded by the Research Council of Norway (RCN) through the programme SAMANSVAR on Responsible Innovation and Corporate Social Responsibility. Before we go into AFINO's mandate, and into the unique context of RRI in Norway, we would like to explain our motivations for writing this book.

Literature on Responsible Research and Innovation (RRI) and Corporate Social Responsibility (CSR) is surely not scarce. The fact of bringing these two topics together is far less common, but it is not the only defining feature of this book. This collection is characterized by some unique features. To begin with, it is the product of a group of authors who have worked together in AFINO over the past 5 years. So, the authors have had the chance of collaborating and sharing their experiences and challenges met in their efforts to understand and implement RRI and CSR in a variety of contexts. AFINO was not primarily a centre that aimed at producing scholarship on RRI and CSR, but rather at promoting their practice, coming up with new methods and nurturing relevant skills, such as transdisciplinarity (see Chapters 2 and 5), and, last but not least, nurturing a deeper dialogue and learning process between RRI and CSR researchers and the main research policy institution: the Research Council of Norway (RCN) (see Chapter 3). This particular role of AFINO positioned the authors close to the implementation level but, at the same time, made them very alert to the policy implications.

Another characteristic element of this anthology is that the authors are all grounded in the unique Norwegian context. Together with the Netherlands and the UK, Norway is one of the three nations in Europe where RRI was embraced by major research funders (see Chapter 3), while in terms of CSR, Norway has a strong public ownership of companies that creates an expectation for transparency and social responsibility (see chapter 1). Furthermore, there is a tradition of collaboration and trust between businesses and the public. In short, both RRI and CSR have been accepted early and supported in the country. This of course could make Norway look like an *uninteresting* case, because it is so atypical and so uniquely privileged. Yet, when we decided to write this book, our mood was far from optimistic. The prospects of CSR and RRI seemed gloomy even in Norway.

CSR seemed to be the victim of the rise of sustainability and of the success of the Sustainable Development Goals (SDGs). RRI seemed for a while – erroneously, see Chapter 3 and the next section – to have disappeared from the Research Council of Norway's agenda, and the Council itself was in full turmoil. Furthermore, we felt that many of our initial expectations had not been fully met through our work in AFINO. In particular, we would have liked to engage more thoroughly in an ongoing learning process and dialogue with the RCN, on the challenges of implementing RRI in practice, and despite several fruitful discussions together, the shared learning between AFINO and the RCN is, in hindsight, somewhat scarce and fragile.

These feelings and thoughts convinced us that there was a book to write. We wanted to explore where these challenges, doubts, or indeed fragilities around RRI and CSR come from. We thought that looking at them from the privileged perspective of Norway could reveal the core challenges, those that, even in very favourable circumstances, are bound to emerge. Because our position is close to the practice and implementation but also mindful of policy issues, it was natural to focus on what happens at the level of concept translation into practice, and in particular at the gaps between aspirations and actual results, or between intentions and outcomes, as well as the resulting fragilities in the implementation of RRI and CSR. This focus on fragilities is the most characteristic feature of this collection.

Most of the experiences and the reflections collected in this anthology revolve around this central theme, but they do it at different levels. Some stay close to the first-hand experience and explore the issues, solutions, and struggles at the practical level. Some start from issues that emerged in practice or in case studies to move the reflection from the practice to the operational level: to the implementation mechanisms and the contexts and institutions in which and through which they operate–from research projects to the law. In some cases, the reflection reaches the level of policy and politics. What is often at the centre of the reflection is the dialectic between the forces that promote RRI and CSR and the resistances and contingencies that twist the achievement of the objectives. So, the book does not develop theories or propose new interpretations of concepts, nor does it provide solutions and practical guidance, although we conclude with some policy-oriented recommendations. The common thread along the chapter is the attempt to understand the dynamics of translation and implementation, to bring into the light the critical junctures where things may derail or at least need attention, resourcefulness, or just some good luck. To talk of good luck is not just a pun to lighten the discourse, but it is also a reminder of the role of contingency in engaging with complex systems and a reminder of the limits of rationality, both in planning and in explaining events in retrospect.

Whether the authors are trying to illustrate a fragility or challenge, to understand its origins, or to illustrate how they have been tackled or how things developed in unexpected ways that raise new fragilities and questions, we believe that all chapters contribute to bringing to light the richness of the domains of research, innovation, policy, and business. This book offers the reader a journey through RRI and CSR in

Norway, where they will encounter many tensions and challenges, different styles and theoretical backgrounds, thematic affinities as well as analytical dissonances, many doubts, ambivalence, and open questions. We trust that readers will emerge from the journey with a better sense of what makes RRI and CSR important concepts, despite their flaws and imperfections, and where their fragility comes from. It is a journey that we believe will be valuable for all those interested in RRI and CSR but also for those more generally interested in the relations between science, policy, business, and society. The governance of R&I is not a hard science but more akin to an art or a craft. Arts and crafts need the ability to rethink and to change ideas, methods, and approaches. We therefore invite the readers to see more shades, to ask new questions, and to see the learnings around fragilities as sources of resilience.

The unique Norwegian context of RRI and of AFINO

As written above, this book looks at the current Norwegian context of implementation of Responsible Research and Innovation (RRI) and Corporate Social Responsibility (CSR) and is grounded in the experience of the AFINO Centre, funded by the Research Council of Norway (RCN). It is important to understand first why AFINO was funded by the RCN, as it illustrates the unique focus on and ambition for RRI in Norway. The Norwegian context stands in high contrast with the European context, where, as some of our colleagues wrote in their recent book on the implementation of RRI in an EU context: "RRI is dead (at least as an EU policy concept), and it is not yet clear what will come after" (Völker et al., 2024; p.x). But note that there was a time (in 2022–2023) when many (including ourselves) thought that the same was true for Norway. This worry has now been dispelled.

As of today, RRI is very much alive and supported by the Research Council of Norway (RCN), with its recently updated web pages on RRI, which encourage RRI to be implemented in the RCN-funded projects, and with the outstanding call for Research Centres for Artificial Intelligence, where the RCN will channel 850 million NOK and is expecting attention for RRI principles in the applications. The RCN's commitment to responsible research in a broad sense can be traced back to 2008 (Gulbrandsen and Rynning, 2016), with the launch of the ELSA programme on Ethical, Legal, and Social Aspects of new and emerging technologies (2008–2015). This programme was key in developing the concept of Responsible Research and Innovation (RRI) and contributed to shaping the RCN's main strategy document that asked for RRI to be incorporated into the projects funded by the RCN. After the ELSA programme, the SAMANSVAR programme on Responsible Innovation and Corporate Social Responsibility was developed in 2014, with a focus on research about – and research as – socially responsible innovation (Gulbrandsen and Rynning, 2016). Over that period, the RCN established its framework for RRI (RCN, 2019), based on the four dimensions outlined by Stilgoe and colleagues (2013): anticipation, inclusion, reflection, and responsiveness.

The commitment of the RCN to RRI, as they explain it themselves, comes from an awareness that science, innovation, and technologies transform social, economic, and political structures in profound ways (RCN, 2023). From this entanglement, sociotechnical imaginaries and futures are created, which give a direction to research, innovation, and policy (Jasanoff and Kim, 2015). This has led the RCN to engage with so-called third-generation R&I policies (Arnold et al., 2019) and to define itself as a social actor that should be involved in discussions about the transformative power of research and innovation, and where these processes are taking us, and reflect on the potential social, environmental, economic, and ethical aspects, as well as impacts and benefits of new research and innovation, in particular, who might be benefiting from it and for whom might it be detrimental (RCN, 2023). To clarify the difference, the first generation of R&I policy "involves funding research and essentially delegating the choice of theme and quality control to the scientific community in the expectation that societal benefits will eventually appear", the second generation of R&I policy "focuses on funding research and innovation in order to get specific societal benefits, especially economic growth" (Arnold et al., 2019; p.3), and the third generation of R&I policy focuses on addressing grand societal challenges in responsible and sustainable ways. This demands that the broad spectrum of R&I actors is involved in R&I processes, including researchers and innovators, industry and private sectors, civil society, and, as argued by the RCN, funding and research policy institutions. This is a demanding way of organizing and governing R&I processes, which leads to important challenges. The main challenge is the 'transformational challenge': namely, how to concretely achieve the desired structural transformations towards more sustainable, fair, and robust economies and innovation systems through R&I processes that are able to address complex and uncertain grand societal challenges. In concrete terms, this transformational challenge translates into difficulties inherent to establishing a shared vision regarding the goal and direction of the transformative R&I process, the lack of coordination across different policy levels (from regional to national and European), or again a lack of adaptive policy portfolios and spaces for experimentation and learning about how to address complex and uncertain grand societal challenges[1] (Arnold et al., 2019).

Learning about how to engage in third-generation R&I policies is fundamental, and a central objective of the RCN in supporting RRI has been to establish such networking and learning arenas that build on the varied experiences of R&I actors and create new partnerships. It is precisely for that purpose that the AFINO centre has been funded[2]. AFINO was funded as a way to experiment and test the ground for con-

---

1 For an extensive mapping and discussion of the different challenges and failures inherent to third generation R&I policies, ranging from directionality failures, demand articulation failures, and policy coordination failures, see the Technopolis Report by Arnold et al., 2019; pp.11–12.
2 Similarly, three years before AFINO, in 2016, the RCN funded the Centre for Digital Life Norway (DLN), which has been a pilot for testing transdisciplinary cooperation and RRI in the context of biotechnology research and innovation.

cretely engaging in the third generation of R&I policy and nurture partnerships and methods to navigate the associated challenges of such R&I governance. The objective of AFINO was therefore to build new networks between research, innovation, private sectors, and research policy institutions and develop expertise on how to ensure that R&I processes address grand societal challenges in responsible and sustainable ways, giving attention to societal values, needs and impacts, and ethical dimensions of scientific developments and technological innovations. Across the different chapters, it will become visible that AFINO has looked at the implementation of RRI and CSR in different domains, ranging from developing skills and knowledge through, for instance, training for post-graduates (see Chapter 5) to more concrete domains of application, such as involving patients and users in medical innovation projects (see Chapter 8), or looking at the impact on CSR of some relevant Norwegian laws (see Chapters 9 and 10), to only name a few.

The content of the book

Every chapter in this book opens with an abstract of its content where readers can find a summary of their main arguments. Here we provide an overview to specifically present how each chapter contributes to the main theme of this anthology – the fragility of RRI and CSR. The chapters are organized into three sections. The first section, titled 'The emergence of Responsibility', provides context by situating RRI and CSR in Norway within long-term trends in the relations between society, science, innovation, policy, and business. Siri Granum Carson, who is the AFINO Centre director, opens the book with **Chapter 1**, titled "The institutionalisation of social responsibility in Norwegian business and research – moral progress, moral decay, or both?" This chapter interprets the rise of CSR and RRI as responses from the business and research sectors to their legitimacy and role being challenged by society at large. The growing public awareness of the role of business and research in tackling – and generating – environmental and social problems put pressure on them and brought attention to their processes and purposes. This perception of being under public scrutiny opens up to different (com-)possible developments of CSR and RRI. They can contribute to more reflective and transparent practices, making science and business more attentive to social and environmental concerns, but they can also become cosmetic practices that affect the public presentation of business and science and fail to enact the deep transformations that the challenges of our times require. So, here we are presented with a classic fragility: if a change is more apparent than real it may easily turn into a problem rather than a solution. But, on the other hand, it is hard to tell whether small changes are the beginning of an incremental transformation and they need tending and care rather than harsh and impatient criticism.

In **Chapter 2**, titled "From value-freedom to responsible research and innovation? – Post-normal and transdisciplinary pathways", Matthias Kaiser relies on his

readings of history and philosophy of science to go through the emergence of the social responsibility of science – which he extends to the ethics of science – in its many expressions. He portrays the social responsibility of science as an attempt to overcome the dogma of value-free science and address tensions between knowledge and power in research and innovation processes. The author illustrates the development of social responsibility with examples from the Norwegian context (referring, for instance, to the CO_2 capture project in Mongstad). Commenting on RRI as one form of social responsibility of science, Kaiser highlights tensions between scientific knowledge and its relation to existing power structures in society. He argues that these tensions create fragilities when engaging with responsible research, and that there is a need to more closely examine those fragilities, as a starting point to discuss a new take on the ethics of science, moving from a theoretical and rather abstract enterprise to a more practical type of ethics, focusing on the processes of research and innovation (characterized by tensions, doubts, failures, and dilemmas) rather than an overwhelming concern for outcomes.

Chapter 3, titled "Norwegian engagement with RRI and the propagation of RRI by the Research Council of Norway", is a transition from contextualization to a more empirical analysis. Christian Wittrock and his colleagues show through a bibliometric analysis that Norway occupies an internationally outstanding position in the production of academic literature on RRI. The analysis then moves to the organization that dominates research funding in Norway – the Research Council of Norway (RCN) – and illustrates in detail the initiatives and programmes through which it promoted RRI. The study also considers the evaluation of the programmes promoting RRI, their relative scale in relation to total funding, and the view of advocates of RRI within the research council that meanwhile experienced a serious political crisis. This critical part of the analysis provides insights into the fragility of RRI. In this respect, the key contribution of the chapter is its ability to expose the limits of indicators of success, which, taken in isolation, can give very different impressions about how successful the uptake of RRI in Norway has been. The broad range of the analysis still raises a host of important questions for further studies, which the authors meticulously illustrate. This difficulty in achieving a comprehensive and well-grounded evaluation of the success of RRI clearly poses questions about how a policy can be furthered and sustained rationally, when feedback is incomplete and far from unequivocal. This is probably a more disturbing fragility than the more apparent one represented by the exposure of the research funder to unpredictable political attacks.

The second section of the book, titled 'Contexts of fragile responsibility', discusses some specific challenges encountered by RRI and CSR. The authors' analysis is located at the operational, policy, and political levels and presents issues that are not exclusive to the Norwegian context but have wide relevance. The section opens with **Chapter 4**, titled "The elusive transformation of research and innovation. The overlooked complexities of value alignment and joint responsibility". This chapter brings the analysis of two challenges for RRI at the philosophical level. Giovanni De Grandis pos-

its a contrast between what he sees as an ambitious transformative agenda of RRI and uptake and institutionalization that are still marginal, homeopathic in his own words. This gulf between ambitions and achievements is explained in part by the inadequacy of the means devoted to achieving the ambitious aims of RRI, which were largely left to the limited resources of junior researchers. But the bulk of the chapter is an ambitious philosophical analysis of the difficulties of achieving value alignment and collective future-oriented responsibility, two ideas that De Grandis sees as fundamental pillars of RRI. To put it very simply, in both cases the changes needed to align value and create a sense of joint responsibility for future outcomes hit against the existing socio-economic reality of R&I and how this latter shapes, at a deep level, the values of situated individuals and the obligations and constraints of organizations. The changes needed for transformative RRI are often incompatible with the running logic of existing institutions, and this explains why the chapter suggests scaling out RRI to "relocate" RRI projects outside existing infrastructures.

In **Chapter 5**, titled "Navigating tensions around RRI in higher education", Anne Blanchard and Erik Bjørnerud contribute with their own experience in designing and teaching a series of PhD courses on the topic of RRI, as part of the AFINO research school. They share the tensions met by the post-graduate participants that create ambivalences and fragilities when early career researchers decide to engage in RRI but lack support for it. Those tensions, expressed by the PhD course participants over the years, centred on navigating short-term temporalities, the expectation for quick results, narrow merit-based criteria, and the gap between intentions and outcomes in RRI projects. They mostly stem from conflicting demands from research and policy environments and lead early career researchers to internalize narrow criteria for what counts as a 'successful' RRI project and what could be accomplished in the duration of a PhD or post-doctoral project. What the authors argue is that a key to navigating these tensions, gaps, and fragilities is to nurture a caring and supportive RRI community, based on slow spaces for building relationships and reflections, for being attentive to the sometimes-conflicting demands stemming from research, policy, and society and for reconnecting with the essence of RRI, as an iterative learning process with deeply transformative aims. The authors conclude with early reflections on how exploring these tensions is helpful in shaping new hybrid and anti-fragile partnerships across RRI policy and research environments specifically, in order to discuss and address more profoundly the tensions and gaps in RRI projects.

The second section ends by turning to CSR and to a more global outlook. In **Chapter 6**, titled "Companies squeezed between autocratic and democratic regimes", Atle Midttun looks at how Corporate Responsibility (CR) is now operating in the context of a global political economy that has become increasingly marked by a new, bipolar rivalry between democratic and autocratic states – this polarization is at the root of the fragility he explores in his chapter. As the world moves towards a bipolar contestation between democratic and autocratic regimes, aggravated by the war in Ukraine,

the author argues that it is time to adapt CR to these new bipolar realities. In particular, Midttun claims that this shift affects the very concept of CR, which calls for a major revision. In order to discuss this, the chapter relies on three case studies of companies that have been exposed to controversies stemming from the democratic-autocratic divide, including the Nordic case study of the apparel company H&M, squeezed between human rights based pressures from Western stakeholders and Chinese authorities pressures over social conditions in the Uyghur region. In this context of polarisation and tension, creating a fragile context for the implementation of CR, the author argues for a cautious and nuanced interaction between governments and businesses, with CR being based on openness and cooperation.

The third section of the book, titled 'Practices: fragile or robust?', looks at some of the practices and mechanisms through which RRI and CSR are actually pursued and implemented. The first two chapters look at the ideal and practice of stakeholder – or citizen – involvement in RRI, but do it at different levels: the first looks at different understandings and rationales for involvement, while the second delves into the concrete challenges of practicing stakeholder involvement in a context of innovation in digital health.

In **Chapter 7**, titled "Including societal actors in R&D – Different expectations, different responsibilities", Harald Throne-Holst looks at the different roles of societal actors, citizens, and stakeholders in the processes of research and innovation. Building on a history of inclusion and participation and how these came to be demanded at the (Norwegian) policy level, the chapter discusses the opportunities, drivers, and barriers to inclusion in the unique Norwegian context, which has a long-held culture of supporting inclusion and social equality. The author specifically discusses the fragility of inclusion in research and innovation processes, seeing inclusion as a set of methods for participatory initiatives, and discussing their – sometimes tacit or hidden – purposes. Furthermore, the mechanisms to evaluate them are subject to conflicting views on how 'best' to enact these methods in the attempt to navigate different roles, relations, power, values, knowledge claims, and ways of knowing. Chapter 7 finally outlines four points that should receive particular attention when engaging in inclusion and participatory processes: (i) managing expectations, (ii) clarifying the purpose, (iii) designing the process carefully and collaboratively, and (iv) taking stock of where accountability is situated.

In **Chapter 8**, titled "Do You Value Responsible Innovation?", Elin Oftedal, Tatiana Iakovleva, and Matthias Kaiser take a closer look at user participation in the context of the development of digital health – or e-health – technologies for the care of elderly people in Norway. Drawing on qualitative research methods, such as interviews, 'user cafés', and participatory workshops, the authors try to map the complex value landscapes of stakeholders involved in research and innovation around e-health technologies, including end-users – specifically elderly people relying on healthcare technologies, industry representatives, and local policy-makers. Their analysis reveals the complexities and fragilities of stakeholder engagement around e-health, characterized by strong power imbalances

and often conflicting aims between well-being and profit. The authors highlight the importance of understanding the value landscape in shaping communication and collaboration within the innovation processes. They propose that responsible innovation (RI) should incorporate a thorough understanding of the cognitive, affective, and conative dimensions of the different stakeholder values. By integrating these insights into innovation processes, the authors argue for the possibility to support more inclusive, equitable, and aligned innovations within e-health.

Chapters 9 and 10 bring us back to CSR and are both focused on the actual impact of pieces of Norwegian legislation. As with the previous pair of chapters, here, too, we start with a contribution that looks at the systemic or operational level, and we then move to the shop floor level of implementation in particular companies. In **Chapter 9**, titled "Have law and social science trivialised the Concept and practise of whistleblowing in Norway 2007–2023?", Kristian Alm and Heidi Karlsen explore the consequences of the 2007 Norwegian legislation that protects whistleblowing and of the social scientific research on whistleblowing. They claim that the social value of whistleblowing lies essentially in its function of exposing malpractices within organizations that are detrimental to the public interest and hence in its enabling corrective action. However, their analysis reveals that, in Norway, both legislation and social scientific research have worked with and consolidated a broader understanding of whistleblowing that includes expression of internal and personal issues. This evolution – trivialization, in the author's own words – of whistleblowing has generated fragility at different levels. First, it has limited the specificity of social scientific knowledge. In the face of a recorded increase in retaliation against whistleblowers, we are currently unable to see its relative impact on its public interest use as against its personal interest use. But this fragility leads to another one, namely the lack of knowledge for informing action to support the public interest function of whistleblowing. So, what they point out is a paradox: two key instruments of social action (the law and social science) have actually brought about a conceptual vagueness that hampers social action.

In **Chapter 10**, titled "Acting on the Norwegian Transparency act: interpretation and implementation", Caroline Ditlev-Simonsen looks at a much younger piece of Norwegian legislation, the 2022 Transparency Act that compels medium and large businesses to publish information about how they deal with the risk of human rights and fair working conditions violations in their supply chain. Her focus is on the contingencies that affect the reception and compliance strategy within individual firms. Her study shows significant differences – steeped in the firm culture, resources, and personnel – in responding to the demands of the law and complying with it, thus indicating that hard law – as compared to soft forms of regulation – may not be an effective instrument to promote more homogeneous and consistent practices of CSR. Here the problem is not the lack of positive impact of the law – firms seem to have made improvements in overseeing their value chain – but the variability in their responses. So, here again, the contingent circumstances make the law a more fragile instrument than it is often assumed for bringing about conformity and standardized behaviour.

Throughout the book, the contributors have explored very different issues, ranging from social demands for transparency and participation to the resistance of vested interests, from funding strategies to individual values, from coordination problems to systemic constraints, from tensions within researchers' purposes to global rivalries, from stakeholder involvement to the impact of legislation. They have used an equally varied array of methods and approaches. This diversity reflects the complicated life-cycle of governance and policy concepts like RRI and CSR, as well as the more specific concepts through which they are translated and implemented. They move across different levels and reach many and diverse contexts, and at every move, fragilities, risks, contingencies, unknowns, and resistance occur. It is difficult not only to manage the journey but also to capture accurately and to evaluate its actual impacts. So, neither intentions nor knowledge pass across levels – ideation, policy, translation – without distortions and blurring. This happens in Norway too. The important point is that the variety and variability of the circumstances in which policy concepts are developed, translated, and adapted do not lend themselves to produce law-like regularities. This, of course, happens in many domains of human life, but when phenomena are observed, recorded, and reflected upon, a repertoire of organized experience and learning can be built and contribute to navigating new situations and contexts of fragilities. This is exactly what this book attempts to do.

References

Research Council of Norway (RCN) (2019). A Framework for Responsible Innovation – under BIOTEK2021, IKTPLUSS, NANO2021 and SAMANSVAR

Research Council of Norway (RCN) (2023). Responsible research and innovation as a method. https://www.forskningsradet.no/en/research-policy-strategy/rri/ (Accessed May 2024)

Arnold, E., Aström, T., Andréasson, H., Nielsen, K., Wain, M., Tofteng, M., & Røtnes, R. (2019). Raising the Ambition Level in Norwegian Innovation Policy. Final Report: technopolis group.

Gulbrandsen, E., & Rynning, H. (2016). Ansvarlig forskning og innovasjon i Norges forskningsråd. Forskningspolitikk. https://www.fpol.no/ansvarlig-forskning-og-innovasjon-i-norges-forskningsrad/

Jasanoff, S., & Kim, S. H. (2015). Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power. University of Chicago Press.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a Framework for Responsible Innovation. Research Policy 42 (9): 1568–80. https://doi.org/10.1016/j.respol.2013.05.008.

Völker, T., Slaattelid, R., & Strand, R. (2024). Translations of Responsibility: Innovation Governance in Three European Regions. Routledge, NY and London.

Section 1: **The emergence of responsibility**

Siri Granum Carson

Chapter 1
The institutionalization of social responsibility in Norwegian business and research: moral progress, moral decay, or both?

In this chapter, I argue that Norwegian organizations face challenges towards their social legitimacy due to an increasing complexity of the 'social contract' between the state, civil society, and organizations. Further, I argue that Corporate Social Responsibility (CSR) and Responsible Research and Innovation (RRI) may be viewed as organizations' attempts to manage these challenges by explicitly addressing issues concerning their social and environmental impact. In this context, both CSR and RRI may be understood as 'legitimizing practices', i.e., practices aiming to *gain*, *sustain,* or *regain* the social legitimacy of business and/or research organizations (Carson, 2019). These practices often take the form of engaging a broader set of stakeholders in order to secure attention to public needs and interests. In the last part of this chapter, I discuss whether the institutionalization of social responsibility should be viewed as a form of moral progress towards more mature organizations or if – conversely – it may be viewed as an insufficient response to a looming accountability crisis of these organizations. The overarching theoretical framework of the chapter is (neo-)institutional theory, but with a normative core of integrative social contracts theory (Donaldson and Dunfee, 1999).

1 Social responsibility and the social contract theory

The starting point of this chapter is the following question: What drives the institutionalization of social responsibility in, respectively, the business and research and innovation (R&I) context in Norway? A further question to be explored is whether there are significant similarities, differences, and overlaps between the context of business and the context of R&I. The final question of this chapter is whether this institutionalization process constitutes a strengthening or weakening of the organizations' actual abilities to be socially responsible.

The two key terms of the chapter are corporate social responsibility (CSR) and responsible research and innovation (RRI). These are both *essentially contested concepts*, in the sense that they are evaluative, complex, and open-ended or vague (Gallie, 1956; Okoye, 2009). While it will lead too far to discuss the different definitions and frameworks that have been proposed for these two concepts, the following chapters

(especially Chapters 3-5) provide some discussions of the definitions of RRI. As a starting point in this chapter, I offer my own, quite open-ended definition building on a previous book chapter on CSR as legitimacy management (Carson, 2020). I argue that CSR and RRI may be defined as the managing of social legitimacy in, respectively, the business context and the research and innovation context. More specifically, they may be viewed as legitimizing practices, addressing legitimacy challenges in the light of a renegotiation of the social contract of organizations.

This definition naturally brings up the need to define a further concept, namely legitimacy, or more specifically, in this context, social legitimacy or organizational legitimacy, referring to the acceptance of organizations in a society. A famous definition of social legitimacy is:

> [A] generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions (Suchman, 1995, p. 574).

The social legitimacy of an organization is often – and especially in certain sectors such as petroleum and mining – referred to as having a 'social license to operate' (Moffat et al., 2016). Another way to express this basic idea is that the organizations enter into a kind of social contract that implies a coherence between the social values expressed or supported by their activities and the commonly accepted values of the society they are part of (Cho and Patten, 2007). In the case of incoherence, the contract is compromised, and the company's legitimacy is threatened, which in turn may make it difficult or impossible for the organization to reach its goals.

At the basis of social contract theories in political philosophy lies the idea that societies are based on a (usually implicit) agreement that living well together means surrendering parts of our individual freedom to a state authority. While social contract theories go back in time as far as Hellenism and have since developed in a number of directions, the approach in this chapter refers to one particular version developed within business ethics. The approach is called 'integrative social contract theory' (ISCT) (Donaldson and Dunfee, 1995; Donaldson and Dunfee, 1999) and is set up with the intention to integrate empirical and normative aspects of management theory. What emerges is a *weakly normative theory* aiming to give practical guidance and direction to business managers on how to navigate between universal ethical norms, context-dependent social expectations, and harsh economic realities.

According to Donaldson and Dunfee, there are basically two kinds of social contract that should be considered in business:

– The first kind is normative and hypothetical, a kind of "macro" social contract that pertains to all participants in economic activities. It consists of "hypernorms", which are fundamental rules that take precedence over other contracts and are based in religious, political, and philosophical convictions.
– The second kind is actual, "micro" contracts that exist between members of economic communities. These may be explicit or implicit and are binding for mem-

bers of a particular community, be it an enterprise, an industry, or an economic system, to the extent that they are compatible with the hypernorms.

What Donaldson and Dunfee seek to build with their ISCT is a theory that goes beyond traditional business ethics by taking the socio-cultural context of business into account. They point out that, while families, or villages, are formed *naturally,* businesses have an *artificial* character, in the sense that they are created with a specific purpose, and therefore also with rules and structures that are arbitrary and vary significantly between cultures, industries and single companies. The ISCT is designed to provide a theoretical basis for allowing for such legitimate differences while at the same time recognizing universal norms and principles. This also allows for changes over time:

> As social contracts change, so too do the challenges for business [. . .] In subtle, far-reaching shifts, managers and members of the general public have gradually redefined their view of the underlying responsibilities of large corporations (Donaldson and Dunfee, 2002, p. 1855).

ISCT is designed specifically for the business context, and the integrative approach aims to capture the dynamic nature of private companies and their obligations to society. Another area where contractarian approaches have been promoted in recent years is the context of research and innovation. The concept of RRI – responsible research and innovation – emerges in the European context as a result of a perceived legitimacy crisis of science (von Schomberg 2013; Owen et al., 2021). Already towards the end of the 20th century, leading scientists evoked the social contract concept as a warning about broad changes demanding attention from the scientific community:

> Urgent and unprecedented environmental and social changes challenge scientists to define a new social contract (Lubchenko, 1998, p. 491).

Similarly, American social scientist David Guston argued for "a new social contract for science", necessitated by a lack of trust between science and society (Guston, 1994; Guston, 2000). The old or original contract pointed towards a clear division of labour, where the state would fund research and otherwise secure the academic freedom needed for science to flourish, while scientists in return would deliver the knowledge needed to fuel technological development and economic growth. Whether driven by a breach of public trust or accountability (due to a failure to deliver significant returns to society, perceived risks or negative effects, or suspicion of scientific misconduct) or a breach of trust on the side of the scientific community (for example, due to increasing state control and/or decline in direct funding), a renegotiation of the contract becomes necessary, detailing a new and more collaborative relationship between the state and the scientific community.

2 Social responsibility in the business context

The concept of CSR is a term that, in its different varieties, seeks to capture what private companies owe to society, over and beyond making a profit. But what exactly is the social responsibility of a business organization? A classic definition is given by Archie B. Carroll, claiming that in order to be socially responsible, a business must meet economic, legal, ethical, and philanthropic expectations given by society at a given point in time (Carroll, 1991). Carroll illustrates his point with his so-called CSR pyramid (see Figure 1), where the social responsibilities are shown in the form of layers based on how fundamental each element is for the organization. The bottom layer is the economic responsibility, since no business can survive without paying close attention to the bottom line, followed by the layer of legal commitments, which are similarly necessary to respect in order to stay in business. The next layer of ethical responsibilities, by which Carroll means observing social values, principles, or rules that are not codified into law, must also be respected for a company to be tolerated by society. In other words, all three bottom layers consist of mandatory commitments. Only the top of the pyramid, the philanthropic responsibility, is discretional and may be fulfilled to the extent the company seeks to increase its connection with or legitimacy within the society.

PHILANTHROPIC
Responsibilities
Be a good corporate citizen.
Contribute resources
to the community;
improve quality of life.

ETHICAL
Responsibilities
Be ethical.
Obligation to do what is right, just,
and fair. Avoid harm.

LEGAL
Responsibilities
Obey the law.
Law is society's codification of right and wrong.
Play by the rules of the game.

ECONOMIC
Responsibilities
Be profitable.
The foundation upon which all others rest.

Figure 1: The CSR pyramid (Source: Carroll, 1991).

The strength of Carroll's approach is that it shows how the social responsibilities of a company, rather than being set in stone like the Ten Commandments, are dynamic and context sensitive.

If we look at how CSR is presented in EU policy documents, we see that it has evolved from emphasizing only Carroll's top layer of philanthropy towards a more complex definition encompassing all four layers. In a report from 2001, CSR is defined as:

> a concept whereby companies integrate social and environmental concerns in their business operations and in their interaction with their stakeholders on a voluntary basis (EC, 2001, p. 7).

Ten years later, the new definition offered by the European Commission is considerably broader:

> CSR refers to the responsibility of enterprises for their impact on society (EC, 2011, p. 6).

In more detail, the report lays out the elements of CSR as the integration of social, environmental, ethical, human rights, and consumer concerns into the core strategy of the companies, with the aim to 1) maximize value, not only for their owners/shareholders, but also for their stakeholders and society at large, and 2) minimize adverse impact of their business operations.

The changes in the conceptualization of CSR point towards broader changes in what we may call the social contract of organizations, drawing on constitutional theories of the firm and specifically on integrative social contract theory (Donaldson and Dunfee, 1995). The point of departure for this analysis is the claim that business-society relations are increasingly complex, and in response to this complexity, companies address issues concerning their social and environmental impact in order to manage their social legitimacy (Carson, 2019). Globalization of the economy is a direct driver of this increasing complexity, leading to a governance gap in the sense that the regulatory authority of the nation-state loses some of its grip on the organizations. While opening up for new business opportunities, this also constitutes a threat to the social legitimacy of the organizations:

> [O]rganizations as actors in the transnational realm face an increasing 'legitimacy gap'. They make decisions whose consequences transcend any particular time or place – and thereby the regulatory apparatus of the state (Beck and Holzer, 2007, p. 4).

Ulrich Beck argued that economic globalization is a key factor of this complexity, where corporations increasingly meet expectations that they cannot satisfy (Beck, 2005). Several sociologists have pointed to complexity as a hallmark of (late) modernity (e.g. Morin, 2008; Beck, 1992; Luhmann, 1984). The emergence of global capitalism is widely recognized as a reason to renew our conception of democracy and citizenship (cf., e.g., Habermas, 1998; Beck, 1999; Keane, 2009). This entails that the traditional role of private business organizations, as monitored and controlled through governmental regulations,

and as responsible primarily for building private wealth and thereby supporting a general economic growth, is challenged. Palazzo and Scherer (2011) use the concept of 'political CSR' to describe a new situation where private companies – specifically multinational corporations operating in a global economy – must be recognized as co-governing institutions with a responsibility to promote social goods.

In this book, we are specifically exploring the Norwegian or, more generally, the Nordic context of responsible research, innovation, and business. In the literature, some attempts have been made to describe a 'Scandinavian' or 'Nordic' model of CSR, characterized by relatively flat hierarchies in decision-making and a consensual political culture, a strong social-democratic welfare state, and well-functioning partnerships between business, government, and labour organizations. Strand and Freeman (2009) coined the idea of a 'cooperative advantage' of Scandinavian companies, referring to a high level of trust, both within companies and between the companies and their stakeholders.

Another characteristic that has been pointed out as part of a Nordic or Norwegian form of CSR is that it is more government driven than the American, primarily business-driven, model (Albaredo et al., 2009; Gjølberg 2010). In the Norwegian context, this can be explained by the significant state ownership in large companies which, around the millennium, increasingly established activity outside of Norway. This led the Norwegian government, encouraged by civil society actors, to initiate campaigns and committees to promote CSR as a means to close the perceived governance gap in the global economy in relation to topics like corruption and human rights. In the domestic context, the CSR agenda was initially perceived as superfluous, given the high level of regulation in areas like workers' rights and environmental reporting.

From around the Millennium shift, CSR increasingly became relevant as a strategic tool also within the domestic context in Norway, as a response to public expectations and aiming to enhance public trust and corporate reputation. This was partly a response to an international trend where companies in general became more expressive about social and environmental values. It was also a response to the changing situation in Norway, where leading companies such as Hydro, Statoil, and Telenor and smaller enterprises producing clothes or other goods moved (parts of) their operations abroad. This created a need to 're-legitimize' their business operations, e.g., from national suppliers of economic stability and employment to socially and environmentally responsible actors in an under-regulated global economy (Carson et al., 2015). In parallel, many of the topics brought up under the CSR agenda underwent a transformation from voluntary engagement via soft law to novel forms of legal regulation. In 2013, a change in the Norwegian accounting law required that companies over a certain size should report annually, not just on their economic and environmental results but also on their social responsibility.

A further example of this institutionalization of corporate responsibility is detailed in chapter 10 of this book, on the development of the Norwegian Transparency Act, which was implemented in July 2022. The objective of the law is to promote cor-

porate respect and responsibility for basic human rights and decent working conditions in relation to the production of goods and services, by demanding public access to information about how companies handle these issues.

While corporate social responsibility is frequently mentioned in the context of these laws and regulations, the term *corporate sustainability* is today more widely used than CSR, both by industry actors themselves, by policy-makers, and by the research community. Already in 1999, John Elkington launched the concept of the triple bottom line of business, indicating that companies must balance their economic interests with their impact on society and the environment in order to be sustainable. By the time the Sustainable Development Goals were adopted by the UN in 2016, the concept of sustainability was thoroughly absorbed into the business context, pointing to a recognition of the considerable social and environmental footprints of private companies and, consequently, their shared responsibility for sustainable development. Arguably, though, the concepts of *responsibility* and *sustainability* are complementary rather than synonymous in the sense that the former points towards certain (whether legal, political, or moral) obligations to contribute to the latter.

3 Social responsibility in the research and innovation context

How do we secure that research and innovation, rather than posing risks and causing harm, promote social and environmental sustainability? Responsible (Research and) Innovation (RI/RRI) is, like CSR, a contested and diversely defined and used concept. It is often invoked in relation to governance of emerging technologies. According to René von Schomberg, a fundamental idea behind launching RRI as a cross-cutting issue in the Horizon 2020 framework program was that technology has become "democratized in its use and privatized in its production" (von Schomberg 2013, p. 4). On the positive side, this development promotes the accessibility of technological advances to a broader layer of people, exemplified by how the use of mobile phones has spread in developing countries. The negative side is that the privatization of production leads to a loss of control, exemplified by the futile attempts to control the development and spread of genetically modified organisms. RRI is launched as a broad framework by which to address both the positive and negative side of technological development, promoting the need to engage stakeholders throughout the whole innovation process in order to secure a sustainable outcome.

In many ways, RRI as a research and practice field developed organically from ELSI and ELSA studies – while in other ways it may be seen to develop in opposition to this regime. ELSI signifies *ethical, legal, and social implications* and the original acronym was developed under the auspices of the Human Genome Project in the late 1990s. It was gradually replaced, at least in the European context, by the slightly

broader term ELSA – ethics, legal, and social *aspects* of new technologies. Around 2010, the term RRI gained visibility, specifically as a policy concept in the context of the 8th EU framework program, Horizon 2020. While the thematic relation to ELSA was clear, in the sense of ensuring that technological innovation is aligned with societal goals, with RRI, the focus shifted more towards the entrepreneurial and industrial aspects of research and innovation and was tailored more towards modern, collaborative research where strong economic interests and competition in a global market in itself constitutes a major ethical concern (von Schomberg, 2013; Rip, 2018; Forsberg et al., 2021).

According to the OECD, RRI as a policy direction is a response to Europe's perceived societal challenges and crisis in public trust, which implies that ethical and societal dimensions should get a more prominent role in research and innovation frameworks (OECD, 2016). RRI should, according to the European Commission, constitute a "paradigm shift in how we think, live and interact together, as well as a paradigm shift in what the role and place of science should be" (EC, 2009).

The RRI agenda was driven by a perceived social legitimacy gap resulting from the so-called European paradox presented by the European Commission in their green paper on innovation from 1995. The paradox consisted in the fact that, while Europe has invested heavily in publicly funded science, the desired innovation outcome and following economic growth resulting from these investments have been less than satisfactory (de Saille, 2015; cf. EC, 1995). According to de Saille (2015), the RRI agenda was launched with the double objective of defending the considerable use of public funding in research and innovation to secure a willingness to support this priority in the future, as well as an attempt to dissolve the paradox by improving the uptake of the new technology in society. Hence, at the root of the European RRI agenda lies the desire to promote and align two (potentially conflicting) aspirations: 1) to boost economic growth through R&I, and 2) to secure that R&I moves society in the right direction by making it more inclusive and responsive. This double agenda may be seen in parallel with the 'creating shared value' strategy, launched as a version of (or rather alternative to) CSR by Porter and Kramer (2006) and defined as enhancing the competitiveness of companies by promoting social goods in the society in which they operate.[1]

The CSR and RRI agendas share the key objective to engage with the broader community to secure a just and sustainable outcome. In Norway, both the CSR and the RRI agenda were met with open arms and promoted through several policy initiatives, specifically also by the Research Council of Norway (RCN) through the SAMANSVAR ("co-responsibility") program (2015-2020).[2] An interesting aspect of the SAMANSVAR pro-

---

1 Albeit the Creating Shared Value strategy is primarily targeted at individual companies, while the European RRI agenda is directed at the R & I system as a whole.

2 See chapter 3 for an in-depth analysis of the RCN initiatives for promoting RRI.

gram was that it featured the explicit ambition of bringing together the perspectives of CSR and RRI within Norwegian research and innovation policy, cf. the revised program plan:

> The main objective of [SAMANSVAR] is to contribute to meeting the global societal challenges through responsible technology development and socially responsible business [. . .] There will be synergy effects from seeing responsible innovation and corporate social responsibility in relation to each other (RCN 2018, p. 4-6, my translation).

This is admittedly not a unique Norwegian phenomenon but mirrors a development in Europe where the term responsible innovation increasingly is taken up in the business context (cf. Lubberink et al., 2017), featuring overlapping topics from the CSR agenda such as stakeholder engagement, but focusing on social responsibility in the development of new and emerging technologies.

4 RRI, stakeholder engagement, and new governance

The RRI ideal of early involvement of a broad range of stakeholders is inherited from previous frameworks such as ELSA and TA (technology assessment). The same ideal is from early on at the core of the CSR agenda (Freeman, 1984). In both cases, the engagement of stakeholders promotes social legitimacy by securing attention to public needs and interests. Further, we have noticed a parallel reference to a perceived need to redefine the social contract of, respectively, business and science, due to an increasing complexity which emerges in line with globalization.

This point can be related to the concept of "new governance" – or simply "governance" – pointing towards new forms of organizing society. (New) governance signifies non-hierarchical, network-based forms of governing involving a broad range of stakeholders – cf. the classic definition offered by Rhodes (1996, p. 652: "Governance refers to 'self-organizing, interorganizational networks'". Governance signals a movement from top-down government to more reciprocal structures, a movement that is reflected across policy agendas in Western societies from the late 1990s and is seen to be caused by an increasing complexity of contemporary social conflicts (Giddens, 1991)).

New governance is fuelled by globalization, necessitating new ways of thinking about international relations at a point where state authority loses some of its grip and must share in governing with diverse social actors such as private business organizations, research institutions, and NGOs (Bevir, 2012). Røiseland and Vabo (2016) point out that one central characteristic of governance is that discourse becomes a central form of a governing scheme which is less hierarchical than both state authority over all sectors and negotiations within a "free" market structure. Governance by discourse typically includes stakeholder engagement and other forms of soft regula-

tion, which we have identified as integral elements of both the CSR and the RRI agenda. This is connected to the overall social-contract-theoretical framework of this chapter, pointing towards changes in the ways societal actors hold each other accountable. Different versions of social contract theories in business have, for example, been evoked to lend support to stakeholder theory (cf. Francés Gomez, 2018).[3]

While the thematic and methodological scope of both ELSA and RRI is broad, the long-standing traditions of stakeholder engagement as a method by which to secure responsible governance are especially strong in the fields of emerging technologies, not least in biotechnology. We find a striking example of an early institutionalization of this idea in the Norwegian Gene Technology Act from 1993. The rapid development of gene technology is expected to have a significant societal impact, over and beyond, posing concrete solutions as well as risks to the environment and human health. In most countries, including in the EU, regulation of gene technology is limited to avoiding detrimental effects on health and the environment. However, the Norwegian act goes beyond risk management and stresses that the deliberate release of GMOs should represent a "benefit to the community" and enable "sustainable development" (cf. Sections 1 and 10 of the Act). The act implements a mandatory approval process which includes so-called non-safety concerns– including questions of ethics, societal utility, sustainability, and cultural heritage.

While risk assessments usually come in the form of quantitative threshold assessments where a product is either sufficiently safe or not, non-safety concerns are gradual and qualitative and involve the weighing of costs and benefits which may be unequally distributed between different stakeholders (Forsberg et al., 2019). This has led some to argue that these assessments by necessity are vague, arbitrary, and subjective (Zetterberg and Björnberg, 2017). However, others have argued that these shortcomings may be addressed through inclusive, deliberative processes where all relevant arguments are heard before a decision is made by competent and trusted people (Myskja and Myhr, 2020). In other words, the governance of such technologies should incorporate public engagement, both in regulatory discussions and in concrete technology development projects, in line with the general principles of RRI (Stilgoe et al., 2013).

---

3 Albeit, as critics have pointed out, the integrative social contracts theory of Donaldson and Dunfee lacks a deliberative understanding of social contract, leading to a "mono-logical" rather than a dialogical view on stakeholder involvement in business (Palazzo and Scherer 2006; Frances-Gomez 2018). This mirrors a distinction in the stakeholder management literature, whereby stakeholders may be involved at different levels spanning from keeping them informed, via consulting them, to a true dialogue where the intention is to exchange arguments in order to reach an agreement (Morsing and Schultz, 2006).

5 Institutionalizing social responsibility – a story of moral progress or decay?

A starting point in this chapter is that the social responsibility of both business and science may be specified in accordance with social contracts theory. The social contracts may be seen to consist in a relatively permanent "macro" contract, as well as more dynamic, implicit, and explicit "micro" contracts. Broad societal changes may necessitate renegotiation of these social contracts, at least to the extent that the social responsibility of business and research organisations are made explicit in new ways rather than being non-articulate and taken for granted (Matten and Moon, 2008).

The social contracts may be further qualified and explicated in terms of social legitimacy, borrowing from institutional theory the idea that social legitimacy is a measure of how organizations successfully respond to the expectations of their surroundings. CSR and RRI may be explicated through policies, actions, or other measures implemented in the business and/or research community with the objective to gain, maintain, or regain social legitimacy. A follow-up question then becomes: Is this, overall, good news? May we stipulate that the institutionalization of social responsibility constitutes a form of moral maturing, by which the societal actors clarify their roles and obligations towards each other, as exemplified by von Schomberg's definition of RRI as "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other" (Schomberg, 2012, p. 9)?

On the other hand, the focus on social responsibility may be viewed as a response to an accountability crisis of respectively business and research, where accountability in this context should be understood as an aspect of social legitimacy (cf. e.g. Rendtorff, 2020). As a comment to the mounting environmental crisis, Ulrich Bech introduced the concept of 'organized irresponsibility' (Beck, 1992), pointing towards a system of designing elaborate systems of ambiguous responsibility in such a way that no one in particular may be held accountable (cf. also Giddens, 1999). The call for new social contracts, respectively, between business and society, and between science and society, may be seen as a recognition of this challenge – in both cases aiming for a more clarified distribution of accountability for social and environmental challenges.

A genuine worry in this context is, however, the issue of window dressing or greenwashing (or, rather, 'SDG washing'), where the public is misled to believe that organizations are making actual moves towards more sustainable practices when, in reality, they do not make any substantial changes. In institutional theory, the term isomorphism is used to describe the institutional pressures that bring organizations to adopt new structures – such as CSR and RRI (DiMaggio and Powell, 1983). A related term is decoupling, signifying that the internal practices of the organizations may become disconnected from their external presentation (Meyer and Rowan, 1977). Decoupling may allow organizations to carry on with business as usual while publishing

nice CSR reports or allow research communities to continue their practices while bringing in a social scientist (or even hiring a consultant) to do 'the RRI bit'.

The phenomena of CSR and RRI may be seen to represent appropriate responses to a complexity that weakens both the nation-states' capacity to control organizations, as well as individuals' capacities to make ethical decisions within these organizations. In this scenario, the collective responsibility of these societal structures to a certain extent takes the place of traditional professional ethics (whether of researchers, entrepreneurs, or farmers). Hence, the recognition of a certain (inevitable) moral decay, or rather incapacity, in the face of complexity is met by ascribing moral imperatives (or at least social responsibility) to organizations who are required to act as responsible 'corporate citizens' (Matten and Crane, 2005). However, a well-functioning institutionalization of organizational responsibility is inconceivable without a certain level of transparency and structural accountability. The institutions by which to secure this are still to be developed.

References

Albareda, L., Lozano, J. M., Tencati, A., Perrini, F., & Midttun, A. (2009). The role of government in corporate social responsibility. Ethical Prospects: Economy, Society, and Environment, 103–149.

Beck, U. (1992). Risk society: Towards a new modernity (Vol. 17). sage.

Beck, U. (2005). Power in the global age: A new global political economy. Polity.

Beck, U. & Holzer, B. 2007. Organizations. International handbook of organizational crisis management, 1.

Bevir, M. 2012. Governance: A very short introduction. OUP Oxford.

Carroll, A. B. 1991. The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders. Business horizons, 34(4), 39–48.

Carson, S. G., Hagen, Ø., & Sethi, S. P. 2015. From implicit to explicit CSR in a Scandinavian context: The cases of HÅG and Hydro. Journal of Business Ethics, 127, 17–31.

Carson, S. G. 2019. The corporate legitimacy matrix–A framework to analyze complex business-society relations. Philosophy of Management, 18(2), 169–187.

Carson, S. G. 2020. Corporate Social Responsibility (CSR) as Social Legitimacy Management. In: Rendtorff, J.D. (ed.) Handbook of Business Legitimacy: Responsibility, Ethics and Society (pp. 897–906). Cham: Springer International Publishing.

Cho, C. H., & Patten, D. M. 2007. The role of environmental disclosures as tools of legitimacy: A research note. Accounting, organizations and society, 32(7–8), 639–647.

DiMaggio, P. J., & Powell, W. W. 1983. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American sociological review, 48(2), 147–160.

Donaldson, T. and T. W. Dunfee. 1995. Integrative Social Contracts Theory: A Communitarian Conception of Economic Ethics, Economics and Philosophy11(1), 85–112.

Donaldson, T. and T. W. Dunfee. 1999. Ties That Bind: A Social Contracts Approach to Business Ethics. Harvard Business School Press, Cambridge.

Donaldson, Thomas, and Thomas W. Dunfee. 2002. Ties that bind in business ethics: Social contracts and why they matter. Journal of Banking & Finance 26: 1853–1865.

Elkington, John. 1997. Cannibals with forks. In The triple bottom line of 21st century. Capston, Publishing Ltd.

European Commission. 1996. Green Paper on innovation. Document drawn up on the basis of COM(95) 688 final, Bulletin of the European Union Supplement 5/95.

EC. (2001). Green Paper: Promoting a European Framework for Corporate Social Responsibility. Directorate-General for Employment, Social Affairs and Inclusion (European Commission).

EC. (2011). A renewed EU strategy 2011–14 for Corporate Social Responsibility. European Commission.

Forsberg, E. M., Hofmann, B., Kaiser, M., Myskja, B. K., Strand, R., & Ursin, L. 2019. Vurderingskriteriet etikk: Veileder for operasjonalisering av vurderingskriteriet etikk i genteknologiloven.

Forsberg, E. M., Strøm-Andersen, N., Bugge, M. M., Brekke, A., & Prestrud, K. 2021. RRI and Value Chain Collaboration for Innovations Reducing Food Waste.

Francés-Gómez, P. 2018. Social contract theories. In The Routledge companion to business ethics (pp. 96–112). Routledge.

Francés-Gómez, P. 2020. Social Contract Theory and Business Legitimacy. In: Rendtorff, J.D. (ed.) Handbook of Business Legitimacy: Responsibility, Ethics and Society (pp. ??). Cham: Springer International Publishing.

Freeman, R.E. 1984. Strategic management: A stakeholder approach. Boston: Pitman.

Gallie, W. B. 1956. ‹Essentially Contested Concepts', *Proceedings of the Aristotelian Society* 56, 167–198 reprinted in M. Black (ed.): 1962, *The Importance of Language* (Prentice-Hall Inc., Englewood Cliffs, NJ), pp. 121–146 (pages cited in the paper refer to the reprint edition).

Giddens, A.1999. Risk and responsibility. *Mod. L. Rev.*, *62*, 1.

Guston, D. H. 1994. The demise of the social contract for science: Misconduct in science and the nonmodern world. The Centennial Review, 38(2), 215–248.

Guston, D. H. (2000). Between politics and science: Assuring the integrity and productivity of research. Cambridge University Press.

Habermas, J. 1998. Die postnationale Konstellation und die Zukunft der Demokratie. Blätter für deutsche und internationale Politik, 7(98), 804–817.

Keane, J. 2009. The life and death of democracy. Simon and Schuster.

Lubchenco, J. 1998. Entering the century of the environment: a new social contract for science. Science, 279 (5350), 491–497.

Lubberink, R., Blok, V., Van Ophem, J., & Omta, O. 2017. Lessons for responsible innovation in the business context: A systematic literature review of responsible, social and sustainable innovation practices. Sustainability, 9(5), 721.

Luhmann, N. 1984. Soziale systeme (Vol. 478). Frankfurt am Main: Suhrkamp.

Matten, D. and A. Crane. 2005. Corporate citizenship: Toward an extended theoretical conceptualization. Academy of Management Review 30(1),166–179.

Matten, D. and J. Moon. 2008. 'Implicit' and 'explicit' CSR: A conceptual framework for a comparative understanding of corporate social responsibility. Academy of Management Review, 33(2), 404–424.

Meyer, J. W., & Rowan, B. 1977. Institutionalized organizations: Formal structure as myth and ceremony. American journal of sociology, 83(2), 340–363.

Morsing, M., & Schultz, M. 2006. Corporate social responsibility communication: stakeholder information, response and involvement strategies. Business ethics: A European review, 15(4), 323-338.

Moffat, K., Lacey, J., Zhang, A., & Leipold, S. 2016. The social licence to operate: a critical review. Forestry: An International Journal of Forest Research, 89(5), 477–488.

Morin, Edgar. 2008. On complexity. New York: Hampton Press.

Myskja, B. K., & Myhr, A. I. 2020. Non-safety assessments of genome-edited organisms: should they be included in regulation?. Science and Engineering Ethics, 26(5), 2601–2627.

OECD. 2016. OECD Science, Technology and Innovation Outlook 2016 Country Profile.

Okoye, A. Theorising Corporate Social Responsibility as an Essentially Contested Concept: Is a Definition Necessary? 2009. Journal of Business Ethics 89, 613–627.

Owen, R., von Schomberg, R. & Macnaghten, P. 2021. An unfinished journey? Reflections on a decade of responsible research and innovation, Journal of Responsible Innovation, 8:2, 217–233.

Palazzo G, Scherer A. 2006. Corporate legitimacy as deliberation: a communicative framework. J Bus Ethics 66(1): 71–88.

Porter, M. E. and M.R. Kramer (2006). Strategy & Society: The Link Between Competitive Advantage and Corporate Social Responsibility. Harvard Business Review 84 (December), 78–92.

Rendtorff, J. D. 2020. The concept of business legitimacy: Learnings from Suchman: Integrating sociological, ethical and critical perspectives. Handbook of business legitimacy: Responsibility, ethics and society, 3–30.

Rip, A. 2018. The past and future of RRI. Futures of science and technology in society, 115–133.

Rhodes, R. A. W. 1996. The new governance: governing without government. Political studies, 44(4), 652–667.

Røiseland, A., & Vabo, S. I. 2008. Governance på norsk. Samstyring som empirisk og analytisk fenomen. Norsk statsvitenskapelig tidsskrift, 24(1–2), 86–107.

Saille, S. de. 2015. "Innovating innovation policy: the emergence of 'Responsible Research and Innovation'" https://www.tandfonline.com/doi/epdf/10.1080/23299460.2015.1045280?needAc cess=true

Scherer, A. G., & Palazzo, G. 2011. The new political role of business in a globalized world: A review of a new perspective on CSR and its implications for the firm, governance, and democracy. Journal of management studies, 48(4), 899–931.

Stilgoe, J., Owen, R., & Macnaghten, P. 2020. Developing a framework for responsible innovation. In The Ethics of Nanotechnology, Geoengineering, and Clean Energy (pp. 347-359). Routledge.

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. Academy of management review, 20(3), 571–610.

von Schomberg, R. 2012. Prospects for technology assessment in a framework of responsible research and innovation. Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden, 39-61.

von Schomberg, R. 2013. A vision of responsible research and innovation. Responsible innovation: Managing the responsible emergence of science and innovation in society, 51–74.

Zetterberg, C. and *Björnberg*, K. E. *2017*. Time for a new EU regulatory framework for GM crops? Journal of Agricultural and Environmental Ethics 30(3): 325–347.

Matthias Kaiser
Chapter 2
From value-freedom to responsible research and innovation? Post-normal and transdisciplinary pathways

The following chapter is an account of the slow emergence of social responsibility in science, struggling to overcome the dogma of value-free science. It is very much a subjective account, based on my reading of the history and philosophy of science. While being aware that other readings of the historical developments are both possible, published and making up interesting reads, I would maintain that there is learning in relating to various such accounts. My account also tries to connect the development at some points to contributions from Norway, which I believe to be an interesting forerunner in debates about RRI and CSR, the topics in the other contributions to this book.

In the following sections I portray the concept of scientific responsibility – and indeed the ethics of science – as caught in the tensions between knowledge and power. Furthermore, I will discuss the ideas of post-normal science and the research challenges of transdisciplinary research as pathways to realizing the social responsibility of science.

1 Introduction

Science takes care of the facts, and the State takes care of the values. Right? Well, this was the dogma of Modernity which many of us were brought up with. It is also the ideology which is still propagated by many intellectual elites in our days. And it is at least implicit in the curricula of many, if not most, institutions for higher education. It is captured in the slogan of a value-free science. Nevertheless, it is a dogma that does not fit well the history of the emergence of knowledge and science as I will indicate in the following section.

In the current science policy discussions, we often find reference to the responsibility of science, in Europe typically designated as *Responsible Research and Innovation* (RRI). Even from an intuitive point, this seems to clash with the dogma of a value-free science. The term responsibility invokes ethics and ethics invokes values. Responsibility

towards whom?[1] With regard to what? And why? Why not just stick to the facts of the world? Obviously, some kind of sea-change might have occurred in our thinking of science and how we conceptualize the very activity of doing research. In the following sections I shall argue that we need to understand the historical development of our modern science in order to appreciate this policy change and its implications. In particular, we need to examine the tensions between scientific knowledge and its relation to existing power structures in society. We also need to critically assess the philosophy of science which led us here. And one needs to appreciate a new take on ethics which has moved from a theoretical and highly abstract enterprise to practical ethics which is prone to stress process versus outcomes.

2 Science and values in history

The first thing to note is that the very term "science" is a relatively recent one to signify our scholarly activities. If we had moved our attention from science to *knowledge*, we might have gained the insight that knowledge accompanies human cultures from the very early beginnings of humanity, of *Homo sapiens*. Place-based knowledge was always decisive in understanding and living within a given environment. Knowledge was often interwoven with mythology and magic. We call this now place-based, or sometimes indigenous, knowledge systems. The very differentiation between knowledge and belief is a result of the Hellenic philosophy of antiquity. It was characterized as the opposition between *episteme* versus *doxa*. Plato (in Meno and Theaetetus) promoted the striving for real knowledge, episteme, rather than the malleability and multitude of doxa (in Gorgias). Had one asked if knowledge in general was value-free either in the form of episteme or doxa, we would have tended towards a negative answer. Knowledge cultures were to be integrated into the positive development of society and enable a value-guided control of our environments.

The second thing to note is that as people began to travel to faraway places, knowledge also travelled to a certain extent. Part of this happened through early migration movements, and later through trade, political expansions, and seafaring colonialism. This was, for instance, true in relation to Islamic knowledge, as, for example, represented by Ibn al-Haytham (AD 965–1040), and with the discovery of the Americas also in relation to, for example, Aztek knowledge (cf. Poskett, 2022).

But later the so-called *Scientific Revolution* (ca. 1550–1750) took place in Central Europe. In a sense, this naming is already a whiggish viewpoint, perhaps an oxymoron even. Those great names which founded what we now can call the *Principles of*

---

1 There may have always been an intra-scientific responsibility concerning purely epistemic values, and related to the scientific community. Here I ask if there was or is a sense of responsibility transcending the scientific community.

Modern Science did not name their activity *science*. They talked about "natural philosophy" (*philosophia naturalis*). The first scientific journal, *The Philosophical Transactions* – still published today by the Royal Society – in its first volume reads like this on the title page: "*Philosophical Transactions Giving Some Accompt of the Present Undertakings, Studies, and Labours of the Ingenious in Many Considerable Parts of the World*". The scholars introduced their view of this new natural philosophy in an environment where scholarly knowledge was already considered well established in universities run mostly by the Church. Thus, they saw a need to differentiate their new kind of knowledge from the established one, in particular, in the foundations of their new academies. The Royal Society, arguably one of the most influential academies in their time, said in its charter:

> Studies are to be applied to further promoting by the authority of experiments the sciences of natural things and of useful arts (1662).

> In order to the propounding and making of Experiments for the Society, consideration shall be had of the importance of any Experiment, to the discovery of any truth or axiom in nature, or to the use and benefit of mankind (1663).

The prospective utility of their enterprise was stressed here. Their member Robert Hooke had proposed an even more explicit mandate in terms of utility:

> The business and design of the Royal Society (is) to improve the knowledge of natural things, and all useful Arts, Manufactures, Mechanick practices, Engynes, and Inventions by Experiments, – (not meddling with Divinity, Metaphysics, Moralls, Politicks, Grammar, Rhetorick, or Logick).

It is obvious how Hooke tried to differentiate the activities of the Royal Society from these other scholarly activities which were captured in the parentheses. Institutionally it was important to stay away from the spheres of political power (the king) and the moral power, namely Churches and their dominant focus on Aristotle's philosophy. Morals and politics were not the business of the protagonists of natural philosophy. Practical utility was, as also Francis Bacon stressed. Nowadays, we would perhaps describe this goal as innovation. So, were these principles of modern science value-free then?

The rhetoric was perhaps value-free but not the historical influence. One can argue that it was precisely in the areas of metaphysics, politics, and morals that the Scientific Revolution had its greatest impact in the following 200 years. The Age of Enlightenment adopted the philosophy of these thinkers as a major driver for the restructuring of society. It was now "reason" in the sense of the critical spirit of the scientific revolution that was supposed to be the foundation of the new nation-states. Progress which Fontenelle (the first live-long secretary of the French Academy) promoted as the characteristic of scientific knowledge, was now lifted to a general principle of historical and societal progress. Condorcet (the second live-long secretary of the French Academy) wrote one of the most optimistic treatises on the progress of humanity (*Sketch for a Historical Picture of the Progress of the Human Mind*, 1794/1795)

while in hiding from the French Revolution, of which he himself was an important spokesperson and representative. At the very end of this book, after sketching all the impressive progress that was already made and which he predicted would be made in the future, he asks what later (1798) was also discussed by Malthus, known as the Malthusian trap or the Malthusian catastrophe:

> Will increased welfare and improved health of man lead to largely increased populations? Will not necessarily there be a time when the number of people has outgrown the natural resources that nature can supply? Is it not reasonable to assume that when resources become scarce, then there will be fight for the resources, war between people?

Condorcet does not think so, and he presents two arguments against it. The first one I call the technology-fix-argument: "Nobody could claim that such a time is imminent, Technological progress may bring the answers". Let me insert a little sidekick to our current political environment here. The technology-fix-argument is often the default response from politicians to our global environmental challenges. Examples abound. For instance, when facing the challenge of climate change due to CO_2 emissions, some governments have readily invested in technology development to capture CO_2 and store it safely (like the Norwegian government with their "moon landing" project at Mongstad). Perhaps it is the material components of such a strategy which account for its attractiveness; workplaces are created, and the hoped-for innovations would create a bonus for global trade.

The second argument I call the ethics-argument:

> People's ethics and morality will progress alongside reason. Our moral duty is not to make sure that unborn life is born, but that those that are born are secured a life in reasonable welfare, dignity and happiness.

For Condorcet the progress of science (equated with reason) and the progress of morality were by necessity intertwined. He could not think of the one without the other. This belief in general progress gripped many thinkers. Arguably the high point of this belief in progress came in 1851 with *The Great Exhibition of the Works of Industry of All Nations* in London, symbolized by the *Crystal Palace*, a progressive symbiosis of steel and glass. The progress of science and technology was to promote a steady progress of society as well, even with a view to global development. Prince Albert expressed this vision of scientific progress this way:

> Science discovers these laws of power, motion, and transformation; industry applies them to the raw matter which the earth yields us in abundance . . .. The exhibition of 1851 is to give us a true test and a living picture of the point of development at which the whole of mankind has arrived at this task, and a new starting point from which all nations will be able to direct their further exertions" (cited in Bury, 1921, p.190).

Thus, science and technology paved the way to a unified humanity at peace with each other. As Bury observed:

> The great Exhibition was at the time optimistically regarded, not merely as a record of material achievements, but as a demonstration that humanity was at last well on its way to a better and happier state, through the falling of barriers and the resulting insight that the interests of all are closely interlocked (Bury, 1921, p. 190–91).

Albert tried to save the Exhibition from the charge of materialism by expressing that the visitors would feel a "deep thankfulness to the Almighty for the blessings which He has [already] bestowed upon us" (cited in Cantor, 2012, p. 441). I would claim that this was also the symbiosis of facts and values, a symbiosis of the progress of science with the progress of morality – at least in the world of beliefs, even religious beliefs, of many. Of course, the realities of colonialism were conveniently ignored and not questioned. In fact, the expansion of colonialism under the cover of serving humanity might have been among the main drivers for addressing scientific progress.

The 19th century saw now also the institutionalization of the system of science. What was initially an amateur activity turned now into a profession. A new phase of natural philosophy had started, "professional science",[2] offering its practitioners both income and social status. The rise of modern universities since 1810 (Humboldt University in Berlin) spread worldwide with the load of colonialism in the baggage. This initiated further developmental stages: "industrialized science" at the end of the 19th century (characterized by new industrial research bodies between universities and industry, such as the Kaiser-Wilhelm Institute in Germany) and later "Big Science" emerging during WWII (with external research objectives, large mixed consortia, and industry-like management; cf. the Manhattan project).

It was also around this time in the early 1800s that two other important developments emerged. First, the term "scientist" was now introduced, marking the transition from the amateur natural philosopher of the Scientific Revolution to the professional scientist. The philosopher and polymath William Whewell wrote in 1840: "We need very much a name to describe a cultivator of science in general. I should incline to call him a Scientist" (from William Whewell, Preface to *The Philosophy of the Inductive Sciences* 1840).

The second important development was the introduction of scientific disciplines as a sub-division of science which now was striving for objectivity. Lorraine Daston (1991) sees the quest for impartiality, the critical and independent assessment of knowledge claims, and eventually the claim of the objectivity of science, appearing during the 19th century. This sidelined all ethics and morality in the sciences. Scientific disciplines as we know them today emerged only in the 19th and 20th centuries (Stichweh, 1992), and they are still continuously producing countless sub-disciplines. Even socially important subjects like human health and medicine adopted the charac-

---

2 I have taken the differentiation into amateur science, professional science, industrialized science, and Big Science from lecture notes that Arie Rip sent me in the early 1980s. Unfortunately, I do not have a precise reference to them.

ter of modern science only relatively late (roughly at the turn of the 19th to 20th century). Social science was similarly not immediately ready to join this prestigious club (Mcintyre, 2019).

With the fin-de-siecle philosophers, the insight that societal progress was not an automatic result of the application of science and technology (1912: the sinking of the Titanic) signalled an end to the popular belief in progress. This was further assisted by the rise of populist dictatorships and the occurrence of two World Wars, which made it abundantly clear that the earlier symbiosis of science and morality was an illusion. When earlier visions of progress faltered, visions of a marriage between science and ethics faltered also. This was partly amplified by influential writers such as Max Weber, who proclaimed that the objectivity of science demanded a strict separation of science and values (Weber, 1917; Sharlin, 1974). However, the rhetoric of a value-free science continued undisturbed to maintain trustworthiness and funding, while the brittleness of this separation between objective science and subjective ethics became slowly more visibly under the surface. But philosophers (of science) followed this rhetoric willingly and radically.

The long tradition of studies of ethics was now deemed by them to belong wholly to the sphere of subjectivity where rationality had no say. With the arguments of Hume and Moore in their baggage, ethics could at best be the result of some spontaneous intuition or express a subjective attitude towards certain acts and events. Ayer stated with no beating around the bush:

> If now I generalise my previous statement and say, "Stealing money is wrong," I produce a sentence that has no factual meaning – that is, expresses no proposition that can be either true or false. . . . I am merely expressing certain moral sentiments (Ayer, *Language, Truth and Logic*, 1936/1970, 107).

With the possible truth or falsity of all scientific statements as the accepted standard for scientific propositions (both in logical empiricism of the Vienna Circle and Popper's falsificationism), ethics had now fallen out of the rationality of scientific discourses. At best, values belonged to the context of discovery, not the context of justification (Reichenbach, 1938). The ideology of a value-free science was in the driving seat.

3 Cracks in the ideology of value-free science

Science after WWII took a new turn towards Big Science. But no ideology can reign without the seeds of its destruction already growing in the underground. This was also the case with the ideology of value-free science. It is only logical that some of the first cracks appeared in the philosophical discussions about ethics, the field which seemed banned by positivism and the new analytical philosophy.

Some early attempts to regain rationality for ethics should be mentioned here, such as Hare (1952) and Toulmin (1950). From our present-day viewpoints, it might be quite interesting, if not even strange, to note how philosophers struggled to argue for reason and rationality in normative discourse and ethics, given, for instance, the important contributions in the 19th century from, for example, Immanuel Kant.

It is with a certain amount of pride that I will now briefly report on the important work of one of my early supervisors and the mentor in Norwegian ethics of science. I am talking about Knut Erik Tranøy (1915–2012). Tranøy came to know both Georg Henrik von Wright, Ludwig Wittgenstein, and C.D. Broad. von Wright elaborated in an article of 1951 on the parallelism between the so-called alethic modalities of necessity, possibility, impossibility, and the so-called deontic modalities of obligation, permission, and prohibition. While the alethic modalities seem safely placed within the sciences, the deontic modalities, despite their parallelism, belonged to ethics and norms and were thus placed outside the realm of the sciences, at least in mainstream philosophy of the time. Tranøy found this problematic. While he accepted the naturalistic fallacy of Hume and Moore, he also realized how scientific insights are directly relevant to the moral choices we make. But Kant's statement that *ought* needs to imply *can* was not easily reconcilable with everyday experiences. In Tranøy's thesis this turns into the following interpretation:

> It is not legitimate/morally right for a person *b* to issue an order for person *a* to do an act *p* unless it is possible for *a* to do *p*.

These insights, coupled with the work on deontic logic, provided a basis for a "rational" approach to normative systems. Tranøy's thesis (On the Logic of Normative Systems, 1953) was in many ways an attempt to build a bridge between science and ethics, the sciences characterized by rationality, and now the ethics, the basis of normative systems, characterized by the same rationality, even logic.[3]

Tranøy followed this path throughout his whole career. In later years he became the first in Norway to teach medical ethics at the University of Bergen, and he expanded his philosophy through papers and a book on the normative aspects of science. Science became, for him, a human activity as many other activities of life. Tranøy was eventually the most important mentor of the rise of an ethics of science in Norway, which led to Norway being the first country which established national research ethics committees for all fields of science and scholarship (1990).[4]

This conception of science as a social practice is also the important starting point for the analysis that Jerome Ravetz provides in his book *Scientific Knowledge and its Social Problems* (1971). Like Tranøy, Ravetz also had some mentors who inspired his

---

3 Tranøy followed here also the philosophy of Thomas Aquinas for whom *veritas* and *rectitudo*, truth and morality, were two sides of the same coin.

4 More on Tranøy's influence on an ethics of science in Kaiser (2003), though in Norwegian.

work substantially. Perhaps the most important one was Stephen Toulmin. Ravetz's book set a different tone in the then-dominant philosophy of science. Like Tranøy, he takes as a starting point a conception of science as a process which is institutionally organized and carried through by individuals who may or may not share the goals of the institution. He devotes a whole chapter to "Ethics in Scientific Activity," where he calls for a new ethic of science, since "unless there is an effective ethic, even more refined than a 'professional ethic', this very delicate and sensitive work will not long continue to be well governed or well performed" (Ravetz, 1971, 313). He warns that the increasing industrialization of science will eventually lead to various forms of corruption, inherent in the conflict between the autonomy of science and its industrialization. He claims there is the temptation to "extol the virtues of the free search for truth to one audience, and to promise useful services to another" (Ravetz, 1971,421). He adds: "the process of industrialization is irreversible; and the innocence of academic science cannot be regained" (Ravetz, 1971, 422–23). If science is not to fall entirely into the claws of corruption, a new science is to emerge, namely what he calls "critical science". Critical science will face problems that are more demanding "than in either pure science or technology" (Ravetz, 1971,429). He foreshadows the challenges of transdisciplinary research:

> The work in critical science involves an awareness of craft skills at all levels, and the conscious effort of mastering new skills. The data itself is obtained in a great variety of ways, from the laboratory, from the field, and from searching through a varied literature, not all of it in the public domain . . . . Indeed, since the problem-situations are presented in the environment, and much of the crucial data must be produced under controlled conditions in the laboratory, work in critical science may overcome the dichotomy between field-work and lab-work which has developed in science, even in the biological fields, over the past century (Ravetz, 1971, P. 429).

He recalls Bacon's 'pronouncement of philanthropic' science, of applying knowledge to the good for all rather than evil, and envisions science as a work of "practical charity inseparable from spiritual redemption" (Ravetz, 1971, 436). In effect, Ravetz called for a science which is socially responsible.

There were some forerunners to both Tranøy and Ravetz who laid some influential groundwork. Marxist thinkers have earlier called for a merger between social responsibility and science. J.D. Bernal was the most influential among Marxist scholars who were concerned with the external, social functions of science. At the same time there was the belief among those thinkers that this could be achieved without loss to the integrity and freedom of science. Science needed to be steered by science policies which directed it in the right directions, since science left on its own would not produce a universal Good. They believed that it could only be achieved under the right political conditions in a truly socialist or Marxist society.

Of course, mention should also be made to Robert K. Merton (1910–2003), who published an ethos of science already in 1942 (reprinted in Merton, 1973). Here he summed the ethic up in four, later in five, basic categories or virtues, the so-called

CUDOS norms of science: Communism (no property rights on knowledge), universalism (no discrimination regarding the sources of knowledge), disinterestedness (striving for objectivity), organized scepticism (peer-review as process of validation), and, later, originality (only the first inventor to be rewarded by fame). It should be mentioned, though, that this ethos was designed to describe the normative basis of normal science, pursued in academic disciplines under a foundational basis of academic freedom.[5] Even though this was basically describing an ethics of science, the ethics was purely internal, not reaching out to external welfare.

However, an early extension to external factors within the scientific research activity occurs in the 1950s. In 1953, Richard Rudner makes the basic point that any scientist makes value judgements qua scientist. Rudner's fundamental argument is that criteria for the acceptance of a hypothesis reflect basic value-judgements: "How sure we need to be before we accept a hypothesis will depend on how serious a mistake would be" (Rudner, 1953, 2). What Rudner has in mind is that standards of proof inject implicit value commitments. This simple claim of Rudner is an early expression of similar claims which were made years later, e.g., by Kristin Shrader-Frechette (1991) about the difference between epistemic rationality and practical rationality, and the related decisive differences between type 1 and type 2 errors; the difference between false-positive and false-negative; or, in everyday language detecting something which does not exist versus not detecting something which does exist. This is not about (external) values that influence whether people choose a career in science, or what science people pursue, or what research people engage in; this is about value-influences from social responsibility which are crucial in the conduct of science itself. This line of argument became especially influential with the publication of a book by Heather Douglas (2009), aiming to show the limitations of the value-free ideal in research.

After Thomas Kuhn (following Ludvik Fleck) seemingly opened the paths for a different embedding of scientific activities in historical contexts, the so-called STS ("Science-Technology-Society" or "Science and Technology Studies") community pursued this line with ever more vigour in analyzing the historical, political, and sociological context surrounding scientific practice and development. The slogan of "social construction" became a focal point in many receptions of this work, first as social construction of reality (Berger and Luckman, 1966) and then as the social construction of technology (Bijker and Pinch, 1987). This was seen by many scientists as an outright attack on their propagated goal of uncovering the truths of the world. Predictably it led to conflicts between different scientific communities, escalating in the so-called "science-wars" (Hacking, 1999). Many natural scientists felt that they were in a battle with social scientists fighting for the "survival" of facts and truth.

---

5 As descriptive tools for the workings of science they were effectively challenged by Ian Mitroff's study (Mitroff, 1974).

But for theorists (like Latour and Woolgar, 1979) studying the workings of science, it became seemingly more difficult to cleanly separate the facts from the values. Both seem intertwined, at least in all issues that matter to the people (cf. also Elliot and Resnik, 2014; Pelley, 2014; Spruyt et al., 2014). Even when something looks like a fact, if one looks closer, one discovers a sea of uncertainty and different value landscapes around it. I shall come back to this point under the heading of post-normal science.

Our whole conception of scientific knowledge and the production of it has changed since these early voices of criticism of the value-free ideal. Where there once seemingly was a need to ascertain a clear separation of the sciences from politics through, e.g., the Weberian "Wertfreiheit" (science as value-free; cf. Weber, 1973), now a need could be seen to ascertain a "social responsibility" of science. Both needs are, in a certain sense, sensible reactions to socio-political realities: first, dictatorships and ideologically in-fected states which competed about the "final truths", then a capitalist economy which sees the only true utilizations of science in its produced market values. The first reality called for a de-ideologization of truths, and the second called for a counterweight in social responsibility.

This is, for instance, mirrored in the call for a new social contract of science, as first voiced by Jane Lubchenko (1998). The background for this is the realization that (a) science needs a dialogue with society at large, and (b) long-term comprehensive sustainability must be worked into scientific developments, or else sciences' efforts to produce innovation and economic growth will fail because of public scepticism or even resistance. Science needs to support politics:

> A better understanding of the implications of the likely consequences of different policy options will allow more enlightened decisions. Many of the choices facing society are moral and ethical ones, and scientific information can inform them. Science does not provide the solutions, but it can help understand the consequences of different choices" (Lubchenko, 1998, p. 495).

4 The rise of ethics of science: research ethics and beyond

WWII was decisive on another count too. It marked the slow emergence of research ethics, first within medical ethics, culminating in the Helsinki Declaration (the first version in 1964, then with amendments up to 2013), via ethics of animal use in scien-tific experiments (the 3 R's: Reduce, Refine and Replace; first propagated by Russell and Burch, 1959), then spreading to other areas of research. Now ethics began to mat-ter for science in the sense that scientists had to worry about the objects of their study, be they humans, animals, or the environment. This was also a field where both non-governmental and governmental authorities were quite prepared to intervene, issue relevant guidelines, and establish oversight committees. Was this happening be-cause of internal pressures from within the sciences? Or was it a pressure from exter-

nal bodies and authorities? Well, the interpretation of the causal factors behind this development may probably vary. Still, I would venture to claim that one decisive factor here was the reaction among the publics. After the Nuremberg trials (Katz, 1996) against those who conducted "medical" experiments in Nazi concentration camps, the public outcry was that medical experiments could not bypass human rights and human dignity (Resnik, 2018; Briggle and Mitcham, 2012). A common morality was to set limits to what one could be allowed to do in the name of science.

Animal rights were in some countries even more the issue of public campaigns than the protection of human subjects in clinical trials. Animal rights and welfare even spurned activists to violently free animals from animal laboratories or threaten scientists working with such animals privately. Often there was plain hatred between the parties in these disputes. This was most outspoken in Anglo-Saxon countries, with the UK at the forefront.[6] Thus, some scientists became the "bad guys" in the eyes of a significant section of the public. And this was clearly a moral sentiment! Therefore, science had to get accustomed to being judged by a set of moral standards which were clearly from outside science itself but viewed as being overarching whatever standards science would proclaim for itself. Just being a good scientist was, in a sense, not good enough in the eyes of the public.

Interestingly, this development of research ethics was soon joined by another public concern: scientific fraud, misconduct, and integrity of research. After the publication of the book by Broad and Wade, *Betrayers of the Truth* (1984), concern arose that fraud might be as common in science as in many other areas of society; visible was perhaps only the tip of the iceberg. In view of Merton's *Ethos of Science* this was fundamentally counter to those mechanisms which were to ensure the proper progress of knowledge. Given that scientific quality assessment is/was the task of a scientific community, scientists had to trust the work of other scientists in order for the knowledge to be reliable. The United States was the first country where this topic of potential fraud in science became a matter of high-level politics. Interestingly, this was mainly because of the spending of public money for research (Mishkin 1999; Kaiser 1999). But other countries followed in the 1990s and later, with the First World Conference on Research Integrity in 2007 in Lisbon, then followed by several others with even bigger global audiences, with seven such conferences by 2023. So, now ethics was going under the skin of the scientist: '*watch out that you are not corrupted by the temptations of success in research, keep a professional ethics of honesty and veracity*'!

Note, however, that such an ethics, both the research ethics and the ethics of scientific integrity, were not in a strict sense a truly *comprehensive* ethics of science. And it did not really touch upon the dogma of a value-free science. While science was conceived as constrained by a common morality and human rights, the inner workings of

---

6 See e.g. the Wikipedia entry on Animal Liberation Front: https://en.wikipedia.org/wiki/Animal_Liberation_Front

science could still be assumed to be governed by the value-free ideal, provided they worked according to the ethos of science. In other words, the task was to focus on the facts of the world, now to be conducted with ethical restraint in designated areas but to leave the societal values to others, basically to the democratically elected bodies of State.

But WWII also held the very seeds to blow up even this conception. The key experience was the Manhattan Project and the construction of the atomic bomb. This revealed the inherent Janus face of scientific knowledge: producing the Good for humanity on the one side and producing the Bad for humanity on the other. The message emerging from the Russell-Einstein Manifesto in 1956 (On, 2012), and thus indirectly emerging from the Manhattan Project, was basically the call to physicists to work towards the peaceful use of atomic energy and abstain from getting involved in the destructive uses of this powerful energy resource. I think we can characterize this as a call for a socially responsible physics.

Yet, the Janus face was soon to get more substance from other movements: to a certain extent from within science, but also to a large extent from the concerned publics. The environmental movement is a case in point. The publication of the book *Silent Spring* by Rachel Carson (1962) started an environmental movement. It now emphasized that, e.g., the so-called Green Revolution, carried through by a lot of scientific work, did contribute to improving the volume of food production around the world, but at the same time, it contributed to environmental degradation through pesticides and herbicides. The rise of ecological research was to counteract these negative impacts, often portrayed as "unintended side-effects" of science and its technologies. Similarly, the rise of risk research, system analysis, and technology assessment was to give us the tools to control these negative consequences. Societal debates about nuclear power plants added to these lists of problems where science was seen as an instigator of bad outcomes or possible catastrophes.

These developments did lead to internal pressures within science towards greater responsibility for the outcomes and uses of science, both in the short term and in the long term. Jane Lubchenko's call for a new social contract for science was clearly an attempt to bring this into science policy. However, resistance came from powerful institutions and politics, even at the end of the 20th century.

Sir Joseph Rotblat received the Nobel Peace Prize on behalf of the Pugwash Movement in 1995. This was one of the reasons why he was invited to be among the keynote speakers for the opening day of the World Science Conference in Budapest in 1999, arranged by UNESCO and ICSU. His opening speech hit the audience like a bombshell.[7] He called for something akin to a Hippocratic Oath of Science (cf. also Ingierd, 2015): "The time has come to formulate guidelines for the ethical conduct of scientists, perhaps in the form of a voluntary Hippocratic Oath" (Rotblat, 1999).

---

[7] I was present then, accompanying our Minister of Higher Education and Science, and I can confirm the reaction of the audience.

The assembly gathered at this opening speech by Rotblat was immediately taken in by this simple message. There was a consensus that this had to be included in the final documents of the conference which were to be endorsed by the assembly 1 week later. The Nature-based conference newsletter reported the next day that the conference had decided to opt for a Hippocratic Oath of science, and the drafting committee started immediately to work on appropriate formulation. The following days were dominated by semantic wordsmithing in the drafting committee – until Wednesday. Then the message emerged from the drafting committee that one country – Brazil – had vetoed the inclusion of such a proposal in the final documents. Apparently, the Minister had realized that such an inclusion of an ethical social responsibility could eventually threaten its secret military research carried out in the country on behalf of another country. That in turn was a substantial source of national income. Once it was realized that ethics can have very practical consequences, it was removed from the political agenda. Interestingly, many other countries followed immediately afterwards, and Rotblat's proposal was dead.[8]

My reading of this story is the following: while by the end of the 20th century, many scientists (among them a few influential and powerful ones) and some scientific bodies were quite ready to explicitly acknowledge that social responsibility was essential, another side of the coin of academic freedom, funders of science, and national governments in industrialized countries and emerging economies were united to restrict the business of science to discovery of facts and working towards technological innovations. They had apparently a common interest in stopping the scientific community to engage in normative functions or debates of values which would have a direct political impact.[9]

5 Participation on the rise

If science were to take on a social responsibility, the question arises: who can speak for these responsibilities or who can actually express the societal needs to which the scientific community should feel bound? Furthermore, if trust in scientific expertise was seen to slowly decline in several fields during the 1980s and 1990s, for instance, in biotechnology, environmental science, agricultural sciences, risk assessments, medicine, technology, etc., what exactly was it that the publics expected from science in

---

8 I have told this story in the collection of my short stories *Talk, Eat, and See the World*. Available at: https://www.amazon.de/Talk-Eat-See-World-Fireside/dp/3752687916/ref=sr_1_1?__mk_de_DE=%C3%85M%C3%85%C5%BD%C3%95%C3%91&crid=2BQBH69PXQE8V&keywords=talk+eat+and+see+the+world&qid=1706220845&sprefix=talk+eat+and+see+the+world%2Caps%2C118&sr=8-1
9 In global hearings on the ethics of science arranged by the UNESCO, representatives of the Foreign Office of the USA regularly read out a letter where the USA protested strongly to the UNESCO engaging in normative questions!

these fields? The answer to that seemed obvious: talk to these publics and listen to them! One such novelty was public understanding of science, a topic which got its own scientific journal in 1992 in the UK to focus on science communication. But even earlier, many social scientists and indeed scientists in different disciplines sought out the voices of the publics to gauge their sentiments and expectations. The call was for inclusion and engagement of the publics, in other words, participation in science (this point is also developed in Chapter 7).

I have dealt with the various sources for this call of inclusion and participation in another publication (Gethmann et al., 2015, chapter on participation). I identified four sources of participation: public administration, ("applied") social science, development aid, and technology assessment. I shall not repeat the details here but mention just one commonality, namely the erosion of trust in scientific expertise coupled with the insufficiency of it in certain contexts, and the lack of efficiency of top-down management.

Let me describe a development in social science, reminding the readers of the sources of so-called *action research*. Kurt Levin recognized during WWII that top-down governance of food policies was not getting a hold on the public endorsement of official recommendations. Obviously, social science lacked what in natural science was termed an applied version of the sciences that could enable scientific insights to interact with the controlled changes of their domains. He therefore instantiated discussion groups with consumers to derive such recommendations and come to a general understanding of their rationale. This turned into action research in the 1960s and 1970s, partially influenced by Marxist ideals of social change. *Action research* in the social sciences became especially popular in the Scandinavian countries but slowly made its way also into other scientific communities. Scientific expertise was deemed as often biased and basically insufficient to initiate socio-political changes. Without a clear articulation of the voices of the affected parties, i.e., the publics, scientific expertise remained contested and largely ineffective in societal change. In sum, developments within some of the sciences after WWII experiences opened the path towards inclusion and engagement with the diverse publics. Participation was seen as a tool to overcome distrust in science and its failure to effectively move society towards reform and change to realize positive developments.

6 Post-normal science

One important contribution in the early 1990s was the publication by Funtowicz and Ravetz (1993), outlining the basics of what they termed "post-normal science". The term was an extension of the Kuhnian term "normal science," which basically characterized non-revolutionary change in academic disciplines. Post-normal science was characterized by its mantra: a) facts are uncertain, b) stakes are high, c) values are

disputed, and d) decisions are urgent. This was initially proposed as the characteristic of science-for-policy but was later recognized as the characteristic of all science dealing with highly complex (not just complicated) matters of societal relevance. Here they extended Weinberg's (1972) notion of "trans-science", i.e., questions that could be formulated in the sciences but could not be answered with the means and tools of science, like health effects of low-dose radiation. They also related to Rittel and Webbers' (1974) characterization of "wicked problems", a typical challenge in planning and social policy: "Problem understanding and problem resolution are concomitant to each other". Wicked problems have no definite problem formulation and are always multi-causal, multi-scalar, and inter-connected; there exist no true or false solutions to them, only better or worse answers; and there is no stopping rule for when a problem can be regarded as solved. This new social reality introduced an impasse between science and action.

Post-normal science called for a more explicit assessment of the uncertainties which are always inherent in scientific knowledge. One of the tools they introduced was the NUSAP[10] scheme for assessment (cf., e.g., Funtowicz and Ravetz, 1990; 1993). Since facts and values are necessarily intertwined in post-normal science, they pleaded for "extended peer-reviews", the inclusion of affected parties and the diverse publics.

The diversity of knowledge claims and the diversity of value landscapes (my term, cf. Kaiser, 2022, and Kaiser, 2024) implied a recognition of the insufficiencies of purely scientific expertise in complex matters of societal decision-making (cf. also: De Grandis, 2016). Funtowicz and Ravetz never proposed post-normal science as a new philosophy of science, or a "school" of thought. Instead, it was intended to characterize some signposts for a framework of understanding how science operates in these complex societal affairs. I interpret this as a basically ethical plea for the betterment of scientific practice. Post-normal science mainly concerns the perspectivity and diversity of dealing with complex issues. Therefore, I view post-normal science as essentially an ethical framework in the production of knowledge.

7 Transdisciplinarity

Post-normal science also paved the way for a new scientific reality and format for research: transdisciplinary research.

So far, I have outlined several developments which all contribute to the slow break-down of the dogma/ideology of a value-free science. It is certainly not the case that the ideal of value-free science has disappeared in the scientific community, but it is also quite noticeable that a significant number of engaged scientists in our days try

---

10 NUSAP is the short expression for the uncertainty communication they proposed: Numeral, Unit, Spread, Assessment, Pedigree.

to break out of what some have called the scientific "silos". And many of them recognize that knowledge and belief from wherever they originate will always have perspectivity and thus implicitly a certain bias (Saltelli et al., 2020). The problems of present times are too vast, too important, and, at the same time, too complex to maintain that scientific silos of isolated experts can come up with solutions that are socially robust and pave the way to a sustainable future.

One of the main challenges in science policy and strategic science policies is that basic – or "pure" – science is often seen as threatened by the instrumentalization and commodification of scientific research. Efforts to improve the social relevance and stress the social responsibility of science are seen as undermining scientific freedom and marginalize basic science as performed in the scientific disciplines. However, this perceived opposition between the goals of scientific research is a misunderstanding of the intentions of institutional reforms, in particular, the move towards transdisciplinary research. The point is not to replace the one with the other but to seek out the best of what disciplinary and transdisciplinary research have to offer. One needs to realize that scientific insights need to be contextualized and reflective in regard to inherent bias. We still want our immunologists to develop the vaccines we need to counter the next pandemic. But we also need to realize that the socio-political complexities of a pandemic cannot be captured by one or a few disciplinary silos. We also need to realize that what is produced within the disciplines is not necessarily a true picture of our reality. It is always an abstraction, an idealization, often cast in the form of models, and even within the disciplines, the claim of knowledge is oftentimes disputed, and knowledge production has a variety of forms. These critical viewpoints do not make basic science superfluous but acknowledge the post-normal insight that quality assessment asks for extended peer reviews.

Typically, the problems start already with the formulation of the problem. Rittel and Webber (1973) pointed out that, for complex societal issues, no fixed or universally agreed problem formulation seems possible. Disciplinary scientists typically start out with a tunnel-view: the vague societal problem is re-formulated in terms which make the problem amenable to the methodological tools the discipline can offer. Another hurdle is the tendency to sideline or exclude sources of knowledge that are not subjected to standard scientific procedures or quality checks. Indigenous and local knowledge systems are an example of this, as are professional or traditional knowledge systems. Scientific hubris in regard to these knowledge systems is not only misplaced and disrespectful but is also counter-productive in overlooking that these systems have produced useful knowledge over time. Furthermore, the assumption that all knowledge should be value-free and should only build on hard facts is mistaken and equally unproductive.

Diversity, contextuality, and perspectivity are the key terms for transdisciplinary research. Large global institutions like the OECD, UNESCO, and the *International Science Council* (ISC) have now supported the move towards transdisciplinary research

as a necessary component to address the grand societal challenges of our time and move towards the realization of the SDGs.

In a co-authored discussion paper Peter Gluckman and I (Kaiser and Gluckman, 2023) have outlined both the essential characteristics of transdisciplinarity and the institutional hurdles and challenges that this research meets still today in our science system. Transdisciplinary research deals with and incorporates societal values and the diversity among them explicitly and constructively. It builds on reflexivity which sees these values instantiated in often conflicting knowledge claims and seeks out mutually acceptable strategies to improve the status quo. In this sense, transdisciplinary research is the embodiment of the opposite of value-free science and embraces the social responsibility of scientific knowledge.

8 RRI

The term "Responsible Research and Innovation" (RRI) is a child of the European Commission and its framework programs for research. More specifically, it is probably the child of mainly one person from within the European Commission: René von Schomberg (cf. Von Schomberg, 2013; 2019). Like the term *Corporate Social Responsibility* (CSR), it highlights responsibility. As I have been trying to indicate in the preceding sections, it took time to recognize responsibility as the other side of the coin, where academic freedom was already inscribed. While freedom is related to the practicing individual (and their institutions), responsibility establishes the bridge to the societal context and delineates an essential ethical norm. Unlike the early proponents of a close relationship between science and ethics, for instance, Condorcet, the new driver towards social responsibility and an ethics of science was from outside the scientific communities. The perceived "sins" of science and the disputed benefits of the technological changes it proposed were the factors that necessitated the changes in funding and science policy. Yet, the political realities surrounding science were – and *are still* – infected by a mistrust that granting too much social responsibility to the scientific enterprise, even in the form of co-responsibility, was granting too much power to scientists. In many countries scientists were perceived as a critical bunch dissatisfied with governments. This political reality demanded pragmatics of reform which portrayed RRI as a linear development from the fact-finding and value-free community of the past to a new and creative partnership with States and their endorsed development goals (like the SDGs). After seeking out ELSA and relations between "Science and Society" or "Science with and for Society" – and all including ethics – RRI was now the new term coming from the top of funding and science policy. These circles gave the impression they discovered and pushed for a better ethics of science, from the top to the bottom, as a response to critical publics, their electorate. Admittedly it was true and still is true that significant parts of the scientific com-

munity oppose taking on moral co-responsibility on the basis that they perceive major moral failures in the policies of States which again and again ignore their recommendations which – in their mind – are purely objective facts to take into account in what often is called *evidence-based policy*. But ever since WWII, there has been a change within other parts of the scientific community that pushes for engagement in policies for societal change, not only technological innovation, as it is seen as merely instrumental to political and economic values and interests. The emergence of RRI within the EU system thus signalled for many a seemingly bureaucratic imposition on how to write project proposals rather than a re-thinking of the basics of the scientific enterprise. While the concept seemed good, the intentions behind it were suspicious. Science was to join in the ranks of well-structured, state-run, and industry-friendly reform rather than becoming an independent agent of social change. As I perceive it, it was not an outcome of academic freedom but rather an outcome of power politics. However, even if the initiative may have come from "above", apparently, many scientists picked it up and strove to make it a bottom-up reality. The Norwegian AFINO network seemed to document this engagement with scientific responsibility.

I admit I am possibly exaggerating what I see as underlying drivers out of perhaps my personal bias, and repeated experiences of how ethics of science became instrumentalized, when not ignored. But I will defend my point that the social responsibility of science and its research are indeed an essential part of our age-old culture of knowledge. Our systems of science were baked into a socio-political framework with strong interests and cultural bias. Our age-old knowledge traditions, on the other hand, pretended at least that knowledge was a social good serving the interests of the people. Looking back to history, we can clearly see how the dogma of value-free science served particular interests. We can also see that an ethics of science cannot be formulated without a clear reference to social responsibility. Social responsibility belongs to the scientific community in the same sense that RRI should not belong to the mechanisms of political and economic powers.

9 Issues with RRI

So, why do we not just continue with RRI and try to make it a real commitment to the works of science? Well, the European Commission has removed the reference to RRI from its central expectations in the new framework programs. Furthermore, as, for example, Giovanni De Grandis and Anne Blanchard point out in their contributions to this book, RRI imposes quite demanding conditions on research, which for many, especially younger scholars, are difficult to meet (see Chapters 4 and 5). This may be due to conflicting disciplinary cultures and diversity in value commitments. Be that as it may, including RRI in standard scientific research is experienced as perhaps too de-

manding a task for researchers who are stuck in their disciplinary silos and in severe competition.

Personally, I believe one serious issue with RRI is that it is commonly perceived as a policy imposed from top-down, and not emerging from bottom-up, as part of the routine training of scientists. Similar with ethics: ethics has largely been reduced to rubber-stamping in standardized forms that follow applications for funding. Learning how to do scientific research means mostly learning specific methods of how to generate and analyze data. The contextualization of research, the reflexivity of one's own bias, tunnel views, and pitfalls do not enter. This can only be an outcome if universities provide good platforms where dialogue between disciplines from all faculties and dialogue with intended users of the knowledge are regular features of education and research. The new ways of transdisciplinary research need to be supported much more coherently if science is to take on the social responsibility that Lubchenco was concerned about. Most scientists believe that their research is for the good of people, but very few anticipate misuse or controversy. It is through dialogue with others that we may learn about other viewpoints, other value-landscapes, and other systems interfering with what we do.

10 Conclusive remarks

The point of this chapter is to acknowledge the diversity of context in knowledge production and maintenance. One of the contentious issues has been and still is the relationship between science and the ethical responsibility of science to society. I tried to argue the case that knowledge and ethical values were, for a very long time, two sides of the same coin, or virtually indistinguishable. We were striving for knowledge for our own good and the good of our community. Neglecting or denying the ethics of knowledge was the result of the political and cultural circumstances of the time. During the Scientific Revolution and the rise of the principles of science, the stress on the possible instrumental value of the scientific enterprise was a pragmatic move to avoid conflict with the existing powers. Thus, abstaining from "morality" seemed to be the consequence. But even then, the focus was on the good for all people; science was perceived as instrumental knowledge and this knowledge as a tool to get humanity out of misery. With the 19th and particularly with the 20th century came the move to stress the supposed objectivity of the sciences versus the subjectivity of the moral world. Modernity introduced the institutional separation between who is to take care of the facts and who is to take care of the values. The former provide the options and the latter decides the course. The systems of science were steered by the political, social, economic, and cultural conceptions of the world we live in. Colonialism set in to subdue or ignore other valuable knowledge traditions. In the 20th century this turned into the powerful ideology of a value-free science. It provided a convenient framework for sheltering po-

litical and economic powers from interference by the growing academic elite. In effect, it was also the mechanism that constrained and subdued the celebrated academic autonomy in terms of abstinence from politics. And for the scientific elite, it provided the promise of increased financial support from the state, the carrot behind the stick that Vannemar Bush's program from 1948 (Bush, 2020) promised.

Will it endure? The prospects of transdisciplinary research as they emerged on the background of post-normal science (Kaiser and Gluckman, 2023) at least contain the seeds to break loose from this dogma and to embrace in research the diversity of knowledge embedded in our value-landscapes (Kaiser, 2024). And this may turn the ethics of science into a centre-stage measure of quality. This is, at least, my personal hope.

References

Berger, P., and Luckmann, T. (1966). The social construction of reality. Routledge.

Bijker, W. E., and Pinch, T. J. (1987). The social construction of fact and artifacts. *Philosophy of technology: the technological condition: an anthology*, 107–139.

Briggle, A., and Mitcham, C. (2012). *Ethics and science: An introduction*. Cambridge University Press.

Broad, W., Wade, N., and Armstrong, H. L. (1984). Betrayers of the Truth. Simon and Schuster Publishers.

Bury, J. B. (1921). *The idea of progress: An inquiry into its origin and growth*. Macmillan and Company, limited. (here cited after the online version: https://www.gutenberg.org/files/4557/4557-h/4557-h.htm#link2HCH0018)

Bush, V. (2020). *Science, the endless frontier*. Princeton University Press.

Cantor, G. (2012). "Science, Providence, and the Progress at the Great Exhibition"; Isis, 103, 439–459.

Carson, R. (2009). Silent spring. 1962.

De Grandis, G. (2016). "Practical integration: The art of balancing values, institutions and knowledge – lessons from the History of British Public Health and Town Planning", Studies in History and Philosophy of Biological and Biomedical Sciences, vol. 56, 92–105.

Douglas, H. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Press.

Elliot, K.C. and Resnik, D.B. (2014) "Science, Policy, and the Transparency if Values" Environmental Health Perspectives 122 (7), 647–659.

Funtowicz, S. O., and Ravetz, J. R. (1990). Uncertainty and quality in science for policy (Vol. 15). Springer.

Funtowicz, S. O. and Ravetz, J. R. (1993), "Science for the Post-Normal Age", Futures 26/6: 568–582.

Gethmann, C.F., Carrier, M., Hanekamp, G., Kaiser, M., Kamp, G., Lingner, S., Quante, M., Thiele, F. (2015). Interdisciplinary Research and Trans-disciplinary Validity Claims, Springer: Cham, Heidelberg, New York, Dordrecht, London. DOI 10.1007/978-3-319-11400-2.

Hacking, I. (1999). *The social construction of what?*. Harvard University Press.

Hare, R.M. (1952). The Language of Morals. Clarendon Press, Oxford

Ingierd, H. (2015). Technology and commitment. Integrity In The Global Research Arena, 85.

Katz, J. (1996). "The Nuremberg Code and the Nuremberg Trial. A reappraisal". *JAMA*. **276** (20): 1662–6.

Kaiser, M. (1999), "Development of International Guidelines for Research Ethics (Commentary on Mishkin)", Science and Engineering Ethics, 5 (1999), 293–298.

Kaiser, M. (2003), "Forskningsetikk i Norge – Norge viser veien?", i: Forskningsetikk, Knut Ruyter (ed.), Gyldendal Forlag: Oslo, 2003.

Kaiser, M. (2022). "Taking value-landscapes seriously", in: Transforming food systems: ethics, innovation and responsibility, D. Bruce and A. Bruce (eds.), Wageningen Academic Publishers, Wageningen, pp. 46–51.

Kaiser, M. and Gluckman, P. (2023). Looking at the future of transdisciplinary research. International Science Council, April 2023. DOI: 10.24948/2023.05. Available at:https://council.science/publications/looking-at-the-future-of-transdisciplinary-research/

Kaiser, M. (2024). "The idea of a theory of values and the metaphor of value-landscapes". Humanities and Social Science Communications, Springer/ Nature.

Latour, B., and Woolgar, S. (1979). *Laboratory life: The construction of scientific facts*. Princeton university press.

Lubchenco, J. (1998). Entering the century of the environment: a new social contract for science. Science, 279(5350), 491–497.

Merton R.K. (1942/1973) The normative structure of science. In: Merton Robert K, editor. The sociology of science: theoretical and empirical investigations. Chicago: University of Chicago Press; 1942. 1973.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago press.

Mishkin, B. (1999). «Scientific misconduct: Present problems and future trends", Science and Engineering Ethics, vol.5, 283–292.

Mitroff, I. I. (1974). Norms and counter-norms in a select group of the Apollo moon scientists: A case study of the ambivalence of scientists. *American sociological review*, 579–595.

On, F. Y. (2012). The Russell-Einstein Manifesto.

Pelley, J.L. (2014), "Science and Policy: Understanding the Role of Value Judgements", Environmental Health Perspectives 122(7), A 192.

Poskett, J. (2022). Horizons. The Global Origins of Modern Science. HarperCollins Publishers, NY.

Resnik, D. B. (2018). *The ethics of research with human subjects: Protecting people, advancing science, promoting trust* (Vol. 74). Springer.

Reichenbach, H. (1938). Experience and prediction: An analysis of the foundations and the structure of knowledge. University of Chicago Press, Chicago and London.

Rittel, H. W., and Webber, M. M. (1974). Wicked problems. Man-made Futures, 26(1), 272–280.

Rotblat, J. (1999). A Hippocratic Oath for scientists. Science, 286(5444), 1475–1475.

Rudner, R. (1953). "The scientist qua scientist makes value judgements" Philosophy of Science, 20(1), 1–6.

Russell, W. M. S., and Burch, R. L. (1959). *The principles of humane experimental technique*. Methuen.

Saltelli, A., Benini, L., Funtowicz, S., Giampietro, M., Kaiser, M., Reinert, E., and van der Sluijs, J. P. (2020). The technique is never neutral. How methodological choices condition the generation of narratives for sustainability. Environmental Science and Policy, 106, 87–98.

Sharlin, A. N. (1974). Max Weber and the origins of the idea of value-free social science. *European Journal of Sociology/Archives Européennes de Sociologie*, *15*(2), 337–353.

Shrader-Frechette, K. S. (1991). Risk and rationality: Philosophical foundations for populist reforms. University of California Press.

Spruijt, P., Knol, A.B., Vaseleiadou, E., Devilee, J., Lebret, E., Petersen, A.C. (2014), "Roles of scientists as policy adviser son complex issues: A literature review" Environmental Science and Policy 40, 16–25.

Toulmin, S. E. (1950). An examination of the place of reason in ethics. Cambridge University Press Archive. Cambridge

Von Schomberg, R. (2013). A vision of responsible research and innovation. *Responsible innovation: Managing the responsible emergence of science and innovation in society*, 51–74.

Von Schomberg, R. (2019). Why responsible innovation. *International handbook on responsible innovation: a global resource*, R. Von Schomberg and J. Hankins (eds.), p.12–32.

Weber, M. (1917). Der Sinn der »Wertfreiheit« der soziologischen und ökonomischen Wissenschaften. In: M. Weber: Gesammelte Aufsätze zur Wissenschaftslehre. Tübingen 1988, S. 489–540.

Weinberg, A. M. (1972). Science and trans-science. *Science*, *177*(4045), 211–211

Christian Wittrock, Gustaf Nelhans, Anne Ingeborg Myhr,
and Svein Ole Borgen

Chapter 3
Norwegian engagement with RRI and the propagation of RRI by the Research Council of Norway

Country profiles for the uptake and engagement with responsible research and innovation (RRI) show remarkable differences. Likewise, the role of research funders in instigating changes in the science system is widely recognised, but in the field of RRI, few studies detail efforts beyond the British Engineering and Physical Sciences Research Council (EPSRC). However, the Research Council of Norway (RCN) has been one of the few large-scale national funders wholeheartedly embracing RRI. In this chapter, we undertake a descriptive study highlighting Norwegian RRI scholarship and the role of the RCN in furthering RRI in the country. We first compare Norwegian research output on RRI to other major contributors, showing that Norwegian RRI research is surprisingly voluminous. When country size is accounted for, Norway is in the top two, only surpassed by the Netherlands. We then detail the funding schemes originally used by the RCN to further RRI in four fields of research: biotechnology, nanotechnology, ICT and digital innovation, and finally responsible innovation and CSR. We discuss evaluations of the programmes and detail funding provided for RRI activities by the RCN. We proceed by briefly touching on recent developments at the RCN. Despite significant organisational turmoil in 2022, the RCN keeps promoting RRI in very visible ways. However, funding dedicated RRI has been scaled back after a marked uptick in the mid to late 2010s. We finish with suggestions for further research.

1 Introduction

Country profiles—and imaginaries—may be more or less conducive to the uptake of RRI (Doezema et al., 2019; Randles, 2016; Ryan et al., 2021; Wittrock et al., 2021). While our knowledge of such country profiles, and divergence between countries, is growing with respect to implementation patterns, no research has to date sought to gauge the extent to which national research communities have actively engaged with RRI, using a bibliometric approach. This chapter provides initial inroads into that area of country comparison. Relatedly, the furtherance of RRI through the efforts of research funding organisations is an important driver for RRI diffusion in national, as well as global, science systems (Owen et al., 2021; Wittrock et al., 2021, p. 4). Therefore, research programmes promulgated by funders should be a cornerstone of RRI diffusion

and application. However, to our knowledge, only the European Union, the Netherlands Organisation for Scientific Research (NWO), the British Engineering and Physical Sciences Research Council (EPSRC) (now UK Research and Innovation—UKRI), and the Research Council of Norway (RCN) have wholeheartedly embraced RRI in a consistent way (Daimer et al., 2023; Wittrock et al., 2021). To this list comes the—comparatively—smaller Telethon Foundation in Italy, which specialises in rare diseases (Neresini & Arnaldi, 2018).

With respect to large national funders, the efforts of the EPSRC are well documented in a plethora of scientific works and reports (see e.g. Macnaghten & Owen, 2011; Owen et al., 2021; Wittrock & Forsberg, 2019). However, the undertakings of the NWO and the RCN so far remain documented largely in research reports (Egeland et al., 2018; van der Molen et al., 2018). In the case of the RCN, Egeland et al. (2019) additionally show how the RCN undertook RRI implementation as an organisational learning trajectory rather than, for instance, blueprint emulation. *To shed further light on engagement with RRI in Norway, and the RCN's involvement with RRI, we thus undertake an exploratory study with three foci.* First, we place the Norwegian case in the broader landscape of RRI research in terms of scientific output, considering published scholarly work on RRI. Such bibliometric analysis is commonly taken as a measure of the interest in some concept—in our case RRI (see Benders et al., 2007). This step documents that the Norwegian embrace of RRI appears to be one of the strongest worldwide in terms of measures for output and impact, particularly when the size of leading countries is taken into account. Second, we detail the undertakings of the RCN, which by far is the largest funder of research in Norway. We highlight the four flagship funding programmes the RCN originally used to promote RRI in the Norwegian innovation and science system and show the development in funding provided for projects with an RRI component. Third, we turn to recent developments within the RCN, which is characterized by encompassing organisational upheaval, but recently also an amplified effort to further RRI. While the three foci selected cannot say much about the success of RCNs pursuit of RRI, they do say something about the interest in RRI among Norwegian scientists, the considerable effort of the RCN to promote RRI, and the general weight of RRI within the RCN.

Methodology

In our analysis of Norwegian research output on RRI, we used bibliometric analysis of published scholarly work (Zupic & Čater, 2015). We employed the Web of Science (WoS) database and search terms designed to include as many relevant publications

as possible.[1] We thus followed but also significantly expanded on the work by Liu et al. (2022). Our search string is informed by but less inclusive than the one employed by Randles et al. (2022, p. 259), as we are interested in the core of RRI research and not in the genealogy of RRI. Our study contains 1,689 scientific works. Following the recommendations of Strang and Wittrock (2019), we searched both author-supplied keywords, abstract, and title. We consider published works from 2003 to 2023 (inclusive). According to the RRI genealogy study by Randles et al. (2022, p. 250), the former coincides with the first recorded use of the search terms as a central part of the scientific contribution (Hellstrom, 2003). The latter is the end of data collection for our contribution. The publication record for 2023 is not yet complete in the database used. Thus, it can be assumed that the number of entries for 2023 will increase when all publishers deliver data to our database. As a policy concept, RRI success is not straightforwardly manifested in publication output. However, publication records are credible proxies for interest in some concepts, in particular among publishing scientists (Benders et al., 2007; Strang & Wittrock, 2019). But the extent to which the documented interest in RRI delivers on the outcomes sought after by the RCN is not shown by the analysis. Likewise, our analysis does not link publication output to the efforts of the RCN in a direct manner.

In our description of the RCN's efforts to further RRI, and the subsequent changes in efforts to promote RRI, we drew on desktop research and interviews with both RCN staff and key stakeholders in Norwegian RRI (expert interviews). In addition, we had extensive email exchanges with the informants before and after interviews. In all but one case, interviews were recorded and transcribed verbatim. The processing of the thus available material elicited responses to two concerns: (1) the role of the flagship programmes that were designed to further RRI and (2) recent developments at the RCN with regards to RRI. In total, we conducted ten interviews covering most staff that were central to the development of RRI within the RCN and selected central experts. We view these as 'institutional entrepreneurs', often referred to as 'RRI Champions' in the RRI literature (e.g. Randles, 2016). Institutional entrepreneurs are actors—or groups of ac-

---

[1] We used the search strings: ("responsible innovation"), ("responsible research"), ("RRI" AND "innovation") and ("RRI" AND "responsib*") and controlled for false positives through several iterations, including screening of abstracts using Rayyan. We constructed a list of 192 terms abbreviated RRI by browsing the content of searches and employing a set of abbreviation lists from web resources to weed out false positives. A fifth search string, ("RRI" AND "research"), was considered, but since it yielded a large number of false positives, we used the residual search results that were found when it was used together by removing all results from the original set of search strings. This yielded 12 additional entries. We experimented with both the SCOPUS database and the WoS. WoS provided the opportunity to focus solely on author-provided keywords which is not possible in Scopus. Thus, in Scopus available keywords may be back-fed into the database for example by a librarian or by some automated procedure (see, e.g., Strang & Wittrock, 2019). We thus selected WoS to comply with our search strategy. Additional programmes used to assess data were the Bibliometrix R-package with the Biblioshiny interface. Further data processing was done using HistCite and VOSviewer, but we only report the main findings with respect to volume in this chapter.

tors—who initiate and participate in divergent change activities (Battilana et al., 2009). To detail the RCN's funding activities with respect to RRI, we interrogated the database publicly available on the RCN website, using the keyword facility (The Research Council of Norway, n.d.).[2] Following efforts to democratise science, all informants were given the opportunity to comment on the chapter.

2 Norwegian contributions in the context of global RRI scholarship: A bibliometric analysis

Figure 1 shows that RRI has been a dormant topic in terms of publication output up until 2013. Around 2014, this changes abruptly. From 2014 and onwards, there is a steep rise in publications discussing the topic. RRI is thus a discourse and science topic, which appears still on the rise. This timing largely coincides with the publication of the well-known EPSRC framework by the trio Jack Stilgoe, Richard Owen, and Phil Macnaghten (Owen et al., 2012; Stilgoe et al., 2013). Bibliometric analysis is a common choice of method in management fashion studies, used to gauge interest in some innovative concept or idea (Clark, 2004; Strang & Wittrock, 2019).
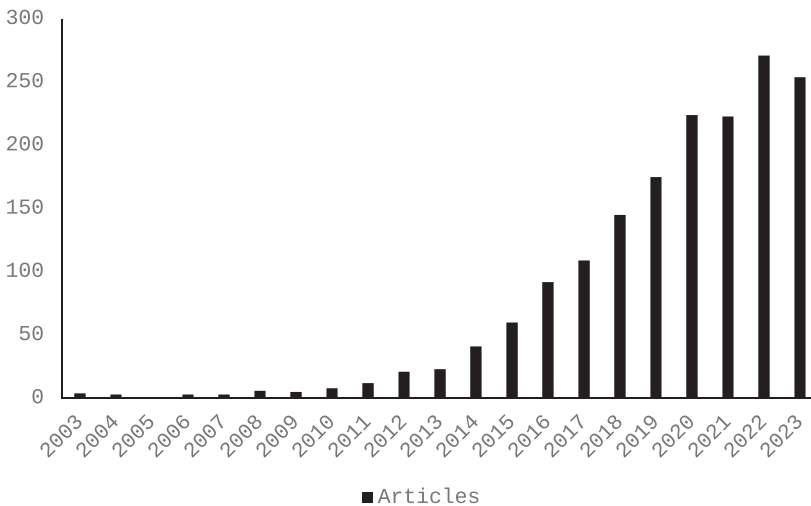

Figure 1: Global publications in our database discussing RRI 2003 to 2023.

Norwegian RRI publication output largely follows the global trend shown in Figure 1, though there is a burst in 2017 and a slump in 2020. In terms of publication output,

---

2 Our search returns projects which use 'RRI' as a keyword in either title or abstract.

Table 1 shows that Norway is in the top ten globally in our dataset of English language scientific work, a fact that may be taken as a measure of scholarly interest in RRI in Norway.

Table 1: Research output and impact of output from top contributors.

Country	Count research output	Total Citations Ranking in ()		Average article citations Ranking in ()	
USA	728	3477	(3)	16.30	(3)
UK	723	7929	(1)	31.70	(1)
NETHERLANDS	634	4887	(2)	21.20	(2)
GERMANY	309	917	(5)	8.10	(8)
CANADA	276	1125	(4)	13.60	(4)
SPAIN	245	757	(6)	9.80	(9)
ITALY	218	411	(10)	6.80	(11)
FRANCE	189	339	(11)	12.60	(5)
AUSTRALIA	179	412	(9)	8.10	(8)
NORWAY	152	654	(7)	10.20	(7)
AUSTRIA	122	468	(8)	10.60	(6)
CHINA	110	248	(12)	5.80	(12)
BELGIUM	109	208	(13)	8.30	(10)

The table shows all countries with more than 100 publications in our database of 1,689 entries. Note that entries may have mixed country authorship.

Norway is a Scandinavian country with a population of only approximately 5.5 million inhabitants (Statistisk sentralbyrå, 2024). It is thus surprising that Norway is among the top contributors to RRI research in terms of published scholarly work. In raw output measures, Table 1 shows that Norway is the tenth largest contributor. Norway by far surpasses the other Nordic countries. Where the United States is the largest contributor, with 728 scholarly works in the WoS database, Norway has produced 152 works during the period considered. However, output may also be considered in relation to impact, typically measured by citation scores. Here Norway also fares well. This is equally true, if we consider impact in the form of average citations per published work. As Table 1 shows, Norway surpasses Germany, Australia, Spain, Belgium, Italy, and China in the top ten with regards to average citations per published work.

The fact that Norway is in the top ten becomes even more impressive when figures on research output are put in perspective. Table 2 shows that the overall tertiary education sector is one of the smallest in our sample of the 13 top RRI contributors, with more than 100 published works.

In Table 3, we contrast the output from main net contributors using the number of inhabitants per country, the number of students enrolled in tertiary education, and the number of people employed in tertiary education. The two latter serve as proxies for the size of the university sector, drawing on the most recent available datasets

Table 2: Measures of relevance for RRI output comparison.

Country	# Inhabitants in mill.	# Tertiary education in millions. / 2018 (World Bank data)	# FTE in tertiary education in thousands/ 2021 (OECD data)
CHINA	1,412.000	44.935	n.a.
USA	331.900	18.941	1,072.99
GERMANY	84.607	3.127	266.31
FRANCE	67.750	2.618	*111.09
UK	67.330	2.467	159.90
ITALY	59.110	1.895	101.00
SPAIN	47.420	2.051	147.02
CANADA	38.250	1.622	n.a.
AUSTRALIA	25.690	1.677	n.a.
NETHERLANDS	17.530	0.889	60.34
BELGIUM	11.590	0.515	**21.1
AUSTRIA	8.956	0.430	31.34
NORWAY	5.553	0.288	27.01

The table shows relevant corrective measures for the evaluation of research output for all countries with more than 100 publications in our database of 1,689 entries. These are inhabitants in millions, the number of students in tertiary education in millions and the number of full-time equivalent staff in tertiary education in thousands. *France: value is for public institutions only **Belgium value for is from 2019. N.a. = not available.

from the World Bank (World Bank, 2018) and the Organisation for Economic Co-operation and Development (OECD, 2021), respectively.

In terms of published works relative to country size, the Netherlands stands out. But Norway follows on its heels and by far surpasses both Austria and the UK. Keeping in mind that the focus on RRI was spearheaded by United States, UK, and Dutch scholars, it is surprising to find Norway by far surpassing the UK in terms of research output relative to country size and as an undisputed number two globally.

If we consider the publication output relative to students enrolled in tertiary education (any degree after high school), the Netherlands stands out again. But, as shown in Table 3, Norway by far outperforms the UK, Austria, and Belgium, which are the other countries ranking high by this measure. When considering the output relative to FTE in tertiary education, the Netherlands once again distances itself from any other country, but Norway again comes second, this time with both Belgium and the UK at its heels.

Drawing on this comparison, we suggest Norway may be one of the most successful countries globally, in terms of RRI scholarship, only surpassed by the Netherlands. Norway thus appears exceptionally concerned with RRI and has contributed significantly to research on RRI, despite its small size.

Table 3: International comparison of RRI scholarship, output relative to other measures.

Country	Relative output to # Inhabitants in mill		Relative output to # in tertiary education in mill. Ranking in ()		Relative output to # FTE in tertiary education in thousands Ranking in ()	
NETHERLANDS	36.17	(1)	712.75	(1)	10.50	(1)
NORWAY	27.37	(2)	526.42	(2)	5.62	(2)
AUSTRIA	13.62	(3)	283.59	(4)	3.89	(5)
UK	10.74	(4)	293.05	(3)	4.52	(4)
BELGIUM	9.40	(5)	211.43	(5)	5.16	(3)
CANADA	7.22	(6)	170.08	(6)	n.a.	
AUSTRALIA	6.97	(8)	106.72	(9)	n.a.	
SPAIN	5.17	(7)	119.40	(7)	1.66	(7)
ITALY	3.69	(9)	114.97	(8)	2.15	(6)
GERMANY	3.65	(10)	98.78	(10)	1.16	
FRANCE	2.79	(11)	72.17	(11)	1.70	
USA	2.19	(12)	38.43	(12)	0.67	(8)
CHINA	0.08	(13)	2.44	(13)	n.a.	

The table shows publication output from all countries with more than 100 publications in our database of 1,689 entries, relative to inhabitants in million, the number of students in tertiary education in millions and the number of full-time equivalent staff in tertiary education in thousands, based on the best available measures.

3 The Research Council of Norway: Supporting the furtherance of RRI through four flagship programmes

The Research Council of Norway (RCN) is the result of a merger of five smaller state-governed funding agencies responsible for pursuing national research goals, as these are decided by the government and parliament in Norway (Egeland et al., 2018). The RCN is the main research funding body, and it administers research funds from almost all the Ministries in the Government and funds research in all disciplines: basic research, applied research, and innovation. Consequently, the RCN's strategy and plans for calls set the agenda for research policy in Norway (Solli, 2023).

Though the organisation has later changed, four flagship research programmes became central in the RCN strategy to promote RRI: BIOTEK2021, NANO2021, IKT-PLUSS, and SAMANSVAR (Gulbrandsen & Rynning, 2016). However, RRI thinking was an early part of the large biotechnology and nanotechnology programmes of the RCN, and all four programmes predate the announcement of their centrality for the RRI

agenda in 2016. The SAMANSVAR programme seeks to foster interdisciplinary research with a focus on Corporate Social Responsibility and RRI. The RCN saw these four programs as arenas for experimentation and learning *in collaboration with the research environments financed through the programmes*. The RCN envisioned the RRI journey as one of mutual learning (Egeland et al., 2019). However, projects with an RRI element—however large or small—have been funded in other programmes too.

After an evaluation of the activities of the RCN, the programme structure was abandoned in favour of a portfolio structure in 2019. Three of the four programmes were joined under the portfolio of 'enabling technologies'—sometimes labelled 'converging technologies'—and the SAMANSVAR programme was incorporated in 2021. The main argument for the organisational change was to curb silo effects. Our informants do all agree with respect to the importance of the boards of the programmes—or later portfolios. These constitute important learning arenas for the RCN and provide opportunity to engage closer with researchers and experts from both Norway and abroad. However, our informants do not agree with respect to the merits of changing from a programme structure to a portfolio structure. Some seem to think that there were too many boards in the programme structure and that the portfolio structure enabled learning across fields. Others hold that the programme structure was essential in ensuring that participants had the necessary in-depth knowledge and language from the focal field. We return to later developments below.

3.1 The BIOTEK2021 programme: solving societal challenges in a responsible manner

The primary objective of BIOTEK2021 is to generate biotechnology that contributes to value creation and innovation in order to solve societal challenges in a responsible manner (Forskningsrådet, 2013). The secondary objectives are:

1. Develop the generic elements within biotechnology, thus enabling Norwegian research groups in academia and industry to compete at an international top level (i.e. Scientific Excellence)
2. Address the various needs and special features of each sector in a manner that activates synergies and fosters cooperation (i.e. Differentiation)
3. Ensure that support is provided to areas in which biotechnology is essential for value creation and industrial development that benefits the society (i.e. Innovation)
4. Ensure the responsible development of technology that addresses global societal challenges in the areas of health and sustainable food and industrial production (i.e. Societal challenges, RRI)
5. Establish conditions that promote cooperation, constructive task distribution and highly focused research activity within Norwegian biotechnology research (i.e. Collaboration)

6. Communicate with specified target groups to ensure that biotechnology research and development are in line with the societal needs (i.e. RRI)

Thus, in the BIOTEK2021 program, RRI was pointed out as integral to several secondary objectives. This is in keeping with the National Strategy for Biotechnology 2011–2020, which states that the ethical, legal, and other social aspects of biotechnological research and development activities need to be more integrated into projects and programmes (Norwegian Ministry of Education and Research 2012). In the BIOTEK2021 program it is mandatory for project proposals to have an RRI project component. Hence, the programme has a requirement for competence building with regards to RRI through activities in projects.

The programme has funded the Centre for Digital Life Norway (DLN), a large-scale network project with a focus on building competencies in interdisciplinary—or better transdisciplinary—research in the biotech sector (Centre for Digital Life Norway, 2024). DLN constitutes an important step in the promotion of RRI. All activities under the Digital Life initiative must be underpinned by the principle and practice of RRI. In addition, the centre offers advice on RRI to projects, including tools to facilitate RRI thinking. One example is the so-called 'walkshop', where participants hike in one of Norway's mountainous regions while contemplating and discussing the future of biotechnology (see Wickson et al., 2015).

The BIOTEK2021 programme was evaluated in 2017, simultaneously with the NANO2021 Programme. The overall conclusion is:

> There is no doubt that RCN and the BIOTEK2021 as well as the NANO2021 programme must be in the forefront internationally when it comes to the implementation of an RRI-perspective. [. . .]. The RRI-framework developed by RCN and particularly the DLN [Digital Life Norway] is an inspiration for other funding bodies across the world (Technopolis, 2017a, p. 55).

However, the BIOTEK2021 programme evaluation also discusses tensions and polarisation emanating from the programme:

> The external experts also note that the RRI theme seems to have created polarisation, requiring further efforts in this area. While the RCN's RRI framework is based on an integration model of the science and society relation, criticisms are based on a separation model. Subscribing to either one of these models is ultimately a political question. One way of dealing with such a conflict is to demonstrate how the RRI agenda can be useful to science. Another is to find ways of demonstrating that the RRI agenda is not something new, but is built on the responsibilities already exercised by scientists and takes its point of departure in what scientists already do (Technopolis, 2017a, p. 44).

The latter solution is recommended as the most respectful one. However, the tension captures well discussions about how RRI appears to have failed in securing a compelling sale pitch (Åm, 2019b; Ribeiro et al., 2017). The evaluation also elaborates on the capacity of the programme to spread a language related to RRI. For many applicants,

the notion of RRI was unknown before applying for funding. The report cites an applicant in this way to make the point:

> I was little aware of the RRI concept before the proposal process that led to the funded project. The process/project and its role in the DLN [Digital Life Norway] has led to a better understanding of the RRI concept, including revealing its chances, but also its challenges. (Funded project applicant) (Technopolis 2017a, p.35).

While the BIOTEK2021 programme is praised for its innovativeness, it is clear from the mentioned polarisation that acceptance of RRI as an institutionalised logic for biotech research and innovation was still a far cry at the time, despite any efforts of the programme (see Tolbert & Zucker, 1996). Some (successful) funding applicants relate to RRI as the brainchild of a particular clique and lament that how RRI can contribute to projects has not been explained in a convincing manner or sufficiently explained altogether. Other projects do in fact appear to engage with RRI and send their PhD candidates to the RRI school under the auspices of DLN (Technopolis, 2017a).

DLN represents a significant part of the total funding volume in the BIOTEK2021 programme. The report on the self-evaluation by DLN finds that there have been little to *"no structural changes to facilitate [. . .] the way research proposals are assessed. [T]hey are assessed ex ante, with scientific merit being evaluated separately from aspects like responsible innovation"* (Centre for Digital Life Norway & Research Council of Norway, 2019, p. 11). It also finds that RRI is often treated as a mere add-on rather than an integral part of research projects. Lamenting lack of integration is also traceable in the recommendations for future responsible innovation from the DLN self-evaluation (2019). It recommends internships in industry and changes to academic research environments; increased attention to end-users in the formulation and design of projects; regular use of sounding boards; and increased engagement from the perspective of the public good as a way to formulate research agendas. Following the path of RRI as a learning agenda, further learning was needed.

As is a central concern for the RCN, the 2020 *Technopolis* report, discussing DLN, recognises that *"RRI requires new skills for researchers, institutions need to adjust R&I governance structures, and target both processes and products of innovation"* (Varnai et al., 2020, p. 24). The report later laments that DLN appears to have no clear intervention logic while still recognising the substantial efforts of the DLN in crafting innovation and networks. An intervention logic can be modelled in a way where concrete efforts lead to specified outcomes through well-defined processes (Pawson & Tilly, 1997). However, some of our informants question the utility of such an intervention logic, either on the grounds of preferring a (mutual) learning agenda or by questioning the 'cultural fit' in a Norwegian setting (see e.g. Ansari et al., 2010). Their argument is then that such source-path-goal models capture messy—and often partly circular—learning processes poorly (see e.g. Kolb, 1984) and that Norwegian culture clashes with the inherent assumptions. The core of these assumptions is that it is opportune to specify the needed effort, the processes, and the goal(s) that need(s) to be achieved,

thus—potentially—stifling the free pursuit of the task as well as interpretations of what the processes and concrete task should be. Strong versions of such an approach may well inhibit learning and curb creativity (Svare et al., 2023). Some may—at an ideological level—link intervention logics to strong versions of New Public Management worldviews, though this link is not necessarily merited (Johnsen, 2024).

3.2 The NANO2021 programme: investing in research with a positive effect on societal development

The primary objective of the programme is to develop outstanding knowledge and sustainable solutions and innovations based on nanotechnology, microtechnology, and advanced materials to meet the needs of society at large (The Research Council of Norway, 2018). Secondary objectives are defined as five objectives. The programme is envisioned to:

1. Enhance innovation and national value creation based on the application of nanotechnology, microtechnology, and advanced materials
2. Enable Norwegian R&D groups to achieve a position in the international forefront and promote high quality in addition to scientific development and renewal
3. Promote responsible research and innovation in this technology field
4. Increase the attractiveness of Norwegian research groups to encourage national and international companies to establish R&D activities in Norway
5. Monitor and facilitating the use of Norway's membership of the European Synchrotron Radiation Facility (ESRF), including the Swiss Norwegian Beam Lines (SNBL) and the European Spallation Source (ESS), and work in general to promote optimal use of national infrastructure and expertise

A specific objective of the programme is thus to promote RRI in the field. Applicants to the NANO2021 programme are required to describe how relevant research questions will be addressed in relation to HSE (Health, Safety, and Environment), ELSA (Ethical, Legal, and Societal Aspects), and/or other RRI perspectives.

The RCN positions the programme as both a vehicle to fulfil the main strategy of the RCN at the time, termed 'Research for Innovation and Sustainability', and the 'Strategy for the Research Council of Norway for an innovative business sector 2016–2020'. The introductory summary thus states that:

> The programme will contribute to achieve the innovation strategy's objectives by investing in research with a positive effect on societal development, and by enabling companies to take advantage of the opportunities inherent in addressing societal challenges (The Research Council of Norway, 2018, p. 4).

Thus, societal development is seen as a business opportunity.

RRI can be both a research topic in its own right—for instance, assessing the impact of new technology on society—and a practice integrated into research projects. With respect to RRI as a thematic area for research, a priority is put on *"research that expands insight into the impacts of nanomaterials on human health and ecosystems"* (The Research Council of Norway, 2018, p. 7). With respect to RRI activities integral to projects, the program furthers four key elements of risk assessment:

1. Risks associated with the extraction/production of or limited access to raw materials
2. Risks during the research, scaling-up, and/or production phases
3. Risks during the user phase
4. Risks during the demolition/recycling phase

The target of the programme is to allocate 15% of the R&D budget to RRI activities. The RCN states that a joint call with other RCN programmes focused on ELSA projects has been instrumental in increasing the share of RRI in the programme. The programme also put emphasis on social dialogue and meeting places related to the technology in question.

As part of an evaluation of the NANO2021 programme, a survey was conducted that targeted project leaders and partners. This revealed a diversity of opinions, and several of the respondents found it *"difficult to express a clear view on the impact of RRI practices as a result of the NANO2021 programme"* (Technopolis, 2017b, p. 4). To the question about how the NANO2021 program had contributed to the spreading of knowledge or increased awareness about RRI, many of the respondents answered, 'Do not know' or 'Neither agree nor disagree' (Technopolis, 2017b, p. 35). The evaluation did, however, find that there was a high awareness of RRI among researchers within the field of nanotechnology, although they did not use specific RRI terminology. This phenomenon was confirmed by the surveys and interviews alike. Such results call into question if the RRI label is necessary in order to promote what Randles (2016) has termed 'de facto rri'. However, respondents largely agreed that having RRI as a prerequisite in the application was a good way to raise awareness about the terminology. Several of the respondents expressed positive views regarding how individual technology projects have benefitted from including an RRI component. Workshops dedicated to RRI were highlighted as a positive means for nanotechnology researchers to develop their own understanding and practices of RRI.

The evaluation report also discussed the implications of the RCN strategy of categorising RRI as one of several themes rather than as an overarching concern (like in the original IKTPLUSS programme—see section 3.3). In some instances, this causes RRI to be treated as a particular component or issue (such as Health, Safety, and Environment—HSE—and Ethical, Legal, and Societal Aspects—ELSA) organised in dedicated work packages and not as reflective processes, addressing all research aspects in the project, as envisioned in RRI. The evaluators emphasise that the RRI framework developed by RCN points to a transdisciplinary approach. They therefore argue that it

may be counter-productive to report RRI activities as one theme and by numbers (percentage of work effort) in favour of an integrated approach.

3.3 The IKTPLUSS programme: Promoting interdisciplinary research and innovation to build excellence

The IKTPLUSS programme has gone through various iterations. Here we focus on two of the early ones that show significant change. In the planning of the first version, an 'investment logic' is furthered, which is organised into four areas (Forskningsrådet, 2015, p. 2):

1. IKT-GRENSELAND [ICT-BORDERLAND ed.] will contribute to recruitment and stimulate more interdisciplinary groundbreaking ICT research in order to promote more cutting-edge innovations
2. IKT-FRONT [ICT-FRONT ed.] will conduct basic ICT research on relevant research topics and contribute to building capacity and robust research environments
3. IKT-VEKST [ICT-GROWTH ed.] will link research with application and contribute to growth and innovation in the ICT industry, other trade and industry, and the public sector
4. IKT-FYRTÅRN [ICT-LIGHTHOUSE ed.] will address important societal challenges by strengthening research efforts in areas where the utilisation of ICT and ICT research has a particularly high-value creation potential

The RCN as an actor in the research and innovation system is emphasised, including by reference to policy white papers, setting the agenda for Norwegian ICT research and innovation (St. Meld. 30, 2009). The programme plan states that:

> It emphasises the Research Council's role as strategic custodian of public research funding, and tries to balance this with the Research Council's shared responsibility for safeguarding society's and industry's need for ICT research (Forskningsrådet, 2015, pp. 2, our translation).

In addition, three cross-cutting fundamental dimensions are emphasised as being of strategic importance:

1. Co-creation
2. Responsible Innovation
3. Internationalisation

The IKTPLUSS programme was used in a joint call with SAMANSVAR to promote interdisciplinary research in ICT with strong RRI components in a 100-million NOK call. Six projects were funded through this joint call. The guidelines accompanying the call text explicitly place emphasis on the creation of mutual and adaptive learning processes and the involvement of stakeholders (The Research Council of Norway, 2015).

However, contrary to the other programmes, we detail in this chapter, the RRI aspects appear significantly toned down in the IKTPLUSS programme description. In the revised programme plan for 2018 and onwards, the phrasing of the programme became very focused on excellence in both research and innovation and further promoted an investment logic, with a heavy focus on radical innovation (as opposed to incremental innovation). The RRI perspective is not very salient in the revised programme, though it mentions a priority on societal challenges and makes reference to sustainability, as well as to the United Nations' sustainability goals (Forskningsrådet, 2018a). The three cross-cutting fundamental dimensions of Co-creation, Responsible Innovation, and Internationalisation are gone, as is any mention of responsible innovation altogether. Thus, the RRI perspective was effectively replaced by a sustainability agenda.

We strongly suspect this should be seen as an outcome of internal struggles over conceptualisation, as has been documented in other studies of expert bureaucracies (Heusinkveld & Benders, 2005). The divergent pattern is then an outcome of a 'cultural struggle' as signified by the larger RRI discourse and drive to instigate institutional changes in the innovation and research system. As more than one interviewee pointed out: *"The RCN is not a person, body or voice, one may work in the field in different ways and contribute with important things in various ways"*. The RCN consists of people. These people, in turn, have varying theoretical and ideological commitments. They are part of departments, which change over time, as does their enrolment in various departments, their managers, etc. Some leave the RCN, and new staff – and managers alike—bring new ideas as they engage with the ones already being discussed internally. Likewise, signals from the environment change and foster adjustments or entirely new ideas and vantage points, as well as new sources of legitimacy (Strang & Bradburn, 2001). The IKTPLUSS trajectory appears to be a good example of such dynamics. Such changes curb some collaboration and development while enabling other trajectories (Stjernberg & Philips, 1993). In addition, such changes have the potential to hamper innovation trajectories that are dependent on organisational actors, such as institutional entrepreneurs. Practices that are not taken for granted in organisations usually depend on actors to confirm their validity and legitimisation (Battilana et al., 2009).

To our knowledge, the IKTPLUSS programme has not been evaluated. An evaluation of Norwegian Technical Industrial Research Institutes laments that the programme—among others—does not seem to explicitly encourage co-operation between institutes (The Research Council of Norway, 2016, p. 50).

3.4 The SAMANSVAR programme: meeting global challenges through transdisciplinary research

The primary objective of the programme is to contribute to meeting the global challenges of society through responsible technology development and socially responsible business (Forskningsrådet, 2018b). Thus, RRI is more or less enshrined as the

primary objective of the programme. In addition, the programme is clearly stated as following up on the Rome Declaration on Responsible Research and Innovation (Council of the European Union, 2014). Likewise, RRI is featured as the means to the end of meeting the global challenges of society. Three secondary objectives are defined. The programme is envisioned to:

1. Build and strengthen research training and recruitment of young researchers to the fields.
2. Promote transdisciplinary research and co-production between research, innovation, and societal development.
3. Contribute to the development of the Research Council as a socially responsible actor and strengthen the Council's work to promote social responsibility in the research and innovation system.

Point 3 is stressed again when discussing the outcomes of the programme. Here the plan further underscores that: *"SAMANSVAR will also contribute to developing the role of the Research Council and other policy actors as social actors".* (Forskningsrådet, 2018b, pp. 4, our translation). Following this ambition and learning agenda, the programme was originally planned directly under the auspices of the innovation division at the RCN and then furnished with its own board. The idea was that the programme should (also) help the RCN learn about how to further interdisciplinary—or better transdisciplinary—research and do so by cutting across the other programme boards. To our knowledge, this arrangement was terminated, when the RCN reorganised to a portfolio model (see the introduction to section 3).

The programme is explicitly staged as a sequel to the ELSA programmes and as closely tied to the enabling technologies, such as biotech, nanotechnology, and ICT technology. In addition, the programme furthers research in the field of Corporate Social Responsibility (CSR) (see chapter 7). A core idea of the programme was thus to stimulate learning between researchers working with RRI and researchers working with CSR, sharing a concern for responsibility as an integral part of practice (see chapter 1).

In keeping with the understanding of the RCN as a shaper of research and innovation, the programme states that:

> The program is rooted in the emphasis on social responsibility and the Research Council's role as a social actor in the main strategy Research for innovation and sustainability. The program also follows up the Innovation Strategy and the Sustainability Strategy, where responsible research and innovation/RRI are highlighted. In the main strategy, the Council emphasizes that research and innovation must benefit society. This implies both that the research is carried out in a socially justifiable manner, and that the research contribution to meeting the major societal challenges must be emphasised (Forskningsrådet, 2018b, pp. 4, our translation).

The RCN points to two prioritised research areas and a development component in the thematic and professional priorities of the programme:

1. Research on responsible innovation
2. CSR research ("Corporate Social Responsibility")
3. Development of socially responsible research and innovation

In keeping with RRI thinking, extensive use of user participation in projects (see chapter 8) is called for, as well as attention to research ethics. The programme financed a virtual centre envisioned as a hub for competence-building on RRI and CSR in the Norwegian context, namely the AFINO. It thus also financed the book you are currently reading. AFINO is an acronym for Ansvarlig Forskning og Innovasjon i Norge (Responsible Research and Innovation in Norway).

Following up on the RCNs ambitions for the furtherance of RRI in Norway, AFINO has established a research school to:

> help researchers and practitioners address these questions [about RRI and CSR ed.] by being a reflexive platform where the RRI and CSR communities in Norway can learn, reflect, share experiences and discuss the above topics (Blanchard, n.d.; see also chapter 5).

As of present, no evaluation has been initiated for the SAMANSVAR programme. Forsberg et al. (2021) have, however, reflected on the role of SAMANSVAR in research connected to this program and identified a potential conflict between two distinct roles that RRI might have. On the one side, RRI might be seen as a discipline, and on the other side, RRI might be understood as a logic for organising research or for research policy.

The SAMANSVAR programme had its last call in 2018 and was mainly financed by pooling money from other programmes. To date, it is by far the largest programme dedicated to RRI. In 2015 there was a large call in collaboration with IKTPLUSS (see section 3.3 above). In 2021 it was incorporated into the strategic focus area 'enabling technologies', or 'converging technologies', which includes nanotech, biotech, and ICT.

3.5 Funding provided for RRI by the Research Council of Norway

Searching the project database of the RCN reveals that a total of 77 projects that applicants describe as related to RRI have been funded as of May 2024 with a total funding volume of about 67.9 million EUR (799.2 million NOK). The funding volume has shown a steady increase since 2012 and peaks in 2020 with 103.8 mill NOK. The 2023 volume is on par with the volume in 2017. The level for 2024 appears comparatively high, and the current level of funding as of May 2024 is only 10% shy of the total funding volume for 2023. However, since funding depends on applications submitted to calls, it is too early to gauge the development for the year 2024.

While this funding volume does not necessarily reflect projects which have RRI as a central concern, we suggest the steep trend towards the peak in 2020 highlights the RCN's growing interest in the RRI agenda. However, the total funding volume in 2020
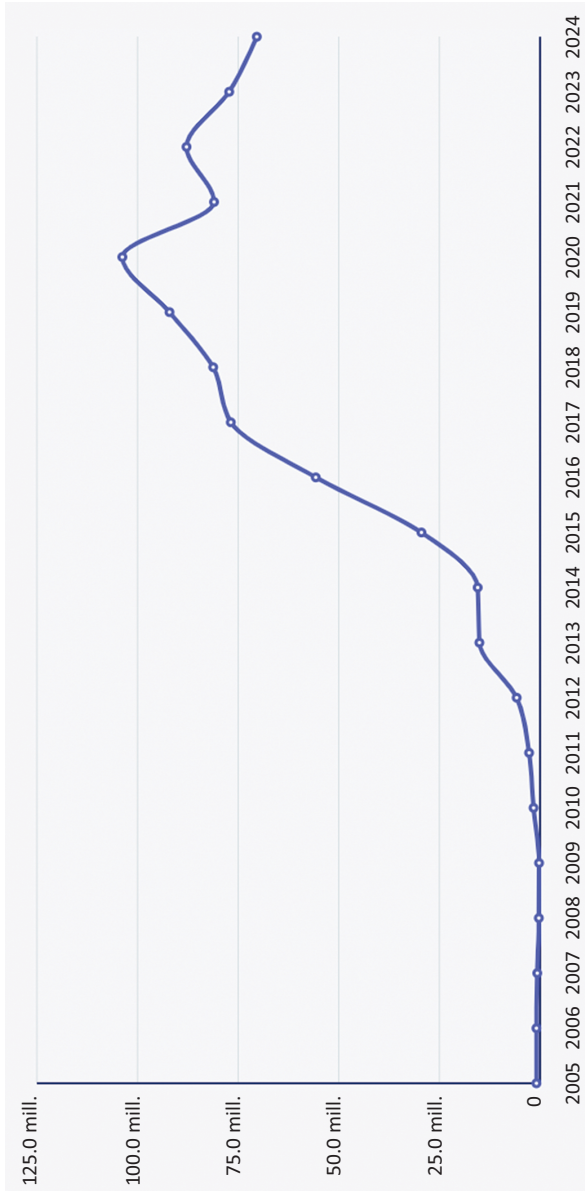
Figure 2: Funding volume of projects which applicants describe as related to RRI, funded by the RCN. (Source: The Research Council of Norway, n.d.), keyword search = 'RRI'.

was 10,400 million NOK, of which 103.8 million NOK were allocated to projects with an RRI element, representing about 1% of the total funding volume. In assessing this percentage, keep in mind that the RCN funds all types of research that usually take place in a country. Even if RRI has been supported and promoted very actively by the RCN, it has never been intended that all research that the RCN funds in its multiple programs should be versed using the RRI criterion. There is information in gauging the relationship between total funding and funding for RRI-related projects, though.

Table 4: The relative weight of funding allocated to RRI related projects by the RCN.

Year	Annual RCN funding volume in mill. NOK	Annual RRI related funding volume in mill. NOK	Percentage of funding volume allocated RRI related projects
2024	13,778.287	70.370	0.51
2023	11,299.665	77.227	0.68
2022	11,495.606	87.830	0.76
2021	11,496.160	80.984	0.70
2020	10,396.686	103.765	1.00
2019	10,056.693	92.052	0.92
2018	9,688.642.	81.197	0.84
2017	9,378.002	76.821	0.82
2016	8,599.939	55.716	0.65
2015	7,835.821	29.464	0.38
2014	6,971.393	15.493	0.22
2013	6,387.573	15.066	0.24
2012	6,163.346	5.804	0.09

The table shows the relative funding volume allocated to RRI related projects, compared to total funding volume administered annually by the RCN. Data compiled from (The Research Council of Norway, n.d.).

Table 4 reveals that the relative weight of funding provided for projects which find it relevant to mention RRI in either the title or project abstract in the project description and which got funded by the RCN shows a steep rise after 2012/2014, with a height in 2020, and a subsequent decline. The relative weight of projects funded after the peak in 2020 is lower than for the year immediately after the four research programmes described were framed as the standard-bearers of RRI research in the country (Gulbrandsen & Rynning, 2016). Thus, in both net and relative fiscal measures, the promotion of RRI by the RCN appears to be waning, with current levels trending below that of 2016 in relative measures. While such measures provide no information on the actual status of RRI use in RCN-funded projects, they do provide cues with respect to the significance of the RRI label in RCN-funded projects. However, a new focus on the ramifications of AI could change the balance again, if the RRI label is used (see section 4.2).

4 Recent developments at the Research Council of Norway

In the RCN strategy for 2020-2024, RRI is specifically promoted within the fields of technology and digitalisation, health and welfare, and as an integral part of the 'green transition' towards a more responsible and sustainable circular economy (The Research Council of Norway, 2020). In addition, one of the five targets of the strategy is 'Ethical and socially responsible research and innovation'. This target is described as *"research and innovation activities are carried out in compliance with recognised research ethics standards and within a socially responsible framework"* (The Research Council of Norway, 2020, p. 22). It is stated that such work includes anticipation of long-term consequences and potential unwanted effects from research undertakings.

Following the 'Long-term plan for research and higher education 2023-2032' issued by the Norwegian government (Meld. St. 5, 2022-2023), the RCN has published a revised strategy for 2024 (Forskningsrådet, n.d.). In this iteration, responsible innovation appears to have less prominence. However, in discussing how the RCN shall contribute to research-based knowledge necessary to meet grand challenges, it is stated that:

> The enabling technologies merge and provide new opportunities, with digitalisation as a key driving force. Mastery, responsible use and development of technologies are crucial for participation, value creation, competitiveness and welfare (Forskningsrådet, n.d., pp. 4, our translation).

With variations, the message is repeated in a section on the goal area 'Environmental, social and economic sustainability'. The document states that:

> Sustainable social development is largely about meeting today's need for just and responsible development within the tolerance limits of nature, ensuring equal opportunities for good health and quality of life, and maintaining an inclusive welfare society. New solutions are needed to rapidly reduce emissions and ensure sustainable management of the use of land and sea. New forms of cooperation across sectors must be developed to reduce inequality, counteract exclusion and maintain trust in democratic institutions (Forskningsrådet, n.d., pp. 11, our translation).

Thus, 'de facto rri' (Randles, 2016) appears still salient at the RCN, while the use of variations over the label or term 'RRI' has been toned down.

4.1 Upheaval, financial crisis, layoffs, and reorganisation

On 16 May 2022, the then-Norwegian minister for research and higher education, Ola Borten Moe, unseated the entire board of the RCN on allegations of financial misconduct at the Council. The issue was that the RCN, in the face of financial cuts in combination with demands of higher spending in order to reduce financial reserves (issued by the previous government), had engaged in auditing practices that were deemed unfit-

ting to budgeting rules in governmental organisations (Loge, 2022; Tonne et al., 2022; Trædal, 2022; Tveit, 2022). The practices had been pursued through many years and were—allegedly—well known both by the relevant public servants and previous ministers. The case caused major upheaval, both politically, in Norwegian research and higher education, and internally in the RCN. As the RCN suddenly was faced with a deficit of almost 3 billion NOK and overspending in the region of 3.5 billion, dramatic measures were needed (Javorovic, 2022; Loge, 2022; Regjeringen, 2022). The final solution had three major components: (1) a cut in research funding, as well as delaying calls and projects; 2) an extra 1.64 billion one-time appropriation to the RCN on the state budget; and 3) reorganisation and layoffs of staff at the RCN (Forskningsrådet, 2022, 2023). In the aftermath of the upheaval, the RCN was asked to save 62.8 million NOK (on top of an already installed cut of 170 million in the period 2017-2022), equalling 8.4% of the running costs (Forskningsrådet, 2023; NTB & Christensen, 2022). In addition, there was pressure to spend more money on research and less on administration. According to press coverage at the time, the cuts also came at a time when the RCN had already been stretched in terms of manpower for a while (Svarstad & Fanghol, 2022). More than 80 people accepted compensation packages to leave, and according to communication from the new board, the downsizing led to a 20% decrease in full-time equivalent (FTE) employees (Forskning.no, 2023; Forskningsrådet, 2023).

Our section 3.5 above, detailing funding allocated projects with an RRI component, shows that such appears unaffected by the financial cutbacks in 2022 due to the upheaval, which was confirmed by our informants. This, we suggest, is in itself a major achievement for the RRI cause at the RCN and points to a willingness to finance RRI-related research. Our informants generally find that the manning down did and does not have much effect on the possibility to keep promoting RRI at the RCN. Though not directly related to the upheaval at the RCN, the number of what we would call 'institutional entrepreneurs' most directly involved in the advocacy for RRI has been reduced from about four in 2016 to about two in 2024. However, as of today, the 'institutional entrepreneurs' report that they have wide support at the managerial level and thus appear successful in having built lasting alliances within the organisation. Likewise, the 'institutional entrepreneurs' emphasise the importance of the portfolio boards and their endorsement of, and interest in, RRI. RRI appears well beyond the ephemeral in the RCN, and RRI thinking is anchored beyond the team of institutional entrepreneurs. Likewise, the remaining 'souls of fire' appear highly motivated and still burning (see Stjernberg & Philips, 1993). Thus, if the team of institutional entrepreneurs have, in fact, succeeded in swaying the RCN for the RRI cause, the reduced group size could have little effect on the long-term commitment of the RCN to RRI. However, if RRI is not yet a taken-for-granted practice, the decimated group of institutional entrepreneurs could prove problematic in the future. Innovations such as the concept of RRI typically need continuous confirmation to have staying power, until it is widely adopted by a relevant group of users (Rogers, 2003). Such continuous confirmation is usually the work of a group of institutional entrepreneurs and not of

single individuals (Battilana et al., 2009). It is beyond our current study to gauge the extent to which RRI has been institutionalised at the RCN. By the word institutionalised, we do not allude to any office of RRI but think in terms of work by institutionalists like Selznick (1957) and Tolbert and Zucker (1996). Citing Tolbert and Zucker (1996), a recent review paper defines full institutionalisation as the situation where: *"[. . .] the practice is taken for granted, possessing a reality of its own, and organizations adopt the practice mostly due to the incomprehensibility of alternatives"* (Naumovska et al., 2021, p. 380). As our analysis suggests, alternatives were comprehensible when the IKTPLUSS programme was refurbished in 2018. Additionally, while the new RCN strategy is saturated with typical RRI concerns, the label 'RRI' is not used much.

4.2 New website and further work to promote RRI

However, recently, it appears there has been an uptick—or at least continuation—in efforts to promote RRI. First, a new website has been launched in December 2023. This website, currently entitled *"Responsible research and innovation as a method",* is to date the most comprehensive web presence of RRI at the RCN, furnished with tools and other material which helps researchers engage with RRI (The Research Council of Norway, 2024). This website is a result of learning through the years, including experience gained in international fora, not least with the development of guidelines in various ERA-NETs (see ERA4Health Partnership, n.d.; EuroNanoMed, n.d.; M-ERA.NET, n.d.), as well as an attempt to engage with some of the criticisms voiced by scientists in, for example, the evaluations of the BIOTEK2021 and NANOTEK2021 programmes, discussed in sections 3.1 and 3.2 (Technopolis, 2017a, 2017b).

The elaborate website is meant to support scientists in their ongoing engagement with RRI throughout a project, regardless of the type of project. Emphasis is put on supporting researchers with a practical guide and associated tools, making the RRI experience as concrete as possible. The website includes advice on how to include RRI in funding applications. Critics may say that the reflective emphasis has been toned down by these efforts to make RRI more concrete. However, keeping the critique of previous evaluations in mind, the website could turn out to be a credible response in the ongoing learning trajectory with RRI at the RCN.

Second, the RCN has had calls in 2022, 2023, and 2024 in the area of 'converging technologies' where RRI is an integral part. In the last two calls, the RCN has developed RRI supplements for the calls. While the large learning arenas AFINO and DLN are unlikely to be continued in their current forms, there has recently been a minor call for a significantly smaller learning arena focussing on 'converging technologies'. Likewise, the RCN has continued international collaboration tied to the ERA-NETs on RRI. Thus, the RCN continues to press for RRI, though calls dedicated to RRI have been scarce since 2018. Notably, a recent focus on AI—and ramifications from AI—issued by the government with earmarked funding for research in the region of 1,000 million NOK

(Kunnskapsdepartementet & Statsministerens kontor, 2023) provides ample opportunity to carry further the RRI torch. This area will most likely become a large field of research. In addition, it appears that ethics continue to be an important theme discussed internally at the RCN. Likewise, the RCN continues to host themed information and networking opportunities for parties interested in RRI.

5 Conclusion

In this chapter, we set out *to shed light on engagement with RRI in Norway and the RCN's involvement with RRI, undertaking an exploratory study with three foci.*

First, we showed that Norwegian interest in RRI is surprisingly strong and world leading in scholarly output. Norway is the tenth largest provider of RRI scholarship globally but the second largest when country size (inhabitants), the number of students in tertiary education, and/or the number of FTE in tertiary education is taken into account, in all cases only surpassed by the Netherlands. Norway thus appears exceptionally concerned with RRI—much more so than, for instance, the UK, United States, Austria, and Germany. Our ambition with this focus was to put our case, the RCN, in perspective.

Second, we highlighted in detail the four flagship funding programmes originally theorised as the standard bearers of RRI by the Council and discussed their evaluations. All but a later iteration of the IKTPLUSS programme has RRI as a salient component. We suggested the divergent pattern of the later IKTPLUSS programme discussed here is a sign of 'cultural struggles' internal to the RCN with regards to the role of the Council in the research system at the time. While the BIOTECH2021 and NANO2021 programmes are praised as world leading in evaluations, the assessments also highlight tensions between the intentions of the RCN and perceptions of scientists with regards to RRI, and not least the relevance of RRI to their conduct. Such concerns are also salient in the evaluations of the DLN—a large-scale initiative to further RRI in biotechnology. Turning to an assessment of funding for RRI-related projects, we showed that the trend is a step rise from 2014 to 2020 and a subsequent decline. The 2023 level is on par with the 2017 level in fiscal magnitude (without taking inflation into account) and with the 2016 level when considering the share of funding allocated to RRI-related projects, compared to the total funding volume. As significant funding of ICT projects with a focus on the potential ramifications of AI can be expected in the future, funding for RRI-related projects may again see growth, but it is unclear if the RRI label or term will be used.

Third, we proceeded to discuss the latest developments at the RCN, marred by public scandal and a financial crisis large enough to affect the fiscal budget of the Norwegian state and significant reductions in staff. The crisis, however, appears unrelated to the RRI agenda and any endorsement of RRI thinking. Likewise, funding for

RRI-related projects appears unaffected—at least at the time of the crisis. Hence, it appears that RRI thinking has survived significant organisational turmoil at the RCN. We discussed how the effect of a reduction of staff being directly involved with RRI would depend on the extent to which RRI is actually institutionalised as a taken-for-granted practice at the RCN and the potential need for reconfirmation of RRI as an innovative approach to science governance. Either way, it is the case that the RRI agenda is still alive at the RCN and that the RCN continues to further RRI in very visible ways. Notably, a webpage designed to help researchers in engaging with RRI has recently been launched. However, while 'de facto rri' thinking is saturating the new RCN strategy, the use of the RRI label has been toned down.

Further research emanating from the explorative study

Our exploratory bibliometric research detailing contributions from countries raises a number of questions with regard to the vast differences uncovered. While Norway stands out, certainly, the Netherlands is a beacon in terms of RRI research. These vast national differences deserve attention, as does the potential role of (national) research funders in the differences unearthed. We suggest comparative studies of countries and the funding available for RRI-related research in various countries as a first step. However, funding is only one way that a research area may be stimulated. Soft (and hard) governance as well as the general country climate is another (Wittrock et al., 2021). Though the vast fond of knowledge on RRI developed in Norway is impressive, the *Technopolis* (2017a, 2017b) evaluations caution us that we cannot take for granted that this knowledge is transformed into 'de facto rri' practices in Norwegian research projects. Such links—or lack thereof—deserve further attention. Organisation science scholars have long been concerned with the extent to which academic knowledge of management and organisation is useful to practice and may serve as inspiration for a deepened understanding within the field of RRI (e.g. Astley, 1985; Daft, 1980; Kieser & Leiner, 2009; Nohria & Eccles, 1997). Another thornier question is if the fond of knowledge produced reflects the ongoing learning—and possibly even change of competencies among researchers—that the RCN envisioned or may point in other directions. If these have indeed evolved over time, it should be possible to elicit changes using not only classic scientometric analysis (see e.g. Liu et al., 2022) but also probabilistic topic modelling (Blei, 2012) or a 'vocabularies' approach (e.g. Ocasio & Joseph, 2005).

The research on the RCN presented here provides just a starting point for assessing their role and impact on the RRI agenda in Norwegian research and innovation projects. We hope to have shown that the RCN is a case that merits much more attention than it has previously received and endorse further studies of funding providers who have sought to promote RRI in national science systems. We touched upon the notion of 'institutionalisation' as this is understood in institutional theory. While the institutionalisation of RRI has been studied at the EPSRC (Owen et al., 2021), conduct-

ing similar studies at the RCN and the Dutch NWO would significantly contribute to our understanding of the role of research funders in impacting the science and innovation system.

Our study does not say much about how policy signals regarding RRI emanating from the RCN through the four programmes and subsequent revisions are 'translated' into actual research practices and acquire meaning in actual research projects funded through the programmes. Studies of this nature do exist (e.g. Åm, 2019a, 2019b; Åm et al., 2021; Borch & Throne-Holst, 2021; Glerup et al., 2017; Solbu, 2021; Völker et al., 2023). However, we suggest comparative studies between programmes and/or countries, as well as a clearer picture of the mechanisms at play in such translation processes, drawing from extant (organisational) theory (e.g. Ansari et al., 2010; Callon, 1984; Carlile, 2004; Latour, 1987; Oliver, 1992; Røvik, 2016) could be promising in assessing the role of funding providers' ability to impact the research and innovation systems they support. The recent work by Völker et al. (2023) appears to be an important step in this direction.

This study regularly touches upon the function and limits of language. Importantly, the role of the RRI label, or in a term used by language theory scholars the 'signifier', vis a vis that which is 'signified', namely the actual practice of RRI or 'de facto rri', features often (see De Saussure, 2011). As Randles (2016) shows, the use of the RRI signifier does not necessarily cover (all) practices of RRI, as expressed by the term 'de facto rri'. If the greater project of the RRI movement is to further 'de facto rri' practices and not the use of the RRI signifier—or label—the relationship between word use and practice deserves more attention. Relatedly, this insight questions the utility of developing some specific RRI framework within the vocabulary of one profession, or discipline, and ask another to conform to this particular language use (see e.g. Mills, 1940). Such insights have significance for the question of research designs capable of capturing 'de facto rri' practices. Likewise, we know that any theorised concept—such as RRI—needs translation in order to be practiced somewhere (Gherardi & Nicolini, 2000), and that the 'transfer' to practice typically involves further theorisation by practitioners (Strang & Meyer, 1993). Therefore, studies detailing how the many discussions in the field about what RRI should—ideally—be influence practices of RRI, and translation processes towards that end, would be useful.

Finally, our study suggests that the RRI agenda survived financial scarcity and organisational turmoil at the RCN, including notable political upheaval. Relatedly, Pansera and Owen (2018) warned that Brexit and political turmoil in the UK could have adverse effects on the priorities of the EPSRC, whereas the follow-up study paints a more positive picture of the state of RRI institutionalisation at the EPSRC (Owen et al., 2021). Such findings call for longitudinal studies of funding providers. Similarly, the fact that only three large-scale national funding providers in Europe and some smaller ones appear to have wholeheartedly embraced RRI begs the question: why these? Detailed accounts of their development trajectories and comparative studies will likely help us better understand how ideas salient to RRI may be furthered through funders.

Such combined efforts may allow us to specify more precisely the pathways to transformation through 'programme theory' which focus on mechanisms at play, taking contextual factors into account (e.g. Pawson, 2013; Pawson & Tilly, 1997), and thus provide realistic theories of change, given salient features of relevant contexts (for example relevant features of the country profile, field of research/innovation, professional vocabularies, etc.).

Empirically sustained studies, capable of eliciting mechanisms with a reach beyond the singular case or organisation, do not appear widespread in RRI research (Klaassen et al., 2018). For instance—building on our material in this chapter—even if the Netherlands, Norway, the UK, and in some cases Austria and Belgium appear highly concerned with RRI, judging from publication data, we lack clear assessments of whether in those countries that fact translates into a higher proportion of research and innovation conforming to 'de facto rri' principles. Likewise, it is not clear to which extent the strategy of the EC in funding an army of researchers to promote RRI through the Horizon 2020 Science with and for Society (SwafS) programme actually resulted in RRI becoming widely accepted as an approach to research and innovation in the EU (Delaney & Iagher, 2020; Wicher & Frankus, 2023). Put differently, even if RRI envisions a change in the science and innovation system globally (von Schomberg, 2013), research in the field has—to our knowledge—not yet built a convincing model to assess pathways of transformation that is based on a systems perspective. We suggest that such work could be enhanced by including further research traditions, which are traditionally concerned with questions about how change may be obtained, such as organisational science in its broadest sense, into our toolbox for the assessment of RRI diffusion, translation, and implementation.[3] Thus, over time, the field of RRI could build genuinely interdisciplinary—or transdisciplinary—knowledge on orchestrated change in the science and innovation system.

---

3 This suggestion should not be seen as an endorsement of 'managerialism' in science or higher education institutions (see, e.g., Deem & Brehony, 2005).

References

Åm, H. (2019a). Ethics as ritual: smoothing over moments of dislocation in biomedicine. *Sociology of Health & Illness*, *41*(3), 455–469. https://doi.org/10.1111/1467-9566.12818

Åm, H. (2019b). Limits of decentered governance in science-society policies. *Journal of Responsible Innovation*, *6*(2), 163–178. https://doi.org/10.1080/23299460.2019.1605483

Åm, H., Solbu, G., & Sørensen, K. H. (2021). The imagined scientist of science governance. *Social studies of science*, *51*(2), 277–297. https://doi.org/10.1177/0306312720962573

Ansari, S. M., Fiss, P. C., & Zajac, E. J. (2010). Made to Fit: How Practices Vary as They Diffuse. *Academy of management review*, *35*(1), 67–92. https://doi.org/10.5465/amr.35.1.zok67

Astley, W. G. (1985). Administrative Science as Socially Constructed Truth. *Administrative Science Quarterly*, *30*(4), 497–513. https://doi.org/10.2307/2392694

Battilana, J., Leca, B., & Boxenbaum, E. (2009). How Actors Change Institutions: Towards a Theory of Institutional Entrepreneurship. *Academy of Management Annals*, *3*(1), 65–107. https://doi.org/10.5465/19416520903053598

Benders, J., Nijholt, J., & Heusinkveld, S. (2007). Using print media indicators in management fashion research [Article]. *Quality & Quantity*, *41*(6), 815–829. https://doi.org/10.1007/s11135-006-9027-5

Blanchard, A. (n.d., n.d.). *Welcome to the AFINO Research School*. Norwegian University of Science and Technology. Retrieved 29th March from https://www.ntnu.edu/afino/the-afino-research-school

Blei, D. M. (2012). Probabilistic topic models. *Commun. ACM*, *55*, 77–84. https://doi.org/10.1145/2133806.2133826

Borch, A., & Throne-Holst, H. (2021). Does Proof of Concept Trump All? RRI Dilemmas in Research Practices. *Science and Engineering Ethics*, *27*(1), 7. https://doi.org/10.1007/s11948-021-00288-8

Callon, M. (1984). Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. *The Sociological Review*, *32*(S1), 196–233.

Carlile, P. R. (2004). Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries. *Organization Science*, *15*(5), 555–568.

Centre for Digital Life Norway. (2024, n.d.). *A national centre for biotechnology research and innovation* Centre for Digital Life Norway. Retrieved 6th April from https://www.digitallifenorway.org/

Centre for Digital Life Norway, & Research Council of Norway. (2019). *Report on the Digital Life Norway egenvurdering seminar*. https://digitallifenorway.org/media/fm/d1a691f938.pdf

Clark, T. (2004). The Fashion of Management Fashion: A Surge Too Far? *The Interdisciplinary Journal of Organization, Theory and Society*, *11*(2), 297–306. https://doi.org/10.1177/1350508404030659

Council of the European Union. (2014). *Rome Declaration on Responsible Research and Innovation in Europe*. (806). European Commision Retrieved from https://digital-strategy.ec.europa.eu/en/library/rome-declaration-responsible-research-and-innovation-europe

Daft, R. L. (1980). The Evolution of Organization Analysis in ASQ, 1959-1979. *Administrative Science Quarterly*, *25*(4), 623–636. https://doi.org/10.2307/2392284

Daimer, S., Berghäuser, H., & Lindner, R. (2023). The institutionalisation of a new paradigm at policy level. In V. Blok (Ed.), *Putting responsible research and innovation into practice: A Multi-Stakeholder approach* (Vol. 40, pp. 35–56). Springer International Publishing. https://doi.org/10.1007/978-3-031-14710-4

De Saussure, F. (2011). *Course in general linguistics*. Columbia University Press.

Deem, R., & Brehony, K. J. (2005). Management as ideology: the case of 'new managerialism' in higher education. *Oxford Review of Education*, *31*(2), 217–235. https://doi.org/10.1080/03054980500117827

Delaney, N., & Iagher, R. (2020). *Institutional changes towards responsible research and innovation: Achievements in Horizon 2020 and recommendations on the way forward* (KI-02–20-259-EN-N). The European Comission. https://op.europa.eu/en/publication-detail/-/publication/582ef256-cbcc-11ea-adf7-01aa75ed71a1

Doezema, T., Ludwig, D., Macnaghten, P., Shelley-Egan, C., & Forsberg, E.-M. (2019). Translation, transduction, and transformation: expanding practices of responsibility across borders. *Journal of Responsible Innovation*, *6*(3), 323–331. https://doi.org/10.1080/23299460.2019.1653155

Egeland, C., Forsberg, E.-M., & Maximova-Mentzoni, T. (2019). RRI: implementation as learning. *Journal of Responsible Innovation*, *6*(3), 375–380. https://doi.org/10.1080/23299460.2019.1603570

Egeland, C., Maximova-Mentzoni, T., Hanssen, A. B., & Forsberg, E.-M. (2018). *RRI-Practice Report from national case study: Norway*, *Deliverable 3.1*.

ERA4Health Partnership. (n.d.). *Responsible Research and Innovation (RRI)*. Instituto de Salud Carlos III (ISCIII). Retrieved 5th May from https://era4health.eu/responsible-research-and-innovation-rri/

EuroNanoMed. (n.d.). *Responsible Research and Innovation (RRI) in proposals to EuroNanoMed III*. Instituto de Salud Carlos III (ISCIII). Retrieved 5th May from https://euronanomed.net/joint-calls/enmiii-rri-guidelines/

Forsberg, E.-M., Thorstensen, E., Dias Casagrande, F., Holthe, T., Halvorsrud, L., Lund, A., & Zouganeli, E. (2021). Is RRI a new R&I logic? A reflection from an integrated RRI project. *Journal of Responsible Technology*, *5*, 100007. https://doi.org/10.1016/j.jrt.2020.100007

Forskning.no. (2023, 10th January 2023, 6:11PM). *87 har søkt om sluttpakke i Forskningsrådet*. Forskning.no. Retrieved 3rd April from https://www.forskning.no/forskningspolitikk-ntb-om-forskning/87-har-sokt-om-sluttpakke-i-forskningsradet/2138642

Forskningsrådet. (2013). *Programplan for BIOTEK2021 (2012–2021)*: *Bioteknologi for verdiskaping – BIOTEK2021*. Oslo S: Norges Forskningsråd Retrieved from https://www.forskningsradet.no/siteas sets/publikasjoner/1253984383600.pdf

Forskningsrådet. (2015). *IKTPLUSS Plan for satsningen*. Oslo, NO: Norges forskningsråd Retrieved from unavailable

Forskningsrådet. (2018a). *Programplan Gjelder fra 2018: IKT og digital innovasjon – IKTPLUSS*. (978–82-12-03684-0). Oslo, NO: Norges forskningsråd Retrieved from https://www.forskningsradet.no/siteas sets/publikasjoner/1254034345381.pdf

Forskningsrådet. (2018b). *Programplan Revidert 2018: Ansvarlig innovasjon og bedriftenes samfunnsansvar (SAMANSVAR)*. (978–82-12-03733-5). Oslo, NO: Norges forskningsråd Retrieved from https://www.for skningsradet.no/siteassets/publikasjoner/1254037286695.pdf

Forskningsrådet. (2022, 6th February 2023). *Informasjon om den økonomiske situasjonen i Forskningsrådet*. Forskningsrådet. Retrieved 3rd April from https://www.forskningsradet.no/nyheter/2022/informas jon-om-den-okonomiske-situasjonen-i-forskningsradet/

Forskningsrådet. (2023). *En vanskelig økonomisk situasjon er løst, men omstillingsarbeidet pågår fortsatt*. Lysaker: Forskningsrådet Retrieved from https://www.forskningsradet.no/siteassets/om-forskningsradet/styredokumenter/melding-fra-styret–juni-2023.pdf

Forskningsrådet. (n.d.). *Forskningsrådets strategi*. Lysaker, NO: Norges forskningsråd Retrieved from https://www.forskningsradet.no/siteassets/publikasjoner/2024/forskningsradets-strategi.pdf

Gherardi, S., & Nicolini, D. (2000). To Transfer is to Transform: The Circulation of Safety Knowledge. *Organization*, *7*(2), 329–348. https://doi.org/10.1177/135050840072008

Glerup, C., Davies, S. R., & Horst, M. (2017). 'Nothing really responsible goes on here': scientists' experience and practice of responsibility. *Journal of Responsible Innovation*, *4*(3), 319–336. https://doi.org/10.1080/23299460.2017.1378462

Gulbrandsen, E., & Rynning, H. (2016). Ansvarlig forskning og innovasjon i Norges forskningsråd. *Forskningspolitikk*, *38*(1), 26–27. http://fpol.no/wp-content/uploads/2016/03/Forskningspolitikk-1-2016_96-dpi.pdf

Hellstrom, T. (2003). Systemic innovation and risk: technology assessment and the challenge of responsible innovation. *Technology in Society*, *25*(3), 369–384. https://doi.org/10.1016/S0160-791X(03)00041-1

Heusinkveld, H. S., & Benders, J. G. J. M. (2005). Contested Commodification: Consultancies and their Struggle with New Concept Development. *Human Relations*, *58*(3), 283–310. https://doi.org/10.1177/0018726705053423

Javorovic, B. E. (2022). *Kaster styret i Forskningsrådet på grunn av «alvorlig økonomisk situasjon»*. Universitetsavisen.no. Retrieved 3rd April from https://www.universitetsavisa.no/forskningsradet-kunnskapsdepartementet-ola-borten-moe/kaster-styret-i-forskningsradet-pa-grunn-av-alvorlig-okonomisk-situasjon/362796

Johnsen, Å. (2024). Trust-based Management in the City of Oslo: Implementing a New Managerial Super Standard? In I. Lapsley & P. Miller (Eds.), *The Resilience of New Public Management* (pp. 243–263). Oxford University Press. https://doi.org/10.1093/oso/9780198883814.003.0011

Kieser, A., & Leiner, L. (2009). Why the Rigour-Relevance Gap in Management Research Is Unbridgeable. *Journal of Management Studies*, *46*(3), 516–533. https://doi.org/10.1111/j.1467-6486.2009.00831.x

Klaassen, P., Rijnen, M., Vermeulen, S., Kupper, F., & Broerse, J. (2018). Technocracy versus experimental learning in RRI: On making the most of RRI's interpretative flexibility. In *Responsible Research and Innovation* (pp. 77–98). Routledge.

Kolb, D. A. (1984). *Experiential learning: experience as the source of learning and development*. Prentice-Hall.

Kunnskapsdepartementet, & Statsministerens kontor. (2023, 7th September 2023). *Regjeringen med milliardsatsing på kunstig intelligens*. Regjeringen.no. Retrieved May 4th from https://www.regjeringen.no/no/aktuelt/regjeringen-med-milliardsatsing-pa-kunstig-intelligens/id2993214/

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

Liu, J., Zhang, G., Lv, X., & Li, J. (2022). Discovering the Landscape and Evolution of Responsible Research and Innovation (RRI): Science Mapping Based on Bibliometric Analysis. *Sustainability*, *14*(14), 8944. https://www.mdpi.com/2071-1050/14/14/8944

Loge, J. (2022, 16th May 1:550 PM). *Foreslår store kutt i Forskningsrådet: Vil stoppe utlysninger, kutte i prosjekter og avslutte sentre*. Forskerforbundet. Retrieved 3rd April from https://www.forskerforum.no/foreslar-store-kutt-i-forskningsradet-vil-stoppe-utlysninger-kutte-i-prosjekter-og-avslutte-sentre/

M-ERA.NET. (n.d., n.d.). *Responsible Research and Innovation (RRI)*. M-ERA.NET. Retrieved 5th May from https://www.m-era.net/other-joint-activities/responsible-research-and-innovation

Macnaghten, P., & Owen, R. (2011). Good governance for geoengineering [Comment]. *Nature*, *479*, 293. https://doi.org/10.1038/479293a

Mills, C. W. (1940). Situated Actions and Vocabularies of Motive. *American Sociological Review*, *5*(6), 904–913. https://doi.org/10.2307/2084524

Naumovska, I., Gaba, V., & Greve, H. R. (2021). The Diffusion of Differences: A Review and Reorientation of 20 Years of Diffusion Research. *The Academy of Management Annals*, *15*(2), 377–405. https://doi.org/10.5465/annals.2019.0102

Neresini, F., & Arnaldi, S. (2018). *RRI-Practice Report from national case study: Italy*, *Deliverable 7.1*.

Nohria, N., & Eccles, R. G. (1997). Where does Management Knowledge come from? In J. L. Alvarez (Ed.), *The diffusion and consumption of business knowledge* (pp. 278–304). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-25899-4

NTB, & Christensen, L. (2022, 7th October 12:4 PM). *Forskningsrådet må nedbemanne*. Forskerforbundet. Retrieved 3rd April from https://www.forskerforum.no/forskningsradet-ma-nedbemanne/

Ocasio, W., & Joseph, J. (2005). Cultural adaptation and institutional change: The evolution of vocabularies of corporate governance, 1972–2003. *Poetics*, *33*(3), 163–178. https://doi.org/10.1016/j.poetic.2005.10.001

OECD. (2021). *Statitstics: Educational personnel by type of institution, management personnel and teacher aides*. The Organisation for Economic Co-operation and Development (OECD). Retrieved 30th March from https://stats.oecd.org/?_ga=2.152785146.273135830.1687262708-1514995489.1687262708

Oliver, C. (1992). The Antecedents of Deinstitutionalization. *Organization Studies*, *13*(4), 563–588. https://doi.org/10.1177/017084069201300403

Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy*, *39*(6), 751–760. https://doi.org/10.1093/scipol/scs093

Owen, R., Pansera, M., Macnaghten, P., & Randles, S. (2021). Organisational institutionalisation of responsible innovation. *Research Policy*, *50*(1), 104132. https://doi.org/10.1016/j.respol.2020.104132

Pansera, M., & Owen, R. (2018). *RRI-Practice Report from national case study: United Kingdom*, *Deliverable 5.1*.

Pawson, R. (2013). *The science of evaluation: a realist manifesto*. Sage.

Pawson, R., & Tilly, N. (1997). *Realistic evaluation*. Sage.

Randles, S. (2016). *Deepening 'Deep Institutionalisation:' Elaborating a Concept and Developing a Typology to Analyse and Contrast the Institutionalization of De-facto responsible research and innovation (rri); and H2020 RRI in Research and Technology Organisations (RTOs)* (Joining Efforts for Responsible Research and Innovation (JERRI), Issue D1.2). Manchester_Metropolitan_University. https://www.jerri-project.eu/jerri-wAssets/docs/deliverables/wp-1/JERRI_Deliverable_D1_2_Deepening-Deep-Institutionalisation.pdf

Randles, S., Tancoigne, E., & Joly, P.-B. (2022). Two tribes or more? The historical emergence of discourse coalitions of responsible research and innovation (rri) and Responsible Research and Innovation (RRI). *Journal of Responsible Innovation*, *9*(2), 248–274. https://doi.org/10.1080/23299460.2022.2061306

Regjeringen. (2022). *Kraftfullt engangsgrep sikrer norsk forskning*. Oslo, NO: Regjeringen.no Retrieved from https://www.regjeringen.no/no/aktuelt/kraftfullt-engangsgrep-sikrer-norsk-forskning/id2948536/

Ribeiro, B. E., Smith, R. D. J., & Millar, K. (2017). A Mobilising Concept? Unpacking Academic Representations of Responsible Research and Innovation. *Science and Engineering Ethics*, *23*(1), 81–103. https://doi.org/10.1007/s11948-016-9761-6

Rogers, E. M. (2003). *Diffusion of Innovations* (5th (pbk.) ed.). Free Press. https://www.simonandschuster.com/books/Diffusion-of-Innovations-5th-Edition/Everett-M-Rogers/9780743258234

Røvik, K. A. (2016). Knowledge Transfer as Translation: Review and Elements of an Instrumental Theory. *International Journal of Management Reviews*, *18*(3), 290–310. https://doi.org/10.1111/ijmr.12097

Ryan, T. K., Mejlgaard, N., & Degn, L. (2021). Organizational patterns of RRI: how organizational properties relate to RRI implementation. *Journal of Responsible Innovation*, *8*(2), 320–337. https://doi.org/10.1080/23299460.2021.1975376

Selznick, P. (1957). *Leadership in administration: A sociological interpretation*. University of California Press. https://www.ucpress.edu/book/9780520049949/leadership-in-administration

Solbu, G. (2021). Frictions in the bioeconomy? A case study of policy translations and innovation practices. *Science and Public Policy*, *48*(6), 911–920. https://doi.org/10.1093/scipol/scab068

Solli, M. (2023). Stakeholder Inclusion as the Research Council of Norway's Silver Bullet. *Etikk i praksis-Nordic Journal of Applied Ethics*(1), 83–97. https://doi.org/10.5324/eip.v17i1.5043

St. Meld. 5. (2022-2023). *Long-term plan for research and higher education 2023–2032*. (5). Oslo, NO: Norwegian Ministry of Climate and Environment Retrieved from https://www.regjeringen.no/contentassets/9531df97616e4d8eabd7a820ba5380a9/en-gb/pdfs/stm202220230005000engpdfs.pdf

St. Meld. 30. (2009). *Klima for forskning*. (30). Oslo, NO: Det Kongelige Kunnskapsdepartementet Retrieved from https://www.regjeringen.no/contentassets/057291a3e4bf4825b23e48edda46a32c/no/pdfs/stm200820090030000dddpdfs.pdf

Statistisk sentralbyrå. (2024, 21st February 2024). *Population*. Statistisk sentralbyrå. Retrieved 5th April from https://www.ssb.no/en/befolkning/folketall/statistikk/befolkning

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568-1580. https://doi.org/10.1016/j.respol.2013.05.008

Stjernberg, T., & Philips, A. (1993). Organizational innovations in a long-term perspective: Legitimacy and souls-of-fire as critical factors of change and viability. *Human Relations*, *46*(10), 1193. https://doi.org/10.1177/001872679304601003

Strang, D., & Bradburn, E. M. (2001). Theorizing Legitimacy Or Legitimating Theory?: Competing Institutional Accounts of HMO Policy, 1970-1989. In J. L. Campell & O. K. Pedersen (Eds.), *The Rise of Neoliberaalism and Institutional Analysis* (pp. 129–158). Princeton University Press. https://doi.org/10.1515/9780691188225-010

Strang, D., & Meyer, J. W. (1993). Institutional Conditions for Diffusion. *Theory and Society*, *22*(4), 487–511. https://doi.org/10.1007/Bf00993595

Strang, D., & Sine, W. D. (2002). Inter-Organizational Institutions. In J. A. C. Baum (Ed.), *The Blackwell Companion to Organizations* (pp. 497–519). Blackwell Publishers Ltd. https://doi.org/10.1002/9781405164061.ch21

Strang, D., & Wittrock, C. (2019). Methods for the Study of Management Ideas. In A. Sturdy, H. S. Heusinkveld, T. Rey, & D. Strang (Eds.), *The Oxford Handbook of Management Ideas*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198794219.013.8

Svare, H., Johnsen, Å., & Wittrock, C. (2023). Does trust-based management reform enhance employee-driven innovation? Evidence from a Scandinavian capital. *Nordic Journal of Innovation in the Public Sector*, *2*(1), 43–58. https://doi.org/10.18261/njips.2.1.4

Svarstad, J., & Fanghol, T. A. (2022, 6th October, 5:55 PM). *Forskningsrådet må nedbemanne:—En tøff og krevende situasjon*. Khrono. Retrieved 3rd April from https://www.khrono.no/forskningsradet-ma-nedbemanne-en-toff-og-krevende-situasjon/723143

Technopolis. (2017a). *Evaluation of the RCN's BIOTEK2021 programme. Final report*. https://www.forskningsradet.no/contentassets/ea9db93ed204480fa15a61682ed66d9e/evaluering_av_biotek2021.pdf

Technopolis. (2017b). *Evaluation of the RCN's NANO2021 programme, Final report*. https://www.forskningsradet.no/contentassets/dc0a2485b388440dadcb0c7a47f042a3/evaluering_av_nano2021.pdf

The Research Council of Norway. (2015). *Guidelines and important considerations relevant to all types of applications in this call for proposals*. Oslo, NO: The Research Council of Norway

The Research Council of Norway. (2016). *Evaluation of Norwegian Technical Industrial Research Institutes*. Oslo, NO: The Research Council of Norway Retrieved from https://www.forskningsradet.no/siteassets/publikasjoner/1254019976134.pdf

The Research Council of Norway. (2018). *Work programme applicable from 2018: Nanotechnology, Microtechnology and Advanced Materials – NANO2021*. (ISBN 978–82-12-03681-9). Oslo, NO: The Research Council of Norway Retrieved from https://www.forskningsradet.no/om-forskningsradet/publikasjoner/2018/work-programme–applicable-from-2018/

The Research Council of Norway. (2020). *Empowering ideas for a better world Strategy for the Research Council of Norway 2020—2024*. (978–82-12-03858-5). Lysaker, NO: The Research Council of Norway Retrieved from https://www.forskningsradet.no/contentassets/0b1fe1ef42e7435fa7d37e3a4d7dfdc6/strategi_2020-2024_en.pdf

The Research Council of Norway. (2024, 26th April 2024). *Responsible research and innovation as a method*. The Research Council of Norway. Retrieved 1st May from https://www.forskningsradet.no/en/research-policy-strategy/rri/

The Research Council of Norway. (n.d.). *Project Bank*. The Research Council of Norway. Retrieved 1st May from https://prosjektbanken.forskningsradet.no/en/explore/statistics

Tolbert, P. S., & Zucker, L. G. (1996). The Institutionalization of Institutional Theory. In S. R. Clegg, C. Hardy, & W. R. Nord (Eds.), *Handbook of Organization Studies* (1st ed., pp. 175–190). SAGE. https://doi.org/10.4135/9781446218556.n6

Tonne, H., Ruud, K., Aronsen, J. M., Aamot, E., Melleby, K., Hansen, M. H., Heggernes, P., Hjemdahl, K., Kåberger, T., Nygaard, E., & Vorland, L. H. (2022, 24th May 2022, 9:20 AM). *En oppkonstruert krise i*

Forskningsrådet. Khrono. Retrieved 3rd April from https://www.khrono.no/en-oppkonstruert-krise-i-forskningsradet/690004

Trædal, T. J. (2022, 20th May, 11:09 AM). *Den forrige regjeringen kuttet budsjettet, men ba Forskningsrådet fortsette å betale ut på samme nivå*. Aftenposten. Retrieved 3rd April from https://www.aftenposten.no/norge/i/XqkagB/den-forrige-regjeringen-kuttet-budsjettet-men-ba-forskningsraadet-fortsette-aa-betale-ut-paa-samme-nivaa

Tveit, M. S. (2022). *Styremøte 5/22, 27. mai – Innkalling* Lysaker: Forskningsrådet Retrieved from https://www.forskningsradet.no/siteassets/om-forskningsradet/styredokumenter/saksdokumenter-til-styremote-27.05.22.pdf

van der Molen, F., Consoli, L., Ludwig, D., Pols, A., & Macnaghten, P. (2018). *RRI-Practice Report from national case study: The Netherlands*, *Deliverable 9.1*.

Varnai, P., Arnold, E., Rentel, M., King, R., & Melin, G. (2020). *Digital Life Norway: AS IS Report* (1). Technopolis Group. https://www.digitallifenorway.org/news/digital-life-norway-as-is_report-18-june-2020.pdf

Völker, T., Slaattelid, R., & Strand, R. (2023). *Translations of Responsibility: Innovation Governance in Three European Regions* (1st ed.). Routledge. https://doi.org/10.4324/9781003371229

von Schomberg, R. (2013). A Vision of Responsible Research and Innovation. In R. Owen, J. Bessant, & H. Maggy (Eds.), *Responsible innovation: managing the responsible emergence of science and innovation in society* (pp. 51–74). John Wiley & Sons. https://doi.org/10.1002/9781118551424.ch3

Wicher, M. J., & Frankus, E. (2023). Research governance for change: funding project-based measures in the field of responsible research and innovation (RRI) and their potential for organisational learning. *The Learning Organization*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/TLO-10-2021-0119

Wickson, F., Strand, R., & Kjølberg, K. L. (2015). The Walkshop Approach to Science and Technology Ethics. *Science and Engineering Ethics*, *21*(1), 241–264. https://doi.org/10.1007/s11948-014-9526-z

Wittrock, C., & Forsberg, E.-M. (2019). *Handbook for Organisations Aimed at Strengthening Responsible Research and Innovation* (RRI-Practice project report, Issue D17.6). Responsible_Research_and_Innovation_in_Practice. https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5c4ef96d9&appId=PPGMS

Wittrock, C., Forsberg, E.-M., Pols, A., Macnaghten, P., & Ludwig, D. (2021). *Implementing Responsible Research and Innovation: National and Organisational Conditions*. Springer Nature. https://doi.org/10.1007/978-3-030-54286-3

World Bank. (2018). *Education Statisitcs Indicators Report*. The World Bank. Retrieved 30th March from https://databank.worldbank.org/indicator/SE.TER.ENRL?id=c755d342&report_name=EdStats_Indicators_Report&populartype=series

Zupic, I., & Čater, T. (2015). Bibliometric Methods in Management and Organization. *Organizational Research Methods*, *18*(3), 429–472. https://doi.org/10.1177/1094428114562629

Section 2: **Contexts of fragile responsibility**

Giovanni De Grandis

Chapter 4
The elusive transformation of research and innovation. The overlooked complexities of value alignment and joint responsibility

RRI is a broad concept that is subject to different interpretations. This chapter focuses on the view of RRI as a transformative ideal for reforming the research and innovation system in the service of public interest. This is the normatively strong view of RRI that has attracted many policy-makers and young researchers but left cold many senior researchers and innovators. The transformative vision of RRI has failed to materialise, and RRI remains a marginal reality, even in Norway, where arguably the conditions were more propitious than elsewhere.

I attempt to explain the failed transformation, focusing on two key objectives of RRI: the ambition of aligning R&I with societal values and the aspiration to steer R&I through generating a shared responsibility for the future (a prospective joint responsibility in technical terms). Alignment proved very hard to achieve because valuing is steeped in well-established practices, habits, and cognitive-emotional frameworks. These cannot be changed at will. Besides, the ambition to make researchers and innovators more responsive to a wider constituency and additional social responsibilities stumbled against what I call moral saturation, namely the lack of capacity and resources to take on additional moral tasks. Furthermore, modern societies are characterised by a pluralism of values and conceptions of the good life, compounded by the lack of methods for composing value conflicts. These problems come back when we look at what it takes to create expanded shared responsibilities: joint intentions and joint commitments. A formal analysis of how these can be generated shows that they need pre-conditions that are seldom obtaining in the real world of R&I.

So, the transformative ambition went against the inertia and entrenched habits of the R&I ecosystem and yet very often the task of promoting the change was given to junior researchers: the most vulnerable and less powerful actors. I conclude that RRI ambition to transform the R&I system is unrealistic. What can be attempted is to develop small-scale experiments outside the mainstream, where institutional barriers and perverse incentives are partially removed or corrected.

1 Introduction

1.1 The personal and the philosophical

This chapter stems from my own experience with RRI, mainly in the AFINO centre, but also in two other projects. Reflections on my personal experience have stimulated a philosophical analysis in an attempt to turn frustration into an understanding of what did not work in RRI and why. RRI is a marginal phenomenon, neither a transformation of research and innovation nor a systemic change. From a system perspective, RRI is homeopathic, that is, overly diluted, and does not have enough active principle to cure the disease of the system. But just like many people, somewhat mysteriously, benefit from homeopathy, many research and innovation projects benefit from RRI. The practice of RRI is often real, valuable, and inspiring. Yet, it remains peripheral.

The last thing I want to do is to devalue the worth and the efforts of good examples of RRI. They fully deserve to be praised, admired, and inspirational: kudos to them. In fact, my analysis of the unfavourable circumstances in which RRI is often carried out makes the success stories even more praiseworthy and should temper the pessimistic tone of this chapter. My pessimism concerns the transformative ambitions of RRI as a policy concept. Transforming institutions like research and innovation is a big endeavour. Change goes against the inertia, and to go against inertia, power and adequate means are needed. It is only fair to ask that science policies are proportional to the available power and means to effect change. Policy goals need to be set and communicated responsibly and soberly. Policy-makers should lead by example, especially when they invoke the concept of responsibility. We don't need promissory science policy incapable of delivering.

Norway seemed to be in a favourable position to implement the RRI agenda. Responsible Research and Innovation was endorsed and promoted by the Research Council of Norway (see chapters 3 and 5). Furthermore, Norway had a well-established tradition of research ethics committee and could count on an active community of researchers with experience in working on the ethical, legal, and social aspects (ELSA) of science and technology. Finally, Norway also had a strong tradition of Corporate Social Responsibility, which, on the business side of innovation, could have prepared the ground for thinking in terms of responsibility (see chapters 1 and 6). Nevertheless, even in Norway, the implementation of the ideal of RRI has met with many obstacles and challenges and has therefore failed to promote the system transformation that it was hoped to effect (at the international level, cf. Cuppen et al., 2019, pp. 152–153; Schuijff & Dijkstra 2020).

1.2 The aim of the chapter

An emerging RRI implementation literature has begun to analyse the institutional, cultural, and material causes of RRI's shortcomings. This chapter can complement the organisational-institutional analyses of the obstacles faced by RRI (Wittrock et al., 2021; Tabarés et al., 2022; Camarinha-Matos 2023). This complementary contribution includes:

1. A philosophical analysis of valuing and values that explains how hard it is to align the values of different groups, communities, and publics.
2. A reflection on the limits of moral commitments: people can only take up a limited set of responsibilities before they reach moral saturation.
3. An argument that to steer research and innovation towards socially desirable outcomes requires the formation of a joint responsibility within a value chain, but many conditions need to obtain for such responsibility to have normative force.
4. A denunciation of the mismatch between the resources, capacities, and incentives needed to perform some key tasks of RRI and the resources actually available to the individuals tasked to implement RRI. Many lack power, time, and means and are nested in an environment driven by logics that do not support RRI. These tensions are also illustrated in Chapter 5 through the experience of the AFINO research school.

My analysis indicates that the transformative ideal of RRI requires substantial transformations of the research and innovation ecosystem, but the knowledge, will, and power to enact it on a large scale are missing, both in Norway and in the EU. This is not a call to disband though; rather it is an invitation to scaling down and scaling out RRI. If changing the research system is beyond current capacity, then perhaps efforts should go towards smaller-scale but well-funded experiments in RRI, outside the R&I mainstream, in the attempt to building capacity, inspirational examples, and alternative systems.

2 RRI as a transformative ideal

2.1 Transformative RRI as a revisionary ideal

RRI is a manifold concept, understood differently by different actors. For simplicity, I distinguish between two versions of RRI: a) the transformative and ambitious version, which aims at restructuring the research and innovation system and bringing it in the service of public needs and values, and b) the subdued and instrumentalist version, which sees it as a tool to reduce the attrition rate of innovation coming from popular resistance and mistrust (the deficit model's new clothes). I focus exclusively on the

transformative understanding of RRI, because it is the only interpretation of the concept that has normative content and expresses a vision. So, *in this chapter RRI is shorthand for transformative RRI*. I call RRI an ideal, according to Elizabeth Anderson's understanding of what an ideal is.

> The core of an ideal consists in a conception of qualities of character, or characteristics of the community, which the holders regard as excellent and as central to their identities. Associated with this core is a conception of admirable conduct or worthy practices and projects that demand the cultivation, exercise, and expression of these qualities (Anderson, 1995, p. 6).

I contend that RRI is a vision of how certain activities should be performed, with what aims, and by what kind of community. So, RRI provides a vision of a community of researchers and innovators who share the aim of serving the needs and aspirations of the public, are committed to empowering the public to participate in setting the agenda, and want their projects and practices to incarnate these ideals. Finally, this vision determines which skills and virtues, obligations, and goals are preeminent for the members of this community. An ideal is a vision of a community with its distinctive values and moral norms. It is a moral vision which includes moral ideals for the various professional roles active within the community.

An important feature of Anderson's conception is her pluralist view of values and ideals: "There is a great diversity of worthwhile ideals", which cannot be combined and endorsed by an individual or by a group, and "different ideals may require the cultivation of incompatible virtues" (Anderson, 1995, p. 7). So, each ideal sets up a different hierarchy of the goods to be pursued and of the qualities and virtues to be cultivated. This point helps explain the moral conflicts between the RRI ideal and researchers and innovators committed to different ideals of their professional pursuits, for instance, researchers committed to value-free science or innovators dedicated to economic growth (cf. Tabarés et al., 2022, p. 292; chapter 2). RRI is one ideal among others in the fields of research and innovation, and actors can legitimately have different ideals, i.e. different ways of structuring their goals and their relative importance and of valuing professional qualities. Notice that ideals can be very far from each other, but they can also have many similarities. For instance, they may have the same pool of values but order their importance differently. This is an important point because some actors may have no objection to the values of RRI but still have different priorities, because other values (e.g. research excellence or competitive advantage) rank higher in their ideals (cf. Borch & Throne-Holst, 2021). For instance, a researcher may believe that the life of her lab depends on carrying on producing excellent research and hence ranking excellence higher than responding to public desires about research priorities while acknowledging that it would be nice to do so. My main point here is that RRI proposes a *revisionary* ideal for both research and innovation. These are normally driven primarily by values internal to each field – e.g. making discoveries, building theories, and creating new technology, new markets, and new business models – and only secondarily or indirectly

care for the public – through policy or market nudges. RRI wants to reverse this order and put public service first.

Focusing on the ideals at work in R&I and in RRI helps in highlighting two important issues. The first is that ideals can emerge in two different ways: as products of slow and unplanned processes of growth, or as intellectual constructs. The ideals currently at work in R&I are of the former type: communities have formed through time and adaptation and formed an implicit *ethos* that guides their behaviour and mutual expectations without being explicitly articulated. By contrast, RRI is an intellectual construct that provides a new ideal which is not yet rooted in existing practices and in the embodied ethical and cognitive frames of people. The second point is that intellectual constructs tend to look more rational than evolved "*habits of affection and behaviour*" (Oakeshott, 1962), while the latter tend to be more entrenched and resilient. So, what looks more compelling when considered from an abstract intellectual point of view often proves weaker than what is embedded in people's habits and sensibilities. However, let me stress that sympathetic understanding of individual responses to existing circumstances does not imply a justification or praise of those circumstances. The criticism of the political and economic forces that makes R&I inhospitable to RRI is outside the scope of this chapter.

2.2 RRI as a project to align R&I to societal values

But let's now look at the ideal of RRI. The most influential definitions are those offered by Von Schomberg and by Stilgoe, Owen, and Mcnaghten.

> Responsible Research and Innovation is a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the ethical acceptability, sustainability and social desirability of the innovation process and its marketable products (Von Schomberg, 2011, p. 9 emphasis removed).

> Responsible innovation means taking care of the future through collective stewardship of science and innovation in the present. [. . .] [R]esponsible innovation can be seen as a way of embedding deliberation on these [i.e. uncertainty, purposes, motivations, social and political constitutions, trajectories and directions of innovation] within the innovation process (Stilgoe et al., 2013, p. 1570).

Similar definitions can be found in several EU documents, for instance, the 2014 Declaration of Rome:

> Responsible Research and Innovation (RRI) is the on-going process of aligning research and innovation to the values, needs and expectations of society. [. . .]

> RRI requires that all stakeholders including civil society are responsive to each other and take shared responsibility for the processes and outcomes of research and innovation. This means working together in: science education; the definition of research agendas; the conduct of re-

search; the access to research results; and the application of new knowledge in society . . . (*Rome Declaration on Responsible Research and Innovation in Europe*, 2014).

Or earlier in the Science in Society programme:

> Responsible Research and Innovation (RRI) means that societal actors (researchers, citizens, policy makers, businesses, civil society, . . .) work together during the whole research and innovation process in order to better align the process and the results with the expectations of society. The following elements characterise Responsible Research and Innovation:
> 1. All societal actors are engaged in R&I, thereby increasing the quality, relevance, acceptability and sustainability of innovation outcomes by integrating society's expectations, interests and values (European Commission C (2012) 4526 of 09 July 2012, p. 5).[1]

These authors and documents stress the emerging need for more collegiate, open, and democratic ways of deliberating about the goals of R&I and the distribution of benefits and burdens. They aspire to create a process in which important and ethically relevant elements of the future that are constructed through science, technology, and innovation are discussed and negotiated broadly within society. RRI is thus a form of distributed governance of the future and the goal is aligning R&I with societal needs and values.

To sum up, RRI is a response to a perceived distance between science and society and to a certain discontent on the part of society. The dissatisfaction comes from feelings that a) experts patronise citizens and take decisions away from them (disenfranchisement), b) R&I disproportionately benefits the elites and leaves externalities to the people (injustice), and c) asymmetry of power biases communication and undermines trust (mistrust). These are the issues to which value alignment is presented as an answer, and they can be summarised as a legitimation deficit (chapter 1). There is also a challenge about the direction of R&I, namely the perception that it needs more steering to meet the most urgent problems and more acute needs. The response could not be governance from a technocratic or bureaucratic elite: this would be rejected by both citizens and researchers and innovators. Moreover, it would lack flexibility and adaptability. So, direction needs to come from devolved governance.

2.3 Building collegiality to mitigate uncertainty

How can the governance of the future be responsible? Since uncertainty of truly complex issues makes it impossible to anticipate consequences, responsibility should not

---

1 Similar statements are found in several documents. One example is quoted in de Saille (2015, p. 158), "societal actors (researchers, citizens, policy-makers, businesses, civil society, . . .) work together during the whole research and innovation process in order to better align the process and the results with the expectations of society". I found very similar statements at various places (e.g. European Commission, 2024).

be grounded on knowledge of consequences. Uncertainty comes at least from two different sources: limited knowledge and strategic interdependence. Limited knowledge is a feature of R&I, which works at the fringes of what is known and understood, and which travels new paths not yet explored. Strategic interdependence is the situation in which different actors respond to each other's decisions and behaviour. The activities of researchers and innovators trigger reactions from other actors, and these reactions are not predictable, because they depend on how actors *perceive* and *interpret* knowledge, innovation, and each other's evaluations and intentions. R&I are uncertainty generators, RRI tries to mitigate both kinds of uncertainty – which feed on each other.

Limited knowledge is tackled through attempts at anticipation. Strategic interdependence is addressed through improved communication and collaboration. To reduce the unpredictability of responses to research and innovation, more actors are involved in producing them. Wide participation and co-production should produce a widespread feeling of ownership and reasonable expectations about how involved actors will act. I believe that this second move is the more important one in mitigating uncertainty and is therefore crucial for the success of RRI.[2] An obvious example is public rejection of a technology. If R&I is done independently from the public, how the public will react is unknown and creates uncertainty, but if the public participates in the process, this uncertainty can be mitigated.[3]

If I am right, there is a move away from a consequentialist conception of responsibility and towards a conception of shared or joint responsibility. It is not a retrospective responsibility, understood as accountability when outcomes unfold. It is instead prospective responsibility, understood as a commitment to steer outcomes in desirable and acceptable directions. Rather than trying to predict the future, the point is to make widely owned decisions, in the production of which various scenarios and risks have been discussed and multiple needs considered. The result of the ideal RRI process may be seen as the construction of an objective collective intention in a public space (see Tuomela 2002). In the final section of this chapter, I try to illustrate the conception of joint, prospective responsibility that is needed for the success of this strat-

---

2 Carrier and Gartzlaff (2020) write that "foresight knowledge is widely considered impossible among a large number of RRI scholars" and this implies that "A 'consequentialist' approach to RRI founders on the tight limits set on predicting the development of knowledge and its social repercussions" (p. 151, see also p. 158).

3 I can bring another example from my previous research on pharmaceutical regulations. When regulators developed an innovative and adaptive approach to evaluating advanced therapies, they worked mostly with two types of stakeholders (industry and patients groups), and they were not able to anticipate the reactions of other stakeholders like technology assessors and payers, who struggled to adapt to the new regulatory regime (De Grandis et al., 2023). This problem was noticed by some scholars who thus advocated a system-wide approach to reduce uncertainty and improve efficiency and sustainability in innovation. This idea of adaptive biomedical innovation (Hirsch et al., 2016; Honig & Hirsch, 2016) is a good example of a strategy for reducing strategic uncertainty.

egy. We will see that building such joint responsibility requires many pre-conditions. Before we do that, we need to take a closer look at the idea of aligning values.

3 The difference between recognising value and valuing

3.1 What can it mean to align values?

I look at valuation and values because I want to understand what it means in practice to align values. Aligning values is one of those verbal expressions that sound attractive and may look like a solution to a problem. But to check if something is a real solution, one needs to understand what the practical, real-life implications of the phrase are. My discussion of valuing and values has this purpose. Accordingly, my emphasis is on *the difference between valuing something and acknowledging a value.* The concept of value is broad and used in a variety of ways in different academic disciplines. My own disciplinary background is in ethics, but I try to incorporate social scientific perspectives. The conception of value that I use aims at reflecting the experience of people – in philosophical terms it addresses the phenomenology of values or the axiological dimension of values, i.e. how values manifest themselves in people's lives and consciousness – but it also aims at showing a problem for the normative use of values, i.e. for the appeal to values to control people's behaviour. There is another possible source of confusion. Values are sometimes used meaning *moral* values and sometimes used more broadly – e.g. including aesthetic or epistemic values. Again, my argument straddles between these uses. In my description of valuing, I use values in a broad sense, and I emphasise especially the values internal to shared practices, which are not only moral values. When instead we deal with the ambition of RRI to align values, I interpret this aspiration as referring mostly to the moral values of the public. Finally, there is also an important distinction between tacit values and asserted values – or, if you prefer, between implicit and explicit values. My analysis stresses the importance of tacit values, while I understand the alignment ambition as working with asserted values. In a nutshell, my argument aims to show that *an alignment on asserted values runs the risk of being shallow and fragile, because it overlooks the power of tacit values embedded in habits and local shared practices.*[4]

Another question is what we mean by the aligning metaphor. I take it to mean that it does not demand a full assimilation of the values of different people. Rather, I

---

4 Some may object that I use a descriptive approach to valuing to criticise a normative use of values. That is exactly what I am doing. Normative prescriptions that are based on an incomplete understanding of the circumstances in which they should operate and of moral psychology are defective. Facts and values do not exist in separate realities; on the contrary, they are intertwined. RRI aims at chang-

interpret alignment to describe a state in which different people can accept each other's main goals and behavioural norms and believe that these are not harming or undermining their own. It is a state that makes room for compatible differences and where conflicts have been smoothed over.

In the next section, I illustrate how values are embedded and may conflict in R&I. In section 3.3, I explain the conception of valuation and values that I adopt, which helps to understand what it means to value something, in what ways it differs from seeing something as valuable, what kinds of attitudes are involved in valuing, and why making values explicit is not an easy task. In section 4, I will discuss some examples of challenges associated with the attempt to align values in different circumstances that present themselves in the context of RRI.

3.2 Communities and values in research and innovation

Let me focus on the domain of research first. Research is a practice that typically takes place within some groups, either because it is carried out in teams or because it is addressed to and assessed by one's peers. So, we can already identify a research team and a peer or disciplinary community as two types of groups that have and practice their own values. Some research teams are interdisciplinary, and that means that they bring together people with slightly different professional values. Some research projects and most innovations aim at developing outputs that will be applied outside the lab or the firm, and hence they need to take into account not only their own professional values and standards but also external criteria and values, like the needs and preferences of the final users or customers, the legal requirements pertinent to some products and services, criteria of economic efficiency, regulatory, quality, and manufacturing standards.

Let me give you an example from medical innovation. Imagine a team of medical researchers and clinicians who are developing a cell therapy for a very severe disease with no effective therapy. They have their values as scientists who want to gain knowledge and communicate it and be acknowledged for their work. They also have their values as clinicians who want to help patients in desperate need of a cure. On top of this, they may be under pressure from their funder and employer to patent their discoveries and to collaborate with pharmaceutical industry to ensure that the therapy they are developing will be available on the market and hence to all the patients in need. Turning their research into a marketable innovation brings with it new interactions and demands. For instance, generating the data and evidence needed for getting marketing authorisation poses some demands on how they should proceed in their clin-

---

ing reality, not at setting some absolute abstract normative standards. For interesting discussions of the relation between descriptive and normative ethics, see Sayer (2011) and Hämäläinen (2016).

ical experimentation. These demands may conflict with their clinical judgement or with their scientific curiosity. For example, they may want to change the manufacturing process of the cell, but that would make the data collected so far unusable as evidence for regulatory approval and delay the marketing of the product. So here different values – scientific curiosity and the need to reach the market quickly both to benefit patients and to allow the firm to recoup its investment and make a profit – may push in different directions. Various decisions taken during the development process may involve conflicts of value. For instance, producing more and better evidence or optimising the manufacturing process may lead to a more accurate targeting of the therapy and to lower production costs but may also substantially delay marketing or even lead to losing the chance of accessing the market if a competitor is granted market exclusivity for that indication.

3.3 Roles as embedded normative constellations

We can look at the different priorities and values that may conflict during research and innovation by considering the different roles (and associated values and obligations) that different stakeholders carry with them. In this context, these will be mostly (but not exclusively) professional roles. Here a role can be simply defined as "a capacity in which someone acts in relation to others" (Emmet, 1966, p. 13). We can take professional roles as useful signposts for a cluster of values, rules, and obligations. So, each role has a "morality" attached to it; such morality is clearly not exhaustive of the ethics of the person behind the role (the *persona*), because, typically, individuals have several roles in their lives. However, professional roles are significant because they contribute quite substantially to the social and personal identity of people and their moral profile. Under the assumption that a person performs her professional role with some integrity, the rules and values associated with the role are practiced on a regular basis and are not just abstract moral requirements and usually become second nature.[5] The depth of the connection between the role and the person is illustrated by the contrast between the professional – for which the identification is deep – and the amateur – for which it is superficial. Roles surely imply some room for individual judgement and some leeway in how to interpret them. However, as Emmet perceptively discussed, the person and the *persona* (i.e. the role) always coexist.

One important factor in the balance between the person and the *persona* is the rigidity of the institutions that host the roles. To be sure, no role is completely determined, but the extent to which individuals may transform roles is not just up to their

---

5 As Herbert Simon noted, an individual "does not live for months or years in a particular position in an organization, exposed to some streams of communication, shielded from others, without the most profounds effects upon what he knows, believes, attends to, hopes, wishes, emphasizes, fears, and proposes" (Simon, 1976, p. xvi).

initiative. Contrary to Richardson's (1999) emphasis on this active reshaping of roles, I contend, following a venerable tradition from Max Weber to Ursula Franklin, that the rise of organisations, technology, and management and more recently of new public management and entrepreneurial universities has tended to restrict individual discretion and latitude (cf. O'Neill, 2002, ch. 3). This is happening also in R&I in the Nordic context, as Ylijoki (2003, p. 330) remarks, "it can be claimed that academics are increasingly 'managed professionals' whose power in managerial discretion has diminished as compared to university managers (. . .). In other words, academics have greater accountability but reduced autonomy" (Cf. Glerup et al., 2017).

3.4 Valuing and the different force of enlivened and notional values

In becoming second nature, roles establish and consolidate patterns: patterns of perception, judgement, emotional responses, and action. These are central features in Samuel Scheffler's account of valuing – in the sense of holding dear, cherishing. This account has the merit of reflecting our experience of valuing more faithfully than alternative accounts and of explaining why disagreements about values do not lend themselves to purely discursive reconciliations. It also explains why people can recognise the importance of a value while failing to act on it when it conflicts with established practices in their professional domain. Conversely, it also explains why, without altering the practices of a professional group, it is very unlikely that their work aligns with external values. In short, it helps to appreciate the importance of the gap between acknowledging values and acting on them through habits and sedimented patterns. In the context of RRI, a relevantly similar account of values has been proposed by Boenink and Kudina (2020), especially with their ideas of "values as lived realities" (p. 456).

In contrast with many accounts that try to reduce valuing to one essential element, Scheffler proposes a multidimensional explanation: a complex of beliefs, dispositions, and susceptibility to certain emotions. Specifically, valuing any X involves:
1. A belief that X is good or valuable or worthy (cognitive element)
2. A susceptibility to experiencing a range of context-depending emotions regarding X
3. A disposition to experience these emotions as being merited or appropriate
4. A disposition to respond to X and what affects it. These responses can be directly in action or in deliberation, i.e. in terms of having certain reasons for action[6]

---

6 On this last point, I have slightly modified Scheffler's account to make it more pragmatic and less intellectualistic.

Let's illustrate this with an example. As a scholar, I value accuracy in research and communication. So, this means that I see accuracy as something worthy and valuable. My valuing accuracy means that I am subject to certain emotional reactions; for instance, I will feel shame if a colleague points out that my references are sloppy and inaccurate, or I would feel annoyed and disappointed if I notice that students do not care about accuracy in their essays. On reflection, I will not consider such emotional reactions as arbitrary or misplaced, as I could do if I get annoyed with my 3-year-old daughter because she is not shaping the cookies well enough. Finally, my commitment to the value of accuracy will also show up in my checking a reference when a doubt arises – automatic action – or in my decision – deliberation – to check my transcription of an interview even if it is late and I would like to go home. Notice that all this can be the case even if I am not reflectively aware of my commitment to accuracy. So, because of its deep link with emotions, valuing has motivational force and prompts actions – with or without deliberative mediation.

A very important feature of this conception of valuing is that it considers the belief in the value of something *necessary but not sufficient* for valuing – it is only one of the four conditions listed above. Therefore, it is possible to see something as good or worthy without valuing it. In other words, one can see that something is worthy and understand why other people value it, without valuing it oneself. I may acknowledge that being a check master or a marathon runner is something valuable, without valuing it. We can express this distinction by calling the act of judging something as valuable without having strong emotional and executive commitments *recognising as valuable.*[7]

To mark the difference between the values that an agent recognises and what an agent values and lives by, I will call the former *notional values* and the latter *enlivened values*. While there is no limit to how many values one can recognise, there are limits to how many goods and values one can value, because it means giving them a special place, a superior ranking, or priority. So, for every agent, the set of enlivened values will always be smaller than the set of notional values. Another important and less obvious feature of this account of valuing is that it is compatible with the fact that valuing does not require a primacy or even an awareness of the cognitive element. Appreciation is not something that has to come first; often it is the result of habituation, first-personal experience, and increasing emotional investment, which eventually provides the grounding for the cognitive judgement of appreciation.

---

7 These judgements can bring with them some weak deliberative implications and sometimes some weak emotional implications. If I recognise that something is valuable, but I am not actively interested in its value, I can still feel some obligations to try not to harm it. There is a sort of negative obligation attached to values, something that we can express as not harming values. Coming to emotions, some people (perhaps those with a very developed moral imagination) may be disposed to feel certain emotions in certain circumstances, if they have explicitly recognised something as valuable. However, these emotions would be vicarious, like those of a sympathetic impartial spectator.

Table 1: Types of valuations and values.

Valuing Enlivened values	Recognising as valuable Notional values
Belief that X is worthy (possibly tacit)	Belief that X is worthy
Strong emotional responses to X	Weak emotional responses to X
Strong cognitive awareness of what affects X	Limited cognitive awareness of what affects X
Strong motivational power	Weak motivational power
Incorporated in deep seated habits and practices (concrete)	Only active at intellectual level (abstract)
May be tacit (habits)	Is always explicit

Consider these examples. 1) I may think that Tai Chi is a weird activity, but then, by accidentally trying it, I begin to like it and then even see the aesthetic beauty, the spiritual, physical, and social value of this discipline, and then completely reverse my initial judgement. 2) I am tempted to eradicate the dog rose bush at the back of my house because it seems to me just a messy bush. But, out of laziness, I leave it. Then I start taking a little care of it, and gradually, it becomes a pastime and I increasingly appreciate the beauty of the flowers. So, I end up valuing it, even if I have never consciously thought that it was worthy. This is not a qualification of the conception: the cognitive element, the belief that something is valuable, is there, but, in some cases, it may not be explicit and sometimes not even reflectively available. For some things that have become so deeply rooted in people's way of feeling and behaving that belief may not be on the surface of our awareness, it could be a tacit presupposition.

The important point is that the process of enlivening a value is not purely cognitive, but it requires changes in how an agent perceives certain relationship with, for instance, some people, practices, or projects. Enlivened values guide our appraisals and behaviour without the need to be present in our awareness. They are a main component of our practical life – and occasionally of our practical reasoning. For this to happen, the value needs, so to speak, to take roots in our practical identity, which means that it needs to change some of the things we do, as well as how we do some things and how we perceive and feel in some circumstances. An enlivened value is not something that an agent can simply declare; it has to be expressed in some deeds, emotions, and attentiveness. Purely verbal enunciation is not a sufficient expression for an enlivened value (cf. Oakeshott, 1962; Dewey, 1983; Boenink & Kudina, 2020).

3.5 Apparent and real value alignment

This account of valuing and values implies that engaging in rational conversation or deliberation may lead participants to converge on recognising some notional values while not leading them to actually value them.[8] The distinction between recognising as valuable and valuing highlights the fact that valuing implies a deeper commitment, which is the result of changes in our cognitive, emotional and behavioural patterns. Hence it points to those experiences, practices, relations, and habits that effect the transition from seeing as valuable to valuing, i.e. to the process of enlivening values, a process that takes time. A notional agreement happens when people discuss values and can all acknowledge that a certain value or cluster of values are important and desirable. They may even agree on how to structure the importance and relevance of a cluster of values in different contexts. However, this process happens at a purely intellectual level and does not automatically embody itself into emotional reactions, habits of attention, and relations and practices incorporating and expressing those values. Only when these further steps are consolidated do the agreed-upon values come to life and inform practices and dispositions. Only at this point has a practical alignment of enlivened values been achieved. This practical alignment does not mean a perfect integration of behaviour, neither within an individual nor within a group or team. When a problematic situation arises, there will be the need for reflection, inquiry, judgement, and hence room for differences, disagreement, and contestation. But when some practical alignment has been achieved, problematic situations are the exception rather than the norm.

4 Aligning values in the context of RRI

In this section, after providing a sketch of the formation of a researcher professional identity, I illustrate some situations in which the RRI ideal of promoting an alignment between R&I and public values creates demanding tasks for which RRI does not provide enough resources. The three vignettes explore:

---

8 A similar point has been observed by Waldorff and Madsen (2023) in an empirical study of the implementation of a new concept in a healthcare setting in Denmark. In spite of the fact that the concept was seen as valuable, its translation into practice met with several obstacles and was only very partially implemented. The authors conclude that adoption is possible only when the translators "believe that the new model aligns with existing logics, not to mention that they have leverage in changing related practices" (p. 444). Their study can be seen as an illustration of how enlivened values (or already established practices) may block actual adoption of new concepts even if they are seen as valuable.

1. Intrapersonal alignment of existing professional (or role-related) values and RRI values
2. Team alignment within a cross-disciplinary collaboration whose members belong to different communities (epistemic communities or communities of practice) with different values
3. Wide alignment between a team, stakeholders, and affected publics

4.1 The socialisation of a young researcher

In sketching the challenge of intrapersonal alignment, I pick the case of an early career researcher who is both pursuing a career within an academic discipline and taking up the responsibility of making a research project compliant with RRI requirements. I assume that this researcher endorses the ideal of RRI as I described it above. I also assume that research work is pursued not barely as a means to earn a salary or a position within society but also as an activity that brings with it intrinsic rewards, like the pleasure of finding or testing things, the sense of contributing to a valuable intellectual or social endeavour, the pleasure of continuing to learn and apply high-level knowledge and skills, and the pleasure of being part of a community of curious and intelligent people. This last point brings us to the interpersonal dimension of research. Becoming a researcher means earning membership in a community and endorsing an ideal as defined by Anderson (see above section 2.1).

Each discipline is characterised by its own epistemic culture, its traditions, schools of thought, established practices, and conventions. To fit within a disciplinary research environment, a researcher needs to be socialised into these cultures, standards, practices, and values to make them one's own. By the time one has become a professional researcher, one has internalised many of these. To thrive and to find pleasure at work, one needs to fit well both within its immediate working environment and within the broader disciplinary community. Furthermore, one needs to be acknowledged as a contributing member who upholds the values and standards of the group. A research job is meaningful and rewarding as long as a researcher feels that a) the epistemic community is true to its authentic values, b) the researcher is living up to the community's standards, and c) the community recognises the status of the researchers as a valuable contributor. When these conditions are realised, the researcher has endorsed and internalised the values of the disciplinary community. The community's goals, rules, conventions, and practices have become enlivened values.

Nevertheless, the process of integration can bring with it several tensions, for a researcher may well admire many features of the epistemic culture and disciplinary practices while feeling uncomfortable or critical about other aspects – for instance, some persisting sexist attitudes, or some insufficient sensibility towards the needs of experimental subjects, or a certain smugness and dismissive attitude towards other disciplines. So, the integration may include some frictions and reservations and

should not be seen as complete and flattening integration. An epistemic community accommodates some internal differences and disagreements and can tolerate some eccentricity. However, it should be noted that while one is still in the process of being fully acknowledged and building a reputation within a community, playing the critical voice is not easy and often not expedient. So, already aligning oneself with a disciplinary community may involve some tensions and ethical probing.

4.2 Intrapersonal alignment: to be a scholar or to be an RRI advocate? A vignette

What happens when a young researcher takes up the responsibility of realising the RRI requirements within a project? Let's assume that the project is an interdisciplinary one and that our early career researcher is a PhD candidate, let's call her Lise, who belongs to a discipline that is contributing to the project. Lise needs to earn her credentials within her disciplinary community, but, at the same time, she needs to carry out her RRI duties. For instance, she has to promote anticipation and reflexivity within the project. Anticipation, according to Stilgoe and colleagues, "faces a tension between prediction, which tends to reify particular futures, and participation, which seeks to open them up" (Stilgoe et al., 2013, p. 1571). Furthermore, "Anticipatory processes need to be well-timed so that they are early enough to be constructive but late enough to be meaningful". It looks like a job for a professional in squaring circles. Perhaps it would be easier to focus on reflexivity. Or maybe not, if we accept that reflexivity "asks scientists, in public, to blur the boundary between their role responsibilities and wider, moral responsibilities. It therefore demands openness and leadership within cultures of science and innovation" (Stilgoe et al., 2013, p. 1571). The sympathetic reader is probably beginning to feel quite sorry for poor Lise. How will she manage to persuade her busy senior colleagues to question their well-consolidated habits and duties, in public, while they are focused on delivering the deliverable of the project and to carry out their many other duties?

Maybe it is better to put her stakes on inclusion; probably people are eager to be involved, but first, she will have to understand what kinds of expectations the funders, the colleagues, and the external participants have. Afterwards, she will have to ensure that there will be room "for public and stakeholder voices to question the framing assumptions not just of particular policy issues (. . .) but also of participation processes themselves" (Stilgoe et al., 2013, p. 1572). So, what will be the success criteria of this big job of recruiting participants in a genuinely inclusive way, of ensuring that they produce pertinent inputs without priming them, and of communicating the results in such a way that funders and colleagues will be responsive to the inputs?

The problem for Lise is that she is in a very weak position to perform the critical conscience of science: she does not have the experience, the status, or the capacity, and she is unlikely to find enthusiastic support from her colleagues. She finds herself in the very uncomfortable position of having to question the very values that she is

interiorising and to challenge her seniors. Furthermore, she is also in the very difficult position of not having very well-specified objectives and, perhaps most decisively, no robust solutions for operationalise the aims of RRI. This puts her in a very vulnerable position because she is asking her colleagues to divert time and energy to activities whose objectives are vague and whose results are extremely uncertain. Her colleagues have prudential reasons and enlivened values that push them to bet on their own habitual and familiar (and much more robust and predictable) disciplinary work rather than on the woolly ideal of RRI.

The risk is indeed that she will discover that the less seriously she takes the ambitions of RRI and the more she focuses on her own research, the more accepted she will be within the project and the better chances she will have to further her career. To have some modest and non-intrusive token initiative in the spirit of RRI will be her ticket to come out of the project sane and with some residual prospects of employment.

4.3 RRI demands and moral saturation

This of course is just a vignette of a hapless case, but unfortunately, it tracks the experience of not a few people I have talked to (and my own experience with another RRI project), both in the sense of helplessness of the junior researcher and in the scepticism of senior researchers. Let me start from this last point. In academic contexts, RRI hits against the barrier of what I call moral saturation: academic researchers are facing expanding demands and growing role-related duties that already stretch their capacities. Ylijoki (2003, p. 331) describes the situation in terms of:

> the co-existence of two value sets. Market-oriented values and ideals have strengthened, stressing, for instance, the importance of attracting external funds, establishing a good reputation within funding agencies, capability for cost-effectiveness and efficiency, and creating and sustaining large collaboration networks within and across academia (. . .). But at the same time traditional academic values and ideals, such as freedom and autonomy, reputation among peers, recognition within the scientific community, publications in highly ranked journals and intellectual contributions to one's field, are also regarded as essential in research work.

Similarly, Glerup and colleagues (2017), in a study of how researchers perceive their responsibilities, found that they already have a full plate and struggle to see the point of RRI. Their perceived responsibilities include: a) responsibility for doing robust and high-quality science, b) responsibility for caring for people and managing research groups, c) responsibility for supporting the organisation within which research is done and for managing its resources carefully, and d) responsibility for respecting citizens' views about what research is legitimate. Unfortunately, often RRI promoters have adopted a normative approach to expanding responsibilities, without taking "the perspective or researchers into consideration" (Wäscher et al., 2020, p. 146). Instead, we "need to ensure that the promotion of responsibility does not alienate those

working in research and innovation" (Glerup et al., 2017, p. 330). So, the central question should be: "how can RRI be embedded into the normal routines of behaviour of professional researchers? Perhaps, more importantly, how can RRI activities be embedded so that they are valued by the research organisation itself?" (Shelley-Egan et al., 2018, p. 1740; cf. Wittrock et al., 2021, ch. 4). Pushing researchers already saturated with responsibilities is counter-productive because, as Williams (2006, p. 213) notes, "people who are over-loaded tend to drastically simplify their sense of responsibility, both so that their tasks are made manageable and so that they do not have to see themselves as failing in their duties".

If we look at the demands on young academics trying to build a career, the picture is again one of overwhelming demands, high competition, and very precarious jobs (see chapter 5). They need to publish, to gain teaching experience, to learn to write successful grant applications, to disseminate their results and build professional networks, to be proficient in the use of digital technologies, to record their achievement and have evidence for them, and to engage with the public. To succeed or just to survive, people need to be strategic and set their priority right (cf. Debowski, 2012). Yet, engagement with RRI is not rewarded enough to pass the strategic priority-setting test (Felt, 2018).

RRI is not entering a field where carefree researchers devote themselves to their beloved studies, oblivious of everything else; quite the contrary: researchers feel burdened by so many demands, duties, interpersonal obligations, financial pressures, and administrative tasks that finding the time for doing research is a challenge (Gill, 2010). As a head of department told me, many academics asked him to be protected from the constant encroachment of administrative and external demands.

The other problem is that RRI objectives are both grand, insufficiently specified, and hard to operationalise. What strikes me about the last point is that often a host of methods and approaches are mentioned by RRI proponents – for instance, by Stilgoe and colleagues (2013) – but even if, for the sake of argument, we grant that they are good enough, they can only be examples or parts of a broader plan. But how broad? Serious reflections on at what scale, on what timescale, under which presuppositions, and through which resources the goals can be achieved are lacking. Often junior researchers with no dedicated training, with limited or no previous experience in these practices, and often with only part of their working time allocated to RRI are left with the burden of implementing practices with a transformative ambition. Furthermore, no attention has been paid to the fact that what has to be changed are highly valued, consolidated practices that define the identity and pride of epistemic communities, as well as securing their status as professionals and experts (Glerup et al., 2017). My analysis of valuing and values shows why it is so hard for abstract values to transform enlivened values: even if usually the former are recognised, they are not embodied as enlivened values that have come to shape the routines, the emotions, the perception, and the identity of agents. To be more concrete, things that easily conflict with enlivened values include: a) opening to wider inputs – often of variable epistemic standing;

b) disrupting closed or non-transparent practices; c) changing evaluative practices – for instance, adding external criteria; and d) introducing new steps in routines and procedures.

To be sure, the causes of failure that I am pointing out are compounded by other very strong causes at the institutional, organisational, political, and economic levels.

4.4 Another vignette

Imagine the following scenario. A new process developed by an engineering department has the potential to make profitable again an extractive sector that has been declining sharply in recent years, leaving several communities centred on mining activity in a deep crisis. If the new process is rapidly adopted, some mining communities can be back in business. This means the continuation of their existence and current way of life. This is much welcomed by the miners but not by some other residents and by young people who are increasingly aware of the negative health and environmental impacts of the mining activity, are uncertain about its long-term prospect, and would very much hope for a conversion to a different economic basis.

Now engineers and social scientists are promoting a project that aims at exploring the viability of the new process in one of these mining communities. But the two departments have very different needs. The engineering department has created knowledge and capabilities and sees the opportunity to establish a research stream in these new processes, but it needs to retain the leading researchers who are also attracted by some offers to develop their process abroad. It also needs to show the impact of its research in terms of industrial collaboration and benefits to the local economy. The social sciences department, on the other hand, is interested in developing skills and experience in large-scale social involvement in innovation and planning. Furthermore, a thorough and in-depth involvement opens the possibility for several research projects in the area that could bring funding and new opportunities for the department graduates and post-graduates who do not enjoy such great employment prospects. So, one of the promoters of the research project (the engineering department) would like to have a quick stakeholder involvement, dominated by the mining industry, the unions, and the local government, all very well disposed towards the testing of the new process. The social scientists instead would like to have a broad citizen involvement exercise that includes more voices, like those of the emerging tourist sector, the young, the women, the environmental activists, the public health experts, and those suffering from chronic diseases associated with the mining job. This will require a longer and more uncertain process and presumably growing personnel involved in the process. This approach though requires a much longer time and very substantial resources to support broad involvement activities with diverse groups – some of which are loosely organised and unused to participate in such exercises – and finally, it is much more uncertain in its outcome, not only in the sense that it is hard to pre-

dict what the result will be but also in the sense that it is not obvious that the results will be robust and would support a clear decision. They may lead to contestation and disputes.

4.5 The challenges of team alignment

This fictional example shows that costs, duration, and uncertainty of the process may be perceived as a threat by some partners and as an opportunity by others. For the engineers, the process needs to be time controlled and predictable, at least on the type of outcome: the new industrial process will be implemented, and the process should only decide how this will be done. To have this kind of outcome you need to pick the right set of stakeholders. For social scientists, by contrast, an expansive process means expansive and more interesting research and work opportunities, and to achieve this result, the set of stakeholders needs to be broader, more diverse, and less united in their goals, values, and expectations.

Researchers who have results that they want to apply express a supply-driven perspective, which prefers a narrow public involvement. On the contrary, researchers who would like research and innovation to be shaped by the needs and demands of people rather than driven by the needs of the market economy or the knowledge production system are endorsing a demand-driven approach that favours a flexible and open-ended model of citizen or stakeholder involvement. Here we see the conflict between the subdued and the transformative understanding of RRI.

While the case presented in the vignette is rather extreme and does not claim to be typical, it still highlights how difficult it may be to achieve alignment even if both research partners can agree to adopt an RRI approach. Both the engineering and the social science departments may be motivated by perfectly legitimate interests and by their own sense of what their main social and scientific responsibilities are. Both feel a responsibility to contribute to the funding stream to their department and to create new opportunities for young researchers, but that produces conflicting interpretations of the scope and timeline of the project. Both feel a responsibility to society at large, but again, they interpret this responsibility differently. The engineers believe that putting to use the result of more than a decade of publicly funded research and to enable a community to remain economically viable represents an obvious way of promoting the public good. The social scientists instead believe that using their capacity to promote a wider involvement of the local community in deciding their future and to help counter-balance the asymmetry of power within the community is their distinctive and dutiful way of contributing to the public good. Both perspectives are grounded in the professional values and in the loyalty to their own community.

4.6 The challenges of wide alignment

Including the voices and values of stakeholders and affected publics forces researchers and innovators to move beyond the moral simplifications that are provided by the professional and disciplinary division of labour. It forces researchers and practitioners to mediate between the needs and values of different human groups. These groups can be very different in terms of their extension and of their ethical and functional cohesion. Two consequences follow. One is that researchers and practitioners must navigate a rich and multilayered moral landscape. The second is that different groups need to be engaged in different ways. Asymmetries in knowledge, power, internal integration, and strategic focus can be very significant and demand very different work and resources to enable a fair and level playing field.

When attempting a broad involvement, RRI operates in a space that is characterised by multiple discordances, of which the most important are value pluralism, uneven group integration, and unequal strategic awareness. By uneven group integration I mean that with respect to their attitude towards the research or innovation in question, some groups are cohesive, and some are not (e.g. in our vignette the mining company is cohesive, while the local community is not). By unequal strategic awareness, I mean that some groups are aware of what is at stake for them, what they want, and how to present their case, while other groups have not yet developed this awareness. It is because of the presence of these three features (value pluralism, uneven integration, and unequal strategic awareness) that I describe this as a mis-composed space to emphasise that it is a space where a mediation needs to be achieved between parties that are not equally ready to engage in a deliberation. This situation makes the task extremely challenging and open to contestation and failure.[9] Furthermore, if we look at this challenge from the point of view of the people that should manage it, we have to acknowledge that it is a very delicate task that requires high-level skills, experience, good judgement, credibility, and great personal integrity. This is clearly not a task that should be given to PhD candidates.

The most disturbing consequences of the difficulty and uncertainty of achieving wide alignment emerge at the political and policy level. At the political level, in our vignette, if the local government is keen on supporting the rescuing of the mining activity through implementing the new technology, is it legitimate to demand a different legitimation through a wide public engagement exercise? If it is legitimate, what are the preferences and values that need to be represented and do they have the same weight? There are some stakeholders who have clear views (industry, local govern-

---

9 This issue raises the question of whether a researcher trying to follow RRI principles is a neutral facilitator, an expert, or an advocate for some (public or underrepresented) interests. As shown in chapter 2, researchers will have different attitudes towards trying to be value-neutral or not. An interesting discussion of similar issues has taken place in the field of town planning around the idea of *advocacy planning* (Davidoff, 1965; Checkoway, 1994; Sager, 2022).

ment, trade unionists, young people wanting a different future for the community, occupational health practitioners), but there are others who may not have pre-formed preferences and visions about the future of the community. Is it right to include them in the public engagement exercise? Shall we look at them as those excluded from the debate, or shall we look at them as those who do not care deeply about the issue? Are the values deeply embedded in the forms of life of those who are working in the mining industry more enlivened and hence more genuine than those advocated by the younger generation with a different vision? Are we having a conservative bias if we give more weight to the values rooted in existing practices? Or are the values that have not been tested by people's experience too thin and speculative to be given the same weight? As Boenink and Kudina (2020, p. 459) correctly point out, "technology is one of the many factors destabilizing moral routines", so the hermeneutic of the existing values needs to be supplemented by actively engaging the imagination. But how to weigh values coming from different timeframes? I suspect that most of these issues may be hard to settle and may remain open to contestation.

The risk is that in circumstances of societies characterised by a plurality of conflicting systems of values, any thorough pursuit of a genuinely inclusive wide alignment may generate inconclusive or time-consuming processes. Such an outcome seems to hinder rather than support innovation. Perhaps it is the right thing to do, but it goes against the widely held belief that a fast flow of innovations is necessary for supporting the economy.

Alternatively, and less ambitiously, RRI could be a kind of value-sensitive technocracy, a process led from the top, looking for the relevant values to incorporate in framing their goals and their moral constraints. However, this option is vulnerable to the accusation of selection bias in selecting and co-opting the stakeholders and the relevant publics. The process is open to include those user or public values that can be accommodated with ongoing R&I tracks, not to those who can contest them and bring them to a halt.

At the policy level, what would be the implications of an RRI process that requires an open and flexible project format? A project whose main purpose is open to redefinition through wide involvement is very hard to assess. A project has to be based on a problem statement, a research question, or an envisioned new product, service, or process. However, transformative RRI implies that these should be but a working hypothesis that is responsive to external inputs. So, what is really needed and worth doing cannot be determined in advance, but it is expected from the (anticipatory, inclusive, reflexive, transparent, and responsive) process. The problem is that the process is intensely costly and uncertain in its duration and outcome. If we don't know how much a process will cost and what kind of results (benefit) it can deliver, how can we decide whether it is worth the effort, and, even more acutely, how can we compare different project proposals at the time of funding allocation? A fair and efficient allocation of research and innovation fundings becomes impossible or, at best, a wildly uncertain guesswork. This is not the only challenge though. A process that is

genuinely open to society's inputs and stakeholder involvement is not only requiring funds for the researchers who lead or facilitate the process: it also needs to attract and retain the participation of the public or of (all or as many as possible of) the stakeholders. Then the threat of exiting the process or of dragging it can be used as an opportunity for strategic bargaining by the stakeholders with more strategic awareness. Finally, there is the question of the cost-benefit balance of stakeholders or public involvement. Because "it has to be acknowledged that any kind of scientific reflection or wider engagement takes planning, time and skills" (Glerup et al., 2017, p. 330; cf. Carrier & Gartzlaff, 2020, pp. 163–164). The question of whether it is worthy arises both for those who are committed to a more participatory science and for the people who are supposed to participate. Both may wonder whether it is worth to open or to join a discussion on values, because "like all things, they have costs, material and psychological, and it is not granted that it is always worth paying them" (Viano, 2002, p. 116; on citizens attitudes see Lowndes et al., 2001).

So far, I have shown that one of the pillars of RRI as a distributed governance of the future – namely value alignment – faces several challenges. Next, I will look at the other pillar: collective prospective responsibility.

5 Joint prospective responsibility

In this final section I discuss the other instrument needed for realising transformative RRI. First, I show that RRI scholars have made a robust case for the need for a prospective kind of collective responsibility. I think that they are right that without it, RRI cannot achieve its ambitions. Next, I describe the features of prospective collective responsibility and outline the steps of a process to construct such responsibility among different organisations. In philosophical terminology this is a case of constructing a joint prospective responsibility among a collection of collective agents, a challenging case for responsibility. I will use a simpler terminology and speak about joint responsibility within a value chain. I will then proceed to build a formal model of its construction. The purpose of the model is not practical; it is not meant as a recipe or a set of instructions. It is rather a theoretical model to highlight the necessary conditions that need to be in place for this kind of responsibility to be possible and workable. Finally, I argue that in light of the challenges of aligning values and of numerous side constraints that most organisations have, the construction of joint prospective responsibility is possible only in some favourable circumstances that currently are seldom found. So, this section shows that the difficulty of value alignment also affects the establishment of the kind of responsibility which is more crucial for achieving the ambitions of transformative RRI. Both pillars of transformative RRI – value alignment and shared responsibility for the future – are fragile indeed.

5.1 Retrospective responsibility is not enough for realising RRI

Many proponents of RRI have stressed the inadequacy of the prevalent conception of responsibility, which is individualist and retrospective (Adam & Groves, 2011; Owen et al., 2013; Spruit et al., 2016; Von Schomberg, 2007; Wäscher et al., 2020). Some have stressed the need to develop a collective model of responsibility (Grinbaum & Groves, 2013; Owen et al., 2012, 2013; Von Schomberg, 2007; in a slightly different vein Spruit et al., 2016 have argued the duty to unionise to build collective agency) and some have stressed the need to develop a prospective (or forward-looking) model of responsibility (Grinbaum & Groves, 2013; Owen et al., 2013; Pellé & Reber, 2015; Spruit et al., 2016; van de Poel & Sand, 2021). I believe that both these dimensions of responsibility are central to a conception of responsibility which meets the ambitions of transformative RRI. The case that stresses the limits of retrospective responsibility has been made quite strongly by von Schomberg (2007), Adam and Groves (2011), Pellé & Reber (2015), and – with some qualifications – van de Poel and Sand (2021). So, I take it as well established that while individual retrospective responsibility is not irrelevant for RRI, it is insufficient to achieve its ambitions. However, van de Poel and Sand (2021) have argued that a properly understood prospective individual responsibility is all that is needed for RRI, and Grinbaum and Groves (Grinbaum & Groves, 2013) also give a substantial role to individual prospective responsibility. Both papers end up emphasising a virtue- and care-based conception of responsibility. I argue against their proposed solution in a forthcoming paper and hence I will only give a cursory explanation of why individual responsibility cannot provide the solution.

5.2 Individual responsibility is also insufficient for RRI

There are three main problems with staking RRI on individual responsibility: one is cognitive, another is about capacity, and the third is ethical. Understanding and foreseeing what the socially desirable outcomes are, as well as the unintended consequences of the R&I process, is not something that is typically available to any individual involved in the process. Indeed, it requires both pooling knowledge and expertise from various actors and proactively inquiring about the needs and expectations of users and citizens, as well as the unexpected uses of innovations. This demands a collective and well-coordinated effort that involves many actors along the R&I value chain. Furthermore, it also requires an allocation of resources (most notably people's time), incentives, and the creation of prerogatives that enable people to take the necessary actions. Outcomes that depend on the system's behaviour and on the dynamics of innovation adoption and diffusion are hardly predictable by individuals within the R&I value chain (Swierstra & Jelsma, 2006; Von Schomberg, 2007; Wäscher et al., 2020). According to Shannon Vallor, we live in a condition of "acute technosocial opacity" in which the interactions between science, technology, and human behaviour and social practices are unpredict-

able, and even more so when technological convergence – i.e. the combined effect of different technologies becoming blended and amplifying their powers and impact – may take place (Vallor, 2016, pp. 6 and 27). Risks, threats, and attractive opportunities are typically not self-evident and do not manifest themselves to individuals (Mitcham, 2003); on the contrary, they need to be teased out by pooling together a variety of expertise and viewpoints and by envisioning scenarios.

The capacity problem refers to lack of sufficient agency or power or resources or legitimacy. Individuals are part of value chains or networks as members of organisations and bearers of roles. That means that their freedom and power to act are both enabled and limited by their roles, existing duties, and "infrastructures of responsibility" (Scheffler, 2001, p. 125; Williams, 2006). The complexity of cooperation in contemporary society makes it necessary that a division of responsibilities between individuals and between organisations is established. The infrastructure of responsibility enables individuals and organisation to develop the expertise and credentials they need to operate effectively and gain trust, because it confers to them limited and manageable domains of responsibility and establishes mutual checks and balances. Overstepping one's boundaries means stepping on somebody else's feet and eliciting hostile reactions. When individuals have too broad responsibilities or when they try to fill gaps, results tend to go wrong.

> A basic reality of the modern world is that unmet responsibilities will not be addressed by individual initiatives, except insofar as those initiatives combine to found or restructure institutions (Williams, 2006, p. 214).

My account of the challenges of intrapersonal and team alignment, as well as my description of moral saturation, confirms the capacity problem. Further evidence is provided by Glerup and colleagues (2017), Spruit and colleagues (2016), and Swierstra and Jelsma (2006).

The ethical problem in relying on individual virtues is that it poses a heavy burden of responsibility on individuals without providing the training and the social environment needed to foster those virtues (Groves and Grinbaum at least acknowledge the lack of relevant moral training offered to researchers and innovators, 2013, p. 134). Despite the enormous academic revival of virtue ethics, studies and publications on how to train people to be virtuous are virtually non-existent notwithstanding the fact that Aristotle warned against the illusion that virtue can be acquired without practicing, just at an intellectual level (Aristotle, 1990, book II). Besides, some advocates of virtue ethics insist that the community or the social environment plays a very important role in acquiring virtues and sustaining virtuous behaviour (Blum, 1996; MacIntyre, 1982; Solomon, 1992; an idea supported by social psychology, e.g., Ross & Nisbett, 1991; for a very interesting discussion of caring behaviour within research groups, see Davies & Horst, 2015). So, if people operating in a given professional environment typically do not display the virtues that ensure the achievement of RRI objectives, it is unlikely that simply urging them to take more responsibilities and being

virtuous will work, just like it will not work to urge your kids to behave if they are hanging out with a bunch of louts.

5.3 Outline of a joint prospective responsibility fitting RRI ambitions

Since retrospective responsibility – especially when understood as blameworthiness and liability – and individual responsibility cannot ensure the transformative goals of RRI, I am proposing an outline of what needs to be the form of a joint prospective responsibility that could do the job. Pellé and Reber actually hinted at the necessity of this kind of responsibility:

> But what is also needed is that a collective and systemic responsibility, which would be one main objective of RI, emerges from individual responsible behaviour (2015, p. 115)

This suggestion comes out of their focus on responsibility in the context of supply chains. I fully agree with the idea that RRI requires a type of collective responsibility that applies at the interorganisational level, which I will call value-chain level. Yet there is an important difference between my goal and their suggestion. To me, the value-chain collective responsibility does not emerge from individual responsibilities; it is rather the other way round. First, a joint commitment to a shared goal and to the means needed to achieve it should be established. Then a collective responsibility will follow, and, finally, from this will emerge the role responsibilities and the contributory responsibilities of both collective actors and individuals (cf. Spruit et al., 2016, p. 882; Schwenkenbecher, 2021, pp. 20–23).

Philosophical discussions of collective responsibility mostly focus on three aspects of the problem: 1) an outcome that is important enough to elicit a moral call, 2) the possibility of responding to the moral call if individuals coordinate and act together – i.e. a prospective joint capacity, and 3) the absence of an agent or organisation that is capable of achieving the same result. The RRI case that we are considering has some similarities and some differences with this frame around which the philosophical discussion has focused. If we consider a network of organisations and teams involved in R&I, what I will call a *value chain*, the outcome could be a bad outcome that elicits a moral call of duty to avert it. For instance, it could cause significant harm to health or irreversible environmental damage. However, in many cases it is not something evident; it may be a risk, or even a more hypothetical eventuality. Furthermore, there will also be cases in which the outcome is not so dramatic, and what is at stake is the possibility of optimising outcomes or reducing externalities or achieving fairer distribution of benefits. So, the first aspect may be less obvious than in standard discussions about collective responsibility. The second aspect presents some important differences in our case. The first big difference is that the coordination problem is not among individuals but

among collective actors – e.g. organisations or teams. The second important difference is that the coordination problem is not about a specific action, but it is about introducing, adapting, and sustaining new processes. The third difference is that the capacity of achieving the intended outcome is more elusive to assess: the causal chain from the initiative to be taken now and the future success is non-linear and uncertain. So, what can be done is simply to attempt to be more capable of achieving a desirable outcome, not the expectation that if we coordinate, then we will definitely succeed. So, the focus is more on creating capabilities than on achieving an immediate result. The third aspect instead is equivalent: there are no other agents or networks that can care for the same outcomes.

Two challenges are prominent here. One is that existing obligations for organisations or teams are difficult to be traded against obligations to bring about uncertain benefits. As illustrated above when I introduced the concept of *infrastructures of responsibility*, collective entities are parts of systems of coordination and division of labour that create mutual expectations, dependencies, and obligations, which cannot simply be dropped. The second is that the entity of the expected gain often depends on the capacity of the organisations and teams to be flexible and revise their current commitments and obligations. Let's put it this way. If all the collective agents in the value chain continue to operate as they do now, we can expect outcomes ranging from X to X-. If actors undergo extensive redesign R+ of their operations, they can expect outcomes ranging from X to X++. But extensive redesign may not be feasible for some or all of them. So, an equilibrium point should be identified where the expected outcomes are still a sufficient gain to justify the costs of redesign. If there is a redesign R, which is currently feasible and could produce outcomes ranging from X to X+ that are considered a sufficient improvement, then a joint responsibility to pursue this redesign and the improved outcome is generated, and this in turn creates responsibilities, roles, and obligations for the collective agents and for the individuals within the organisations and teams. So, the creation of a joint obligation is not built on an existing capacity but on a prospective structure (I borrow this concept from van Lente & Rip, 1998) – i.e. a reconfigured value chain, whose members have also been reconfigured, with expected improved capacity to achieve desirable outcomes. Furthermore, this prospective structure, in the context of R&I, is designed to deal with uncertainty and unknowns and therefore must be flexible and adaptive. The normative basis of this joint prospective responsibility is not based on an actual capacity but rather on a commitment to try to develop that capacity. It is an *ought* not based on a *can* but on a *may (become able)*: a potentiality that requires and demands an effort, an intention, or better still a joint commitment.

5.4 The conditions under which a prospective joint responsibility for a value chain has normative force

Before I proceed to build the model of value chain responsibility, let me make explicit some assumptions.

1. The reason why RRI emerged is that currently the R&I system is producing sub-optimal outcomes: not enough benefits, not equitably distributed, and too many unintended consequences or externalities.
2. One of the causes of the sub-optimal results is that value chains are neither tightly coordinated, nor pursuing the same goals, nor sharing enough information and knowledge. This is the level at which RRI is meant to operate.
3. RRI aims at improving the integration, communication, and purpose-sharing of the whole value chain and of the public who bears the consequences of its outputs.
4. Collective, or, more precisely, joint prospective responsibility, has a goal-oriented nature; it is a mechanism to revise existing obligations and commitments to become better able to pursue desirable outcomes.

It is possible that the construction of joint prospective responsibility is the result of a spontaneous process. However, I will assume that there is an – individual or collective – agent that acts as a broker, whose task is to help a set of collective agents that constitute a value chain and some stakeholders to identify some unachieved potential, some superior outcome, and/or some shortcoming or injustices in their domain. Then an improved outcome is agreed as desirable and worth their effort. I will refer to this as the desirable goal(s). This provides a direction for an RRI project and the focus of the joint prospective responsibility of the value chain. It is the normative, prospective focal point. The members of the value chain form a collective intention and joint commitment to do their best to achieve the goal – including reshaping themselves.[10]

---

[10] This process works rather differently in the case of emerging technologies and well-established value chains. Innovation obviously happens in both cases, although often philosophers and STS scholars have overlooked the latter. While my sketch is meant to cover both cases, by doing so it is imprecise. In the case of emerging technologies, we see that the attempt to steer and oversee them carefully through soft and hard governance is already widespread. In this respect, the external steering is already there, but because of the extreme uncertainties, the process needs to be especially iterative and adaptive (in this it departs from my sketch). For established value chains instead, the external steering is not there, but there are several areas (e.g. chemical pollutants, social media) in which contestation is emerging either from civic activism, from research, from governmental bodies, or even from other private sectors (think of health insurances versus pharmaceutical industry, or law firms identifying areas where consumers may ask for compensations). These are, in my view, very promising areas where the contestation provides a good incentive for the value chain to engage with its critics to avoid reputational damages or compensations.

The next step is for the members of the value chain to explore a range of possible improvements in their internal structures and processes and in their interactions and co-ordination so that they work out the scope of their capacity to act in view of the realisation of – or approximation to – the desirable goal. Let's call this the *virtual improvement scope* that identifies a prospective optimal capacity to act together. This hypothetical target delimits the improvement frontier, which also represents the maximal optimisation of the value chain. It also affects the resources that can be mobilised for improving the value chain, because the expected gain affects the willingness to invest. So, if you want, the virtual improvement scope is the ideal value proposition for the restructuring of the value chain. However, a process of adaptation between the resources that can be committed, the changes that are contextually feasible, and the expected improvement needs to take place. This process determines the *contextual improvement scope*, i.e. what the collective agents can commit to do in the existing circumstances.

It remains to reappraise the goal expected from the contextual improvement scope and to evaluate if it is worth the effort – obviously it may not be. If the – not barely monetary! – cost-benefits balance is still favourable, all the normative requirements to create a joint prospective responsibility are in place. There is a desirable goal with normative force, a set of collective agents with the capacity to act together to pursue it, and a stated joint intention to do so. It follows that all actors in the value chain have compelling moral reasons – hence a responsibility – to contribute to the joint endeavour and to honour the commitment. All the ingredients to create a joint responsibility to pursue the goal are in place.[11]

---

11 I develop the philosophical details of this argument in a forthcoming paper. I have started talking about *collective* responsibility and finished claiming that the model I sketch can establish *joint* responsibility. From a philosophical point of view there is a difference. The reason I have initially used collective responsibility is that it is the concept commonly used in RRI discussions of responsibility. In the end, I have used joint responsibility which is a slightly weaker and less demanding concept, and I believe it is more appropriate when talking about a value chain, which is not a structured organisation with a formal decision-making mechanism but a collection of collective actors interacting. So, it is less controversial to attribute joint responsibility to the members of the value chain than collective responsibility to the value chain itself. For those interested in the philosophical aspects I refer them to Smiley (2014) and Schwenkenbecher (2018 and 2021).

Another philosophical point is the relation between *responsibilities* and *obligations*. One use of responsibility comes very close to the concept of obligations (think, for instance, of how hard to distinguish are the notions of parental responsibilities, duties, or obligations). Again, I explore this more fully elsewhere. Here I just note that I use *responsibility* as a more open-ended and discretionary concept then *obligation*. This means that responsibility can be used for (imperfect) duties that are broader, less defined in scope and implying the possibility of balancing different goods. Obligations instead is used for quite specific and strict (perfect) duties. For instance, one has an obligation to give notice, say, 3 months before resigning, while one has a responsibility to use efficiently and not to waste the lab resources.
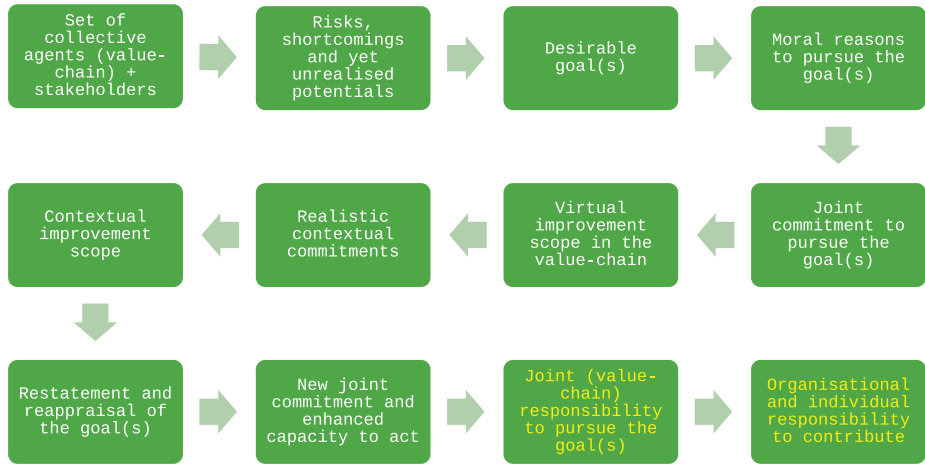
Figure 1: The conditions for having value-chain joint responsibility with normative force.

5.5 The model is not a practical tool

This model serves a theoretical, not a practical, purpose; hence it *is not* a solution for making RRI feasible and capable of living up to its ambitions and promises. It is neither a blueprint for implementation nor a roadmap. My model has mainly the philosophical purpose to show that the notion of a joint prospective responsibility for a plurality of collective agents – a value chain – is coherent and possible. A further point is that *only if* genuine forms of joint responsibility – like the one I sketched – can be made to work can RRI *then* live up to its transformative ambitions. The model attempts to identify the conditions of possibility for a joint responsibility among a collection of collective agents, i.e. the conditions under which such responsibility is normatively compelling and not mere rhetoric. It shows that while the notion is attractive, the conditions for achieving it are very demanding. Whether these conditions obtain in real-life circumstances depends on the specific features of each case and on the affordances of the context. In fact, what I have argued about values and what I have said about joint responsibility depending on prospective structures rather than on actual capacities suggest that for R&I value chains, *both the formation of joint intention and, even more, their execution are very challenging and fragile.*

I conclude that in the current circumstances RRI runs against the odds. My hypothesis is that the transformative ambitions of RRI cannot be realised at the systemic level: too strong are the trends against it – the dominance of the new public management paradigm, the intensification of competition and ranking, governments' obsession with using R&I as an engine of economic growth, the predominance of economic and quantitative indicators, the relentlessness of market competition, and the lobby-

ing power of vested interests. However, I suggest that if at least part of the value chain can be built or relocated outside the institutional contexts more exposed to adverse dynamics, then they may find some interstices where to grow a space (and perhaps eventually an ecosystem) more hospitable to RRI aspirations, to different ways of working, collaborating, sharing, and finding meaning. *Ad hoc,* experimental, small-scale value chains may be the *locus* for giving another chance to RRI and see whether a proof of concept for value chain joint responsibility can be achieved. So, despite the rather pessimistic implications of my arguments, I want to close with a more hopeful note coming from the words of Bob Jessop:

> we might see a revival of communities of scholars and students (serving wider communities too) who explore educational, scientific, and social innovations to make important, disinterested contributions to the intellectual commons and public good (Jessop, 2018, p. 108).

References

Adam, B., & Groves, C. (2011). Futures Tended: Care and Future-Oriented Responsibility. *Bulletin of Science, Technology & Society*, *31*(1), 17–27. https://doi.org/10.1177/0270467610391237

Anderson, E. (1995). *Value in Ethics and Economics*. Harvard University Press.

Aristotle. (1990). *Nicomachean Ethics* (Ross William David sir, Trans.). Oxford University Press.

Blum, L. (1996). Community and Virtue. In R. Crisp (Ed.), *How Should One Live?: Essays on the Virtues* (pp. 231–250). Oxford University Press.

Boenink, M., & Kudina, O. (2020). Values in responsible research and innovation: From entities to practices. *Journal of Responsible Innovation*, *7*(3), 450–470. https://doi.org/10.1080/23299460.2020.1806451

Borch, A., & Throne-Holst, H. (2021). Does Proof of Concept Trump All? RRI Dilemmas in Research Practices. *Science and Engineering Ethics*, *27*(1), 7. https://doi.org/10.1007/s11948-021-00288-8

Carrier, M., & Gartzlaff, M. (2020). Responsible research and innovation: Hopes and fears in the scientific community in Europe. *Journal of Responsible Innovation*, *7*(2), 149–169. https://doi.org/10.1080/23299460.2019.1692571

Checkoway, B. (1994). Paul Davidoff and Advocacy Planning in Retrospect. *Journal of the American Planning Association*, *60*(2), 139–143. https://doi.org/10.1080/01944369408975562

Davidoff, P. (1965). Advocacy and Pluralism in Planning. *Journal of the American Institute of Planners*, *31*(4), 331–338. https://doi.org/10.1080/01944366508978187

Davies, S. R., & Horst, M. (2015). Crafting the group: Care in research management. *Social Studies of Science*, *45*(3), 371–393. https://doi.org/10.1177/0306312715585820

De Grandis, G., Brass, I., & Farid, S. S. (2023). Is regulatory innovation fit for purpose? A case study of adaptive regulation for advanced biotherapeutics. *Regulation & Governance, 17*(3), 810–832. https://doi.org/10.1111/rego.12496

de Saille, S. (2015). Innovating innovation policy: The emergence of 'Responsible Research and Innovation.' *Journal of Responsible Innovation*, *2*(2), 152–168. https://doi.org/10.1080/23299460.2015.1045280

Debowski, S. (2012). *The New Academic: A Strategic Handbook*. McGraw-Hill Education (UK).

Dewey, J. (1983). *The Middle Works of John Dewey, Volume 14, 1899 – 1924*: *Human Nature and Conduct 1922* (J. A. Boydston, Ed.). Southern Illinois University Press.

Emmet, D. M. (1966). *Rules, roles and relations* (Vol. 149). MacMillan. https://ixtheo.de/Record/195014421

European Commission. (2024). *Sustainable wood value chains | Programme | H2020*. CORDIS | European Commission. https://cordis.europa.eu/programme/id/H2020_LC-RUR-11-2019-2020

Felt, U. (2018). Responsible research and innovation. In *Routledge Handbook of Genomics, Health and Society* (pp. 108–116). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315451695-14/responsible-research-innovation-ulrike-felt

Gill, R. (2010). Breaking the silence: The hidden injuries of neo-liberal academia. In R. Ryan-Flood & R. Gill (Eds.), *Secrecy and Silence in the Research Process: Feminist Reflections* (pp. 228–244). Routledge.

Glerup, C., Davies, S. R., & Horst, M. (2017). 'Nothing really responsible goes on here': Scientists' experience and practice of responsibility. *Journal of Responsible Innovation*, *4*(3), 319–336. https://doi.org/10.1080/23299460.2017.1378462

Grinbaum, A., & Groves, C. (2013). What Is "Responsible" about Responsible Innovation? Understanding the Ethical Issues. In R. Owen, J. Bessant, & M. Heintz (Eds.), *Responsible Innovation* (1st ed., pp. 119–142). Wiley. https://doi.org/10.1002/9781118551424.ch7

Hämäläinen, N. (2016). *Descriptive Ethics: What does Moral Philosophy Know about Morality?* Palgrave Macmillan.

Hirsch, G., Trusheim, M., Cobbs, E., Bala, M., Garner, S., Hartman, D., Isaacs, K., Lumpkin, M., Lim, R., Oye, K., Pezalla, E., Saltonstall, P., & Selker, H. (2016). Adaptive Biomedical Innovation: Evolving Our Global System to Sustainably and Safely Bring New Medicines to Patients in Need. *Clinical Pharmacology & Therapeutics*, *100*(6), 685–698. https://doi.org/10.1002/cpt.509

Honig, P., & Hirsch, G. (2016). Adaptive Biomedical Innovation. *Clinical Pharmacology & Therapeutics*, *100*(6), 574–578. https://doi.org/10.1002/cpt.513

Jessop, B. (2018). On academic capitalism. *Critical Policy Studies*, *12*(1), 104–109. https://doi.org/10.1080/19460171.2017.1403342

Lowndes, V., Pratchett, L., & Stoker, G. (2001). Trends in Public Participation: Part 2 – Citizens' Perspectives. *Public Administration*, *79*(2), 445–455. https://doi.org/10.1111/1467-9299.00264

MacIntyre, A. (1982). *After Virtue*. University of Notre Dame Press.

Mitcham, C. (2003). Co-responsibility for research integrity. *Science and Engineering Ethics*, *9*(2), 273–290. https://doi.org/10.1007/s11948-003-0014-0

Oakeshott, Michael. (1962). The Tower of Babel. In *Rationalism in Politics and Other Essays* (pp. 59–79). Methuen & Co.

O'Neill, O. (2002). *A Question of Trust: The BBC Reith Lectures 2002*. Cambridge University Press.

Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy*, *39*(6), 751–760. https://doi.org/10.1093/scipol/scs093

Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013). A Framework for Responsible Innovation. In R. Owen, J. Bessant, & M. Heintz (Eds.), *Responsible Innovation* (1st ed., pp. 27–50). Wiley. https://doi.org/10.1002/9781118551424.ch2

Pellé, S., & Reber, B. (2015). Responsible innovation in the light of moral responsibility. *Journal on Chain and Network Science*, *15*(2), 107–118. https://doi.org/10.3920/JCNS2014.x017

Richardson, H. S. (1999). Institutionally Divided Moral Responsibility*: Henry S. Richardson. *Social Philosophy and Policy*, *16*(2), 218–249. https://doi.org/10.1017/s0265052500002454

Rome Declaration on Responsible Research and Innovation in Europe. (2014, November 25). https://digital-strategy.ec.europa.eu/en/library/rome-declaration-responsible-research-and-innovation-europe

Ross, L., & Nisbett, R. E. (1991). *The Person and the Situation: Perspectives of Social Psychology*. McGraw-Hill.

Sager, T. (2022). Advocacy planning: Were expectations fulfilled? *Planning Perspectives*, *37*(6), 1205–1230. https://doi.org/10.1080/02665433.2022.2040189

Sayer, A. (2011). *Why Things Matter to People: Social Science, Values and Ethical Life*. Cambridge University Press. https://doi.org/10.1017/CBO9780511734779

Scheffler, S. (2001). *Boundaries and allegiances: Problems of justice and responsibility in liberal thought*. Oxford University Press, USA.

Schwenkenbecher, A. (2018). Making Sense of Collective Moral Obligations: A Comparison of Existing Approaches. In K. Hess, V. Igneski, & T. L. Isaacs (Eds.), *Collectivity: Ontology, Ethics, and Social Justice* (pp. 109–132). Rowman and Littlefield.

Schwenkenbecher, A. (2021). *Getting Our Act Together: A Theory of Collective Moral Obligations*. Routledge.

Shelley-Egan, C., Bowman, D. M., & Robinson, D. K. R. (2018). Devices of Responsibility: Over a Decade of Responsible Research and Innovation Initiatives for Nanotechnologies. *Science and Engineering Ethics*, *24*(6), 1719–1746. https://doi.org/10.1007/s11948-017-9978-z

Simon, H. A. (1976). *Administrative Behavior: A Study of Decision-making Processes in Administrative Organization*. Free Press.

Smiley, M. (2014). Future-looking collective responsibility: A preliminary analysis. *Midwest Studies in Philosophy*, *38*, 1–11.

Solomon, R. C. (1992). Corporate roles, personal virtues: An Aristotelean approach to business ethics. *Business Ethics Quarterly*, *2*(3), 317–339.

Spruit, S. L., Hoople, G. D., & Rolfe, D. A. (2016). Just a Cog in the Machine? The Individual Responsibility of Researchers in Nanotechnology is a Duty to Collectivize. *Science and Engineering Ethics*, *22*(3), 871–887. https://doi.org/10.1007/s11948-015-9718-1

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008

Swierstra, T., & Jelsma, J. (2006). Responsibility without Moralism in Technoscientific Design Practice. *Science, Technology, & Human Values*, *31*(3), 309–332. https://doi.org/10.1177/0162243905285844

Tabarés, R., Loeber, A., Nieminen, M., Bernstein, M. J., Griessler, E., Blok, V., Cohen, J., Hönigmayer, H., Wunderle, U., & Frankus, E. (2022). Challenges in the implementation of responsible research and innovation across Horizon 2020. *Journal of Responsible Innovation*, *9*(3), 291–314. https://doi.org/10.1080/23299460.2022.2101211

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press. https://www.google.com/books?hl=en&lr=&id=RaCkDAAAQBAJ&oi=fnd&pg=PP1&dq=vallor+technology+and+the+virtues&ots=DEUoSSk88k&sig=0yO9idb7QmG6TxWHzPpdlFEY924

van de Poel, I., & Sand, M. (2021). Varieties of responsibility: Two problems of responsible innovation. *Synthese*, *198*, 4769–4787. https://doi.org/10.1007/s11229-018-01951-7

van Lente, H., & Rip, A. (1998). Chapter 7. Expectations in Technological Developments: An Example of Prospective Structures to be Filled in by Agency. In C. Disco & B. V. D. Meulen (Eds.), *Getting New Technologies Together*. DE GRUYTER. https://doi.org/10.1515/9783110810721.203

Viano, C. A. (2002). *Etica pubblica*. Laterza.

Von Schomberg, R. (2007). From the ethics of technology towards an ethics of knowledge policy & knowledge assessment. *Available at SSRN 2436380*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2436380

Von Schomberg, R. (2011). *Towards responsible research and innovation in the information and communication technologies and security technologies fields*. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/58723

Waldorff, S. B., & Madsen, M. H. (2023). Translating to Maintain Existing Practices: Micro-tactics in the implementation of a new management concept. *Organization Studies*, *44*(3), 427–450. https://doi.org/10.1177/01708406221112475

Wäscher, S., Biller-Andorno, N., & Deplazes-Zemp, A. (2020). "I Don't Want to Do Anything Bad." Perspectives on Scientific Responsibility: Results from a Qualitative Interview Study with Senior Scientists. *NanoEthics*, *14*(2), 135–153. https://doi.org/10.1007/s11569-020-00365-5

Williams, G. (2006). 'Infrastructures of Responsibility': The Moral Tasks of Institutions. *Journal of Applied Philosophy*, *23*(2), 207–221. https://doi.org/10.1111/j.1468-5930.2006.00330.x

Wittrock, C., Forsberg, E.-M., Pols, A., Macnaghten, P., & Ludwig, D. (2021). *Implementing Responsible Research and Innovation: Organisational and National Conditions*. Springer International Publishing. https://doi.org/10.1007/978-3-030-54286-3

Ylijoki, O.-H. (2003). Entangled in academic capitalism? A case-study on changing ideals and practices of university research. *Higher Education*, *35*, 307–335.

Anne Blanchard and Erik Bjørnerud
Chapter 5
Navigating tensions around RRI in higher education

If one is truly to succeed in leading a person to a specific place, one must first and foremost take care to find him where he is and begin there. This is the secret in the entire art of helping. Søren Kierkegaard, *The Point of View for My Work as an Author*, published posthumously in 1859.

In this chapter, we wish to contribute with our own experience in designing and teaching a series of PhD courses on the topic of RRI, as part of the AFINO research school. We want to share our thinking behind designing these RRI courses, and mostly, the tensions met by the post-graduate participants and our approach to help them navigate these tensions. To this aim, we will, in Section 2, look at what 'teaching RRI' means for us in the context of the AFINO research school and introduce the literature that has inspired us in shaping our PhD courses. We then move onto the challenges of RRI in higher education, and in Section 3, we comment on the *tensions* expressed by the PhD course participants over the years, namely, navigating short-term temporalities, the expectation for quick results, narrow merit-based criteria, and the gap between intentions and outcomes in RRI projects. These tensions, we observe, mostly stem from conflicting demands from research and policy environments and lead early career researchers to internalise narrow criteria for what counts as a 'successful' RRI project and what could be accomplished in the duration of a PhD or post-doctoral project. What we discussed throughout the AFINO research school is that a central key to navigating these tensions and gaps is to nurture a *caring and supportive RRI community*, based on slow spaces for building relationships and reflections, for being attentive to the sometimes-conflicting demands stemming from research, policy, and society, and for reconnecting with the essence of RRI, as an iterative learning process with deeply transformative aims. Finally, in Section 4, we conclude with our early reflections on how exploring these tensions is helpful in shaping new hybrid and anti-fragile partnerships across RRI policy and research environments specifically, in order to discuss and address more profoundly the tensions and gaps in RRI projects.

1 Introduction: a call for RRI in higher education

In the previous chapters, our colleagues have highlighted the importance of aligning research and innovation processes with broader societal interests, needs, and values. They have looked at Corporate Social Responsibility and Responsible Research and Innovation as ways of addressing pervasive, complex, and pressing challenges. This chapter focuses on RRI and furthers our colleagues' reflections by looking at *postgraduate RRI education* as an important way of addressing those challenges and nurturing RRI awareness and practices among scientists. In particular, we will look at the tensions experienced by early career researchers involved in RRI projects and our reflections on how these can foster new hybrid, anti-fragile partnerships around RRI.

1.1 RRI as a way to address global societal challenges

As stated in earlier chapters, RRI emerged in the 2010s as a prominent policy principle to give a socially responsible 'direction' to science (and in particular technoscience[1]). This call for RRI is present in many recent European and Norwegian research policy documents and backed up by a significant body of literature, which encourage research and innovation processes to be increasingly socially responsible and focused on sustainability challenges (see, e.g., Arnold et al., 2019; OECD STI, 2021; OECD STI, 2022; RCN, 2020; Tassone et al., 2018). Focusing on the Norwegian context that is ours, the 2015–2020 Strategy for the Research Council of Norway (RCN, 2015) asserted how the Research Council (RCN) should focus on promoting research and innovation activities that will "yield benefits for society at large in the long term" (p.38). To do so, the RCN should ensure that research and innovation are conducted in a socially responsible way and that greater importance is attached to how research might contribute to solving the grand societal challenges. Following that Strategy document, the RCN developed its framework for responsible innovation (RCN, 2019) based on the four dimensions of RRI of Stilgoe et al. (2013). It frames the demands they place on Norwegian research and innovation (R&I) systems:

1. Acknowledging that technologies can fundamentally change the context they are embedded in, R&I processes are expected to **anticipate** their intricate and dynamic impacts on society.
2. R&I systems need to be **reflexive** about their often-implicit agendas and assumptions and not cover up the uncertainties and limitations of their approaches.

---

1 Here, we understand 'technoscience' as described by Thomas Völker and his colleagues: "sciences that become technologies, such as biotechnology, nanotechnology, and information and communication technologies" (2024; p.10)

3. In addition to engaging in societal dialogue in its different forms, the demand for **inclusion** is now increasingly directed inwards, within the research environment itself. Interdisciplinarity is encouraged, as a way for R&I systems to build a more relevant and holistic knowledge base, as well as reflexively look at one's disciplinary blind spots.
4. R&I activities that are anticipatory, reflexive, and inclusive entail new, more **responsive** governance practices, through, for instance, collaborating with partners that may be affected by a research and innovation process involving different disciplines, policy actors, trade and industry, various interest organisations, and society.

This framework for responsible innovation is based on a fundamental shift from linear models of science-society relations (assuming that there is a straight line from basic to applied research to the development of products and services) to more dynamic and interactive models, sensitive to the context of application, and encouraging new networks across disciplines and society at large. Indeed, when we understand research and innovation processes as embedded in social, cultural, and historical aspects in complex and dynamic ways, then the governance of R&I based on a well-defined distribution of tasks between research, technology, innovation, and policy becomes obsolete (RCN, 2019). Accordingly, this framework places new demands on "the knowledge base, expertise, capacities and skills in the research and innovation systems – both at an individual and institutional level" (RCN, 2019, p.4) and stimulates reflections on how to realise socially responsible R&I processes.

1.2 A strong call for RRI in higher education, but an ambiguous context of application

The ambitions for research and innovation to be increasingly socially responsible represent a challenging turn for higher education. It implies, as we mentioned above, the creation of new knowledge and skills and the organisation of research and innovation processes in different ways. In this context, the theme of *training post-graduates in RRI* has become a central topic for the RRI community (Mejlgaard et al., 2019; Stahl et al. 2023; Tassone et al., 2018), with many considering higher education as a promising way to nurture RRI awareness and practices among researchers. In particular, Tassone et al. (2018) have discussed the need for a new contract between science and society and encouraged a tighter interplay between the academic community (specifically), innovation, and society at large. Along the same lines, Stahl et al. (2023) recently claimed that the integration of responsible innovation in post-graduate education was a key prerequisite to the successful implementation of RRI. This call for RRI in higher education has also been taken up by research policy, with, for instance, the report commissioned by the RCN (Arnold et al., 2019), prompting higher education to address

the failure[2] to engage, in constructive ways, grand societal challenges. This report states that the RCN should make efforts to engage "with the demand side or broader socio-technical systems" (Arnold et al., p.2) and further support "research and innovation that can confront an array of societal challenges" (p.1). The current 2020–2024 Strategy for the RCN (RCN; 2020) follows these lines by stating that ethical and socially responsible research and innovation is one of its five main aims. More precisely:

> *The Research Council will work to ensure that research and innovation activities are carried out in compliance with recognised research ethics standards and within a socially responsible framework. This entails taking steps to anticipate long-term ramifications of these activities for society, including any unintended side effects; exploring opportunities and dilemmas in collaboration with users; and adapting activities based on learning along the way. Gender and diversity perspectives must be carefully assessed and integrated where relevant. Different research fields will require different approaches. For example, broad-based user involvement will be very important in certain fields, but less applicable to others. (RCN, 2020; p.22)*

Teaching RRI at the higher education level is thus considered an important way to nurture RRI awareness and encourage responsible practices in research and innovation among scientists and professionals. However, training post-graduates in RRI faces important challenges, and the strong focus on RRI in higher education in research policy does not prevent some vagueness and ambiguity when it comes to how to put these ideas into practice. A critical challenge lies in the fact that the concept of responsibility is defined differently across various disciplines and by the actors of research and innovation processes. The absence of a universally accepted definition of RRI can pose a challenge in structuring courses in a comprehensive way and might therefore prevent cross-institutional learning (Rip and Voß, 2013; Fisher and Rip 2013; Macnaghten et al. 2014; Mejlgaard et al., 2019; Owen et al. 2013; Ribeiro et al. 2016). In other words, what teaching RRI should look like in practice is fairly vague. As an example of this ambiguity and difficulty in balancing different RRI definitions, the above excerpt from the RCN relies altogether on the four overarching dimensions of RRI by Stilgoe et al. (2013) – 'anticipating' long-term ramifications and 'collaborating' with users through an 'adaptive' learning process – and also on the six actionable RRI keys developed by the European Commission, in particular, 'ethics', 'user involvement', or 'gender perspectives'.[3] To add to the challenges, creating new RRI training programmes for post-graduates is resource

---

2 This failure is also called 'transformational failure' and has motivated a "third generation" of research and innovation (R&I) policy, which is especially concerned with R&I engaging grand societal challenges in 'constructive' and responsible ways. See the introductory chapter of this book or Arnold et al. (2019) and Guldbrandsen (2019) for a description of the first, second, and third generations of R&I policies.

3 In its 8th framework programme 'Horizon 2020', the European Commission has translated the four RRI dimensions by Stilgoe et al. (2013) into the following six keys: ethics, public engagement, science education, open access, gender equality, and governance, with a wish to be more concrete or actionable. What is meant by those different keys, however, in particular 'governance', remains rather ab-

and time demanding; it demands to fit in with existing disciplinary programmes to ensure a greater focus on responsibility in research and innovation while still being relevant to the disciplinary field post-graduates come from (Mejlgaard et al., 2019). Further, post-graduate RRI training is, in some settings, difficult to justify, as RRI is often considered a peripheral activity to 'real' research (Mejlgaard et al., 2016; Stahl et al., 2023). And to finish on a deeper, more pervasive challenge, the resistance and lack of support at the level of institutions (especially universities) can also come from the perception of RRI not only as a waste of time and resources but also as "a threat to the socialisation of students into the ideals of universal, value-free objectivity and the disinterestedness of science" (Mejlgaard et al., 2019; p. 611; see also the discussion of transdisciplinarity in Chapter 2).

Against this challenging background, we contribute, in this chapter, with our own experience in designing and teaching a series of PhD courses on the topic of RRI, as part of the AFINO research school. We want to share our thinking behind designing these RRI courses, and mostly, the tensions met by the post-graduate participants and our approach to help them navigate these tensions. To this aim, we will, in Section 2, look at what 'teaching RRI' means for us in the context of the AFINO research school and introduce the literature that has inspired us in shaping the PhD courses. We then move onto the challenges of RRI in higher education, and in Section 3, we comment on the *tensions* expressed by the PhD course participants over the years, namely, navigating short-term temporalities, the expectation for quick results, narrow merit-based criteria, and the gap between intentions and outcomes in RRI projects. These tensions, we observe, mostly stem from conflicting demands from research and policy environments and lead early career researchers to internalise narrow criteria for what counts as a 'successful' RRI project and what could be accomplished in the duration of a PhD or post-doctoral project. What we discussed throughout the AFINO research school is that a central key to navigating these tensions and gaps is to nurture a *caring and supportive RRI community*, based on slow spaces for building relationships and reflections, for being attentive to the sometimes-conflicting demands stemming from research, policy, and society and for reconnecting with the essence of RRI, as an iterative learning process with deeply transformative aims. Finally, in Section 4, we conclude with our initial reflections on how exploring these tensions is helpful in shaping new hybrid and anti-fragile partnerships across RRI policy and research environments specifically, in order to discuss and address more profoundly the tensions and gaps in RRI projects.

---

stract, but they have nevertheless been translated into indicators whereby RRI projects can be evaluated (see Delaney et al. (2020, p. 23).

2 Teaching RRI in the AFINO research school: starting from the tensions and learning to navigate them

2.1 Three broad ways of approaching RRI in higher education

In this ambiguous context of application, we can roughly delineate three approaches to teaching RRI, in which we will situate our own approach. First, the training of early career researchers in RRI has frequently taken the shape of a 'tool-kit approach' (Hesjedal et al., 2020), consisting, for instance, in introducing the different definitions of RRI, including the European Commission's operationalisation of RRI into six keys (Mejlgaard et al., 2016). Alongside, there has been a growing body of literature offering different kinds of recommendations, overarching principles, and roadmaps for teaching RRI, such as "focus on being relevant", "focus on being reflexive", and "foster an ethics of care" (Tassone et al., 2018).

Second, teaching RRI can go beyond those 'tool-kit' approaches, reaching into more enlightening efforts, for instance, by introducing early career researchers to theories and models of science-society relations and positioning themselves in these. This helps elicit a deeper comprehension of the intricate connections between science and society (Mejlgaard et al., 2016). By encouraging early career researchers to critically look at their own academic discipline and discuss the broader social, ethical, political, and economic aspects of their research, they can arrive at a more subtle, reflexive, and interdisciplinary interpretation of their own responsibilities within this context (Hesjedal et al., 2020; Mejlgaard et al., 2016).

Third, teaching RRI to post-graduates can take these reflexive and interdisciplinary efforts even further, by focusing on 'dislocatory moments' to encourage 'double-loop learning'. What Åm (2019) means by 'dislocatory moments' is the "responses to moments when science–society-related concerns emerge that require broader reflection, societal debate or possible changes of research practices" (p.458). In other words, these moments trigger a realisation that the usual practices in research and innovation no longer work for certain challenges and that these practices can be opened up and changed (Nulli & Stahl 2018). Hesjedal and colleagues (2020) have built on dislocatory moments and explored how they can encourage 'double-loop learning' or, in other words, learning which allows one not only to take in new knowledge but also to rethink, change, and revise one's own practices, values, and ways of knowing.

2.2 Teaching RRI in the AFINO research school

In designing the AFINO research school, we have been profoundly inspired by our colleagues Maria Hesjedal, Roger Strand, Heidrun Åm, and others (Hesjedal et al.,

2020, and Åm, 2019), who have worked with RRI, dislocatory moments, and double-loop learning in the 'sister-centre' of AFINO: Digital Life Norway (DLN).[4] On our side, rather than focusing on triggering dislocatory moments in our research school activities – as it is probably these moments that have brought most participants to us in the first place (see below) – we have wished to *start* from these moments, *start from the early career researchers' feeling of inadequacy and tension* between what they were supposed to deliver in their PhD or post-doctoral project and their critical and motivated commitment to responsibility and care in their research. We wanted to use the AFINO research school as a space to further unpack and explore where these dislocations and tensions were situated, and how we could best navigate them, in the ambitious and ambiguous policy context briefly described above.

This demanded, according to us, to create a deeply reflexive and slow teaching and learning space, to counterbalance the speed and pressures experienced by the participants who are, in their daily work lives, left with little time to connect with one another, share their experienced tensions, and think creatively about how to navigate them. We moved beyond the 'tool kit'–type approach to teaching, as our idea was to build a community of 'resilient RRI researchers' that could share, discuss, and navigate the tensions between research, policy, and society mentioned above and reconnect with what we consider to be the essence of RRI – a critical learning process pursuing deeply transformative aims.

How did we do this in practice? The AFINO centre has been funded by the Research Council of Norway to experiment with putting into practice third-generation research and innovation policy (or transformative innovation: see footnote 2) and draw lessons for how to more constructively address grand societal challenges through RRI. Within the AFINO centre, the research school has had the specific mandate of focusing on transforming research and innovation praxis in the direction of increased sustainability and social responsibility.

Since 2021, the AFINO research school has consisted in a series of yearly PhD courses, designed for early career researchers involved in RRI projects (or having an interest in RRI). Most of the participants were early career researchers (PhD candidates and post-doctoral researchers) part of the AFINO network.[5] The majority of participants had a background in social sciences and humanities. They worked on critical, empirical

---

4 Digital Life Norway (DLN) is a Centre funded by the Research Council of Norway under the same call as AFINO, focusing on biotechnology research and innovation in a transdisciplinary and responsible way. The research schools in both AFINO and DLN share the same mandate: to support transformation in the direction of sustainability and socially responsible research and innovation. For more information, see: https://www.digitallifenorway.org

5 This means that the participants either (a) were working on a project that was affiliated with AFINO (for instance, we had participants from the 'BREAD' project, looking at the responsible governance for addressing food waste) or (b) were part of the 'extended network' of AFINO and took part in its various networking activities (such as workshops, conferences to present their findings, or indeed the research school activities).

projects concerned with, for instance, responsible narrative and qualitative research with local communities and participatory and inclusive science to address challenges stemming from ecosystem collapse and used various perspectives such as deep ecology, eco-feminism, or decolonial and anti-domination approaches to socio-technical futures. They expressed a strong commitment to contexts characterised by uncertainty, complexity, and controversy and were aware that science does not only represent the world but also intervenes in it (Hacking, 1983), notably by contributing to extreme environmental, social, and systemic imbalances and challenges. Their backgrounds, commitments, and interests largely explain why these participants were so finely aware of the importance of making their praxis more caring and aligned with social needs and values, and why they had experienced dislocatory moments beforehand (in particular, we will explore in Section 3 the tension expressed by the participants between wanting to pursue responsible and caring research in an environment that demands fast and measurable results). A smaller number of participants had a STEM background (many of them were affiliated with our sister-centre DLN) and were part of projects concerned with biotechnology research and innovation, AI with medical applications, machine learning, and deep learning for modelling and understanding complex biological systems, CRISPR-salmon and gene-editing for food production, or imaging technologies applied to circular economy, only to name a few. As a whole, the participants in the AFINO PhD courses represented a heterogeneous group with various backgrounds and experiences, different understandings of responsibility, and diverse experiences with moments of tension (see Section 3). There were between 14 and 20 participants in each PhD course.

There have been three AFINO PhD courses so far, with an upcoming one planned for autumn 2024. We will not enter into the detailed programmes and outcomes of each (these are reported on the AFINO website[6]), but what is important to note is that we designed the courses to start from the expressed needs and topics of interests of the participants (for instance, the challenging implementation and evaluation of RRI, or how to balance RRI and the demand for measurable outcomes). The courses were also designed to fit closely to the lived reality of the early career researchers working with (or interested in) RRI, as well as the tensions and challenges they faced in their everyday R&I practices. This was made possible through collecting, several weeks before each PhD course, the participants' expectations and needs for the course. Further, at the end of each PhD course, we would collect the participants' feedback and ideas for themes for the next PhD course. After the first PhD course, and having spent four days together in a retreat-like atmosphere, we got to know the participants fairly well. In addition, there has been a core group of 6–8 participants attending several of the AFINO PhD courses. That contributed to a feeling of community among the partic-

---

6 Reports of the AFINO PhD courses can be found on the AFINO website: https://www.ntnu.edu/web/afino/summer-schools

ipants, and it helped us get a close understanding of their challenges and lived reality so that we could design subsequent PhD courses building on the previous ones.

The first PhD course took place in 2021 in Bekkjarvik (Norway) for a duration of 4 days. It happened not long after the pandemic-related restrictions, and it was particularly important for us to design this first PhD course like a retreat – a slow space to connect with one another, in a very natural and quiet setting, accessible by boat, where the participants could swim, where they were offered yoga classes and walks, and where we ensured that there was enough free time for them to share around their challenges, be creative, or, in other words, build this sense of RRI community. The topic was 'the challenges of transdisciplinary RRI', and we focused on three of those challenges: the concrete implementation of RRI, choosing relevant methods for RRI, and the evaluation of RRI. The days were split between interactive lectures on the topic of those challenges and group work, where the participants were asked to represent RRI-related challenges in creative ways.[7] This first PhD course created an atmosphere of trust, enjoyment, and deep reflections that set the tone for the subsequent PhD courses.

The second AFINO PhD course took place in 2022 in Oslo, over 3 days, and gathered both post-graduates from the AFINO and DLN centres. The topic of that course was grounded in the previous discussions relative to RRI challenges and addressed the question of how to 'engage in critical research within institutions'. In particular, we discussed the mismatch or gap between the initial intentions for engaging with an RRI project (engaging in participatory processes or interdisciplinary research, for instance) and the final outcomes that often don't match those initial intentions and can even be considered a failure (see Section 4). The participants explored the importance of relying on an RRI community when navigating critical, responsible research that doesn't conform to imperatives of narrow criteria for excellence, productivity, or the delivery of rapid and measurable results. To address this topic, we wanted a more heterogenous pool of participants to have access to a broader array of RRI practices and experiences in navigating RRI-related challenges. This is why we co-organised this PhD course between the AFINO (participants mainly from SSH backgrounds) and DLN centres (participants mainly from STEM backgrounds). For practical reasons, as we invited participants from both DLN and AFINO, we decided to organise the course in Oslo and over a shorter period of time. That meant that we were far from the quiet and retreat-like atmosphere of the first PhD course. In addition, because of the more heterogenous group and a setting that did not promote really getting to know each other, participants who attended the first PhD course voiced that the feeling of connecting around shared challenges and tensions was less present this time. However, we all had insights into a broad range of RRI practices, and the group work developed

---

7 All the creative contributions of the participants from the three AFINO PhD courses can be found on the AFINO website, in the 'RRI gardens': https://www.ntnu.edu/web/afino/the-rri-garden

into reflections on the importance of having this community of support when engaging in critical research.

Finally, the third AFINO PhD course, designed, like the first one, as a retreat, took place in Jondal (Norway) over 4 days. Following the same organisation (interactive lectures in the mornings and group work in the afternoon), we also included a walk-shop (Wickson et al., 2015) to take our reflections outside. Having once again collected the ideas and wishes of the participants beforehand, we landed on the topic of 'reinventing RRI', to build on previous discussions about how to pursue critical, engaged, responsible research when also having imperatives for productivity and measurable outcomes. The idea here was to give an opportunity to the participants to make RRI 'their own', and how they could best adapt their representation of RRI in a constructive way to their projects. They therefore each worked on the dimension of RRI that was, according to them, the most important to their project (either inclusion, reflexivity, anticipation, or responsiveness) and reflected on how they could 'reinvent' it to implement it in their project in a relevant manner.[8]

Now that we have shared what was important to us in teaching RRI to postgraduates, and how we organised the AFINO PhD courses, we will look into the specific moments of tension or dislocation that the participants encounter in their everyday practice of RRI. This will then lead us to discuss the gap they experience, between their intentions and the outcomes of their project, and to our concluding remarks on the need for supportive hybrid, anti-fragile RRI communities involving both research and policy environments in an effort to address these tensions and gaps.

3 Moments of tension of early career researchers when engaging with RRI

Before going into the moments of tension experienced by the AFINO PhD course participants, it is important to note that what is reported in this section is the experience of a unique set of participants who took RRI seriously[9] and did not always have a smooth experience with it. The participants indeed all had an interest in RRI, which extended for many to a genuine commitment to RRI (see footnote 9). The courses

---

8 The outcomes can be found in the 'RRI gardens' on the AFINO website: https://www.ntnu.edu/web/afino/the-rri-garden

9 The participants whose projects were affiliated to the Centre for Digital Life (DLN) were so on the condition that they include RRI in their practice. The participants who were in projects affiliated to AFINO had an RRI component built into their projects and attended the PhD courses to learn more about RRI or get help in implementing RRI. The other participants, who were not specifically working on or with RRI, nevertheless had critical topics that demanded their research to be 'responsible' and 'caring'.

were not mandatory but were offered as a space to discuss creatively and practically the theory and practice of RRI. So, the participants' motivation, interest, and eagerness were genuine. It was felt in the creative and interactive atmosphere of the courses[10] and the expressed feeling of a 'regained RRI community': "I think the highlight of the summer school is just to listen and to be here. Everyone contributes collectively to a safe space where we exchange, discuss and listen. Everyone is so engaged, and it is beautiful to be a part of a research community that is capable of that" (participant in the first PhD course, 2021). In addition, participants appreciated the space and time dedicated to reflections on their own context of application of RRI, in their own projects: it was "nice to have a whole different set of eyes to look at my project" (participant to the third PhD course, 2023).

The participants' interest and commitment to RRI seem to naturally fit the demands coming from research policy and funding institutions (see Section 1), with perhaps the most common concern among the participants being the engagement with social actors and concerns and values, or indeed "academic citizenship", where anticipation, future's literacy, and responsiveness also play a key role. However, throughout the AFINO PhD courses, participants shared several other challenges and moments of tension. These tensions are notably between the demands of RRI stemming from policy contexts and the demands that come from the current working of the research system, with its funding cycles, career patterns, and evaluation and reward systems. They relate to how to concretely implement RRI in a context of conflicting demands and expectations, in particular, in terms of temporalities, research organisation, reward systems, and norms of scientific merit.

3.1 Tensions in navigating short-term temporalities and the expectation for quick results

Throughout the PhD courses, participants shared the significant and ongoing pressures they felt to be productive and deliver fast and measurable results as expected outcomes of their PhD or post-doctoral projects. These short-term temporalities reveal a form of policy and governance of research that is shaped by new public management regimes, where short-term goals and achievements are often prioritised to meet the need for rapid results, demands for marketable products, constant auditing, or increasing funding constraints. Short-term temporalities further create the idea that RRI somehow must be a controlled activity, neatly organised in work packages with

---

10 See, for instance, the interviews by Eva Murvold (AFINO communication advisor) of participants who report on their reflections and experiences of the first PhD course: https://www.ntnu.no/blogger/afino/2021/09/02/nurturing-a-vibrant-rri-community/; second PhD course: https://www.ntnu.no/blogger/afino/2022/09/22/the-challenging-gap/; and third PhD course: https://www.ntnu.no/blogger/afino/2023/10/04/with-a-feeling-of-an-rri-community/

planned milestones so that it can result in predictable outcomes, expected to be visible and measurable after only a short-term period. This fragmented and short-term organisation of R&I, and the expectations for quick results, are also discussed in the literature, with, for instance, Felt (2021) explaining how imaginaries around temporalities, structures, and practices play an important role in shaping and giving a direction to academic research. She notably describes research projects as key 'time generators', according to which the production of knowledge is organised into discrete units of time, generally 2–4 years (Felt, 2021). These short-term temporalities explicitly (but also tacitly) permeate and govern the way we do RRI in practice and translate into high expectations for post-graduates, such as numerous, highly cited publications or rapid solutions to address global societal challenges.

This short-term thinking is, however, not supportive of RRI in practice, as the latter requires a longer-term perspective and commitment in order to address complex challenges in a constructive way. In particular, the iterative nature of interdisciplinary or transdisciplinary processes that one might engage with when doing RRI demands to be factored in from the beginning (Borch and Throne-Holst, 2021) and necessitates more time than classical disciplinary work: time for building relations and trust, a common language, a common way of understanding problems and discussing potential solutions (Felt et al., 2015). This tension related to temporalities came clearly through in the AFINO PhD course in 2022, where we addressed the topic of 'undertaking critical research within institutions'. The participants mentioned two particularly central challenges related to navigating these different temporalities and demands for quick outcomes.

First, some participants reflected on how the short-term and fragmented organisation of research supported the idea that RRI in practice must be carefully planned, or that we must have some kind of control over how it will unfold and where we will end up (hence the appeal of fragmented, and thus more controllable, projects into work packages). However, precisely because they are about complex and uncertain grand societal challenges, the outcomes and impacts of RRI projects are unpredictable. In the 2022 PhD courses, a group of participants illustrated this challenge by creating an 'RRI boardgame'[11] where the player – an early career researcher – rolls a die and moves on the board where she/he randomly faces events such as 'lack of funding', 'loss of motivation', 'new collaborations', or 'user (dis)satisfaction' (see Figure 1). The objective was to paint a more realistic (and less linear) image of the lived reality of RRI post-graduates by showing how they must adapt to and address altogether desired, undesired, intentional, and unintentional outcomes in their everyday research practice.

---

11 The template of the boardgame and additional explanations can be found here: https://www.ntnu.edu/web/afino/a-boardgame-to-play-out-the-gap-between-intentions-and-outcomes

Figure 1: The RRI boardgame, by participants in the 2022 AFINO/DLN PhD course.

Second, some participants pointed out the tension between the caring and responsible nature of their projects, aiming for long-term, deep transformative social changes while being funded by institutions governed by new public management and therefore demanding quick, measurable results. These demands have a direct impact on *what counts* – what counts as important research, what counts as relevant outcomes. Indeed, investing time and resources to address complex global societal challenges might not be highly prioritised in a context run by classical academic accounting (Felt et al., 2015). Many participants had very socially engaged projects, looking, for instance, at the built-in biases in AI algorithms, addressing ecosystems collapse and existential risks through narrative research, looking at social justice through crowdfunding approaches to cultural events, or approaching resilience in Arctic communities through decolonial and anti-domination perspectives. The outcomes of such projects cannot be grasped by any fixed measures, and the impacts might only be seen several years after these projects are finished. During the first PhD course in 2021, a group of participants explored these tensions by creating a poetic animation[12] (see Figure 2) illustrating how, even though RRI (depicted as a small bird in the animation) is demanded and supported rather

---

12 The full animation can be seen here: https://www.ntnu.edu/web/afino/rri-implementation

strongly by research policy institutions such as the RCN (depicted in the animation by an elephant), post-graduates nevertheless have to navigate 'intimidating power structures' and 'hierarchies', 'dominant' and 'obstinate' interests, and 'constraining norms' that align with ideals of 'control', 'excellence', 'predictability', and 'speed'. They warned against the fact that in a new public management regime, RRI can quickly become a 'tick-boxing' exercise, which, in essence, has nothing to do any longer with the more committed, caring, and engaged projects of the participants.
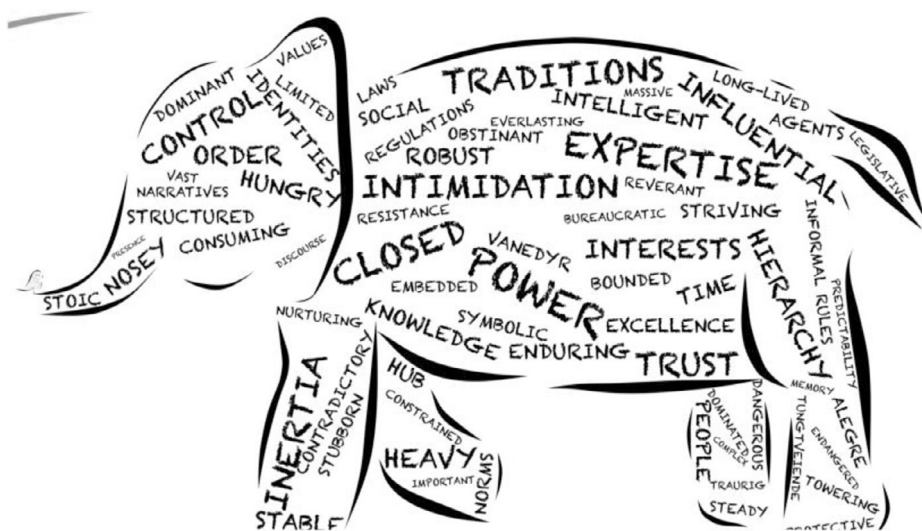
Figure 2: The challenge of RRI implementation, by participants in the 2021 AFINO PhD course.

Short-term temporalities explicitly and tacitly permeate and govern RRI in practice. Early career researchers, whose professional situations are the most precarious and vulnerable in the research and higher education landscape (OECD, 2021), are also the ones who are the most impacted by this short-term organisation of research. Trying to fit long-term RRI ambitions into short-term projects can first denature the responsible and caring qualities of their RRI project, and second, it increases the number of tasks that they must comply with during an already-short project (see also the discussion about 'moral saturation' in Chapter 4). The participants often said that if they wanted to pursue their commitment to RRI, and more specifically to research that is responsible, caring, and more just, in an environment that is often not supportive, they felt 'squeezed' and 'struggling' trying to manoeuvre to fit all their tasks into tight and busy schedules (from publishing at high speed to engaging in participatory research, for instance). The pressure for quick and measurable results nurtures a competitive culture around project-based funding (OECD, 2021), and as reported by some participants, they sometimes feel cornered into strategies of over-promising project

outcomes in terms of quantity or in terms of their 'measurable' quality. Some also clearly voiced some despair and disillusion regarding the fact that research might not be the best path to pursue the ideals of RRI, including social justice, fairness, care, or responsibility. It is therefore crucial to be attentive to where these short-term temporalities come from and how they govern RRI practices, as they generate growing misunderstandings and misalignments between research and policy environments, in between which RRI post-graduates are often trapped and left to disillusion.[13]

3.2 Tensions in navigating narrow merit-based criteria

Not only does the regime of new public management translate into short-term temporalities and logics of control, but it also translates into narrow norms of how researchers are evaluated and rewarded. These narrow merit-based criteria predominantly concern scientific production, teaching, institutional and discipline-based involvement, and impact (Musselin, 2022) and translate into quantitative indicators such as citation indices, journal impact factors, or h-indexes. These have become the current markers of scientific excellence, productivity, and quality (OECD, 2021). However, these metrics only focus on limited (measurable) aspects of research, making excellence a rather narrow concept. In addition, they pose various challenges, including the intensified competition and nurturing of a 'publish or perish' culture (especially among early career researchers whose academic career depend on the track record of publications), an emphasis on individual and disciplinary excellence, a focus on short-term outcomes and impacts, and a focus on measurable and auditable results (OECD, 2021).

These merit-based metrics therefore pervasively govern scientific practices. How we internalise those criteria, and also the short-term temporalities and logics of the research system, became very visible when some participants discussed the tensions they felt within themselves between being post-graduate researchers who wish to pursue an academic career (and therefore have to 'play the game' of meritocratic science) and their engagement in socially responsible, caring research (which bears purpose and values such as feminism, inclusiveness, anti-domination, or social justice).

In addition to creating inner tensions, the way scientific merit and excellence are evaluated and measured does not meet the broader expectations for more socially responsible science. They are simply not favourable to long-term, risk-taking, and less predictable types of research, such as RRI. Therefore, as the report from the OECD

---

13 Even if this is beyond the scope of the experiences reported here, it is worth mentioning that there is a growing body of literature pointing at the pressures on early career researchers which pose a threat to their mental health and wellbeing, to the extent that it might also undermine the integrity of research (see for instance: Cilli et al., 2023; Kismihók et al., 2022; or Learning, 2020)

(2021) stated: "Science is indeed a meritocracy but there is an urgent need to redefine those merits and what constitutes excellence in all its different guises" (p.90).

The question of how to evaluate the merits of RRI projects was a central one in the AFINO PhD courses. Participants felt the criteria for scientific merit mentioned above only reflected a very restricted part of their work and didn't encompass the depth of their RRI research, notably their aim towards deeply transformative social change. In particular, some participants with a bioengineering background felt pushed into focusing on marketable innovations rather than on aiming to deeply transform research and innovation processes towards more social relevance. Many participants also voiced the fact that they felt locked into heavy, entrenched, change-adverse universities, where dominant reward systems and a narrow understanding of excellence did not support responsibility, care, and justice in research and innovation processes. This observation is supported by Mejlgaard et al. (2019), who state that "measures of merit, performance, and success, which are implemented throughout the university sector, nationally and locally, tend to favour traditional components of academic work such as publishing in high impact journals and patenting the results of research and innovation activities. This is not necessarily compatible with the ideas of RRI." (p.609–610)

In the first PhD course in 2021, a group of participants worked specifically on that question.[14] They argued that metrics for evaluating 'responsibility' are context dependent and should therefore be carried out through appropriate qualitative methods, for instance, through a 'dialogic and iterative process' between the different stakeholders of the R&I process, reassessing the framing of the problem at hand, the ways of knowing it, and the potential solutions that could be designed (see Figure 3). Thus, rather than 'rigid guidelines', the participants highlighted the importance of thinking about evaluation as an "open and inclusive discussion".

The question of how to evaluate the merits of RRI was furthered by some participants who not only pointed out the importance of looking at the outcomes, products, and impacts of RRI projects, which are not only impossible to predict but also might only be visible long after the project is finished. Rather, they also suggested paying close attention to the processes of RRI, or what happens during an RRI project: the efforts put into learning, reflecting, opening up, including, and anticipating. This is also mentioned in the RRI framework for the RCN: "the emphasis of impact evaluation is shifting from (end) product to process, and from verdicts/judgements to learning and improving" (p.4). This supports the participants' point above (in Figure 3) that evaluation should take the form of an inclusive discussion, among the RRI actors, to come up with relevant indicators and measures of merit that are best fitted to the context.

---

14 The full presentation on the evaluation of RRI can be found here: https://www.ntnu.edu/web/afino/rri-evaluation
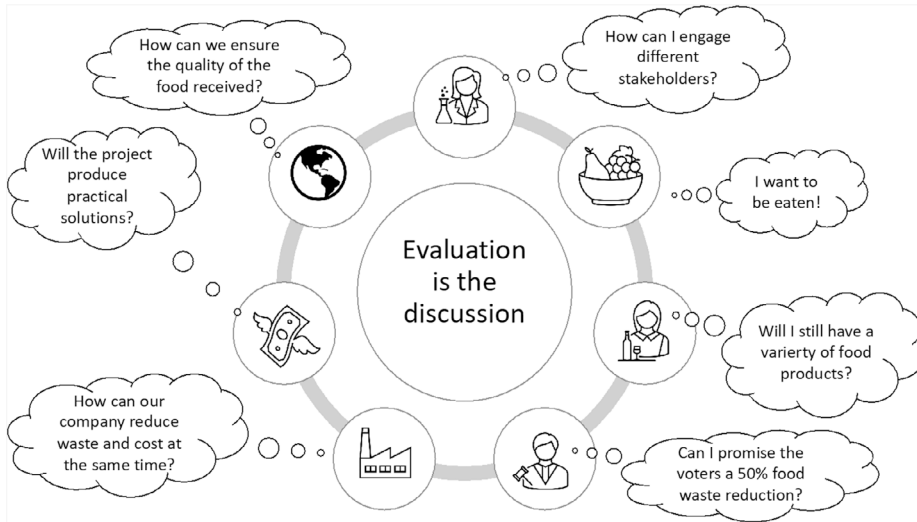
Figure 3: The challenge of RRI evaluation, by participants of the BREAD project, in the 2021 AFINO PhD course.

As we will now see in Section 3.3, the different temporalities, demands for quick results, and narrow merit-based criteria can translate into a gap between intentions and outcomes in RRI projects. These temporalities and logics have taken a grip on research policy ambitions and support the fiction that project outcomes are predictable and expected to be visible and measurable in the short term (Felt, 2021). The demands from research policy institutions, for transformative learning processes and predictable, measurable outcomes, are conflicting, and therefore, despite numerous indicators and roadmaps for implementing and evaluating RRI in practice, fundamental obstacles remain to transform research into outcomes that help address grand societal challenges.

3.3 Tensions stemming from the gap between intentions and outcomes in RRI projects

Most of the participants attending the AFINO PhD courses were driven by a commitment to deep societal changes, towards more responsibility, care, fairness, and justice in particular, and many were engaged in participatory research – through narrative approaches or citizen science, for instance. In this respect, RRI was often seen as a structure or compass to facilitate the journey towards such intentions within their project, something that could help achieve a needed shared understanding of what good processes and outcomes consist of, across the various actors of the R&I processes (Wickson and Carew, 2014), as well as help them structure self-reflections (Stahl et al.,

2023). However, as we have seen above, research is evaluated according to a logic where there is a clear separation between outputs, outcomes, and impacts, which are themselves seen in a linear way (Arnold et al., 2019).

In the PhD course of 2022, on engaging in critical research within institutions, we directly addressed the gap that the participants had experienced between their intentions for pursuing critical, responsible, and caring research and the lived realities and outcomes of their research. The participants were placed in disciplinarily heterogeneous groups, where they produced creative materials to illustrate the gap they felt between their initial, critical intentions for their research and the actual outcomes that were sometimes far from what they had envisioned. The participants all suggested formats that were outside the traditional realm of research:[15] a board game (see Figure 1) to show how the path between intentions and outcomes in research is not linear but subject to unpredictable events, 'street-interviews' of pedestrians to get their spontaneous opinions and values on synthetic meat, thus illustrating the conflicting roles of RRI PhD candidates and post-docs as both researchers and activists, and a live chatroom. The group that produced the live chatroom staged five early career researchers talking about their daily life struggles and doubts in a humoristic way. They used memes to address in a 'light' way some profound challenges (see Figure 4), such as the unavailability of their PhD supervisors or their lack of understanding for deeper motivations for criti-

Figure 4: Excerpt from the gap between intentions and outcomes illustrated by a live chatroom, by participants from the 2022 AFINO/DLN PhD course.

---

15 The participants' full contributions to how to address the gap between intentions and outcomes in RRI research can all be found here: https://www.ntnu.edu/web/afino/the-rri-garden

cal and responsible research, their self-doubts about being 'good enough', and their worries about finishing on time and publishing enough, when they are split between engaging in long-term participatory research and needing publishable results from it, for instance. They concluded by stating that "creating a safe space" for sharing these challenges, "being kind to oneself", and "nurturing a collaborative environment" in research were some of the keys to navigating the gap between intentions and outcomes.

What was common across the contributions from the three groups was the perception of the main cause or source of the gap between intentions and outcomes: it was seen to mainly stem from research environments and funding logics. The idea that from the intentions (or project descriptions), outcomes (or the reality of where the project 'ends up') logically unfold in a linear and predictable way is strong in research and policy environments (Arnold et al., 2019). This logic is visible in the practical organisation of research, which tries to fit real life into strict timelines and Gantt charts, manageable work packages, planned milestones, deliverables, and foreseeable outcomes (Gulbrandsen and Sivertsen, 2018). However, as seen in Section 1, the same policy and research environments also strongly invite to look at global societal challenges for what they are – wicked (Rittel and Webber, 1973) complex and uncertain – and ask to address them in holistic and transdisciplinary ways, with attention to responsible research and innovation processes. The demands are therefore self-contradicting, between encouraging RRI, risk-taking, and transdisciplinary research and yet remaining in a culture of tight micro-management and risk aversion in research funding and auditing.

The exercise of trying to fit into this landscape of conflicting demands implies that post-graduates might engage in risk-taking, critical RRI research, thinking that they will get adequate support for it, when the reality is often that throughout their project, they have to internalise specific (and narrow) notions of what counts as 'success', and what could be accomplished in the duration of a PhD or post-doctoral project. In particular, the participants expressed on many occasions their struggles in complying with indicators and norms for scientific merit and quality (number of publications, h-indexes, journal impact factor), which was seen as incongruent and not representative of their attempts to address grand societal challenges in a responsible, caring, and critical way. This created a sense of failure and self-doubts and a deeper split between the dual role they felt as researchers aiming for high-quality knowledge and 'activists' committed to responsible, participatory, risk-taking research. Further, the participants expressed a lack of consideration for and discussion around unintentional outcomes – when these are met, there is no support system for how to address them, or empathy in considering those outcomes as normal parts of the research process, and this results in themselves, rather than undesired and useless outcomes. This lends to a form of individualisation of responsibilities for impacts, where post-graduates might feel solely responsible for the organisation of their research and the promised outcomes.

The teaching team did not leave the issue here but discussed together with the participants how to address those gaps. As mentioned by the 'live chatroom group', relying on a *caring RRI community* was seen as a key element of navigating the gap

between intentions and outcomes. This caring RRI community would act as a platform for sharing doubts and reflections on the conflicting demands stemming from research and policy environments and the sense of failure they lead to, but also methods for RRI that can benefit navigating between disciplines to gain momentum and robustness. When research is viewed as a collective epistemic and normative enterprise rather than purely an individual effort, the borders between success and failure start to erode (see also Chapter 4). Indeed, if we consider testing hypotheses, the falsification of a hypothesis is an achievement on the collective level; however, on the individual level, one is left with a sense of failure since one has not achieved the promised, measurable outcomes within the project timeframe. The rhetoric of 'high risk, high gain' that is explicitly encouraged (in particular by the European Commission in their ERC programmes) lends this feeling of failure: it might ring true on an aggregate level, but for the individual who addresses global societal challenges characterised by high complexity and uncertainty, a feeling (and sometimes a stamp) of failure is very close.

4 Conclusion: anti-fragile, hybrid partnerships

In this chapter, we have looked at different tensions that post-graduates encounter when they work with RRI. These tensions range from navigating short-term temporalities, expectations for quick results, narrow merit-based criteria, conflicting demands from research environments, and a lack of support in engaging in responsible, critical, risk-taking research. These often resulted in a gap between intentions and outcomes in their RRI projects and an associated sense of failure. What we have tried to highlight, throughout Section 3, is the root of these tensions. They are based in a mismatch between the slow and non-linear nature of RRI processes, the results of which are difficult to predict and measure, and the high expectations stemming from research policy institutions (such as the RCN) for foreseeable, fast, and measurable outcomes. This mismatch is fuelled by the fact that RRI practice demands to embody certain values, such as attention and care to public values, and it thus puts a significant responsibility on individual researchers to reflect these values in their daily work (Mejlgaard et al., 2019). Extending or sharing this responsibility with others is not common practice. However, we have seen through exploring the tensions expressed by post-graduates that there is a need for a broader *caring community around RRI*. This is supported by Broerse (2016), who argued that for RRI to thrive, there needs to be support and enthusiasm both from individual researchers and from higher levels of research organisations, such as research policy institutions (Mejlgaard et al., 2019).

Following this observation, we want to conclude this chapter with our initial reflections on the need for new, hybrid partnerships around RRI, specifically between re-

searchers, innovators, and research policy institutions, to alleviate some of the challenges and tensions met by RRI post-graduates. The tensions between the practice of RRI at the individual research level and the call from research policy institutions for transformative research and innovation need to be highlighted and addressed (Arnold et al., 2019; Gulbrandsen, 2019). In other words, that means taking third-generation R&I policies seriously (see footnote 2) by involving not only RRI researchers and practitioners but also more centrally research policy and funding institutions, along the various steps of R&I processes, in particular in:

1. framing R&I scopes, intentions, realistic objectives and reflecting on the visions that steer us towards new socio-technical futures (here, it is particularly important to challenge the narrative according to which, for advanced economies, excellence in innovation is an absolutely essential element for supporting economic competitiveness and growth. As long as this narrative is central, nothing that slows down the pace of R&I will find the soil to grow);
2. addressing RRI implementation challenges (such as establishing a common understanding of the issue at hand, and language to address it);
3. taking part in quality control and evaluation mechanisms of RRI projects;
4. reflecting on the potential outcomes and broad socio-environmental impacts of these new futures created through R&I projects, as well as how these outcomes might impact the future allocation of research funding; and, last but not least
5. thinking about the way we teach RRI in higher education, not only as an approach or practice that embodies a certain set of values but also as an opportunity to create and nurture shared spaces and extended, hybrid communities for thinking about RRI, its processes, aims, and conditions of implementation.

This is certainly a demanding step for research policy institutions, as it makes them a 'responsible societal actor' (RCN, 2019), centrally involved in R&I processes. Elisabeth Gulbrandsen, former special advisor for RRI at the RCN, illustrates this as traversing the spaces between the "no longer" and the "not yet" (2019). Specifically, she writes:

> *The challenge of addressing Grand Challenges entails situating research processes as "triple loop" learning processes. This involves figuring both "politics" and "research quality" in innovative ways that can help responsible technoscientific cultures emerge, in and beyond the academy. [. . .] What does it take to collectively develop prospectives and figurations to provide directionality for new "trying transformations"? How can we further mutual learning and the development of innovative approaches by researchers, research councils and innovation agencies? (p.26)*

Indeed, the path to new, hybrid partnerships around RRI, and in teaching RRI, is still very explorative. One important feature of those partnerships might be found in Stilgoe's reflections on shared space and slow science (2019), where he argues for 'slow science' as a way to focus on the *quality* of R&I processes, rather than their speed: "we might challenge a one-dimensional fixation on efficiency in science and democratise the discussion of its other qualities" (p.266). Creating space and time for such hybrid

partnerships in the teaching of RRI, by involving innovators and research policy institutions, might indeed extend responsibility, beyond the individual researchers, and encourage a mutual learning based on the different experiences and competences of the different R&I actors. This is also discussed by Borch and Throne-Holst (2021), who argue that if RRI is indeed a new way of doing science, more reflections, dialogue, time, and resources are needed at the political level to facilitate that transformation. The space and time for hybrid partnerships arguably allow a critical rethinking of the explicit and tacit temporalities, logics, underlying values, and expectations that surround RRI (one's own, the project members, supervisors, [inter]disciplinary communities, research policy institutions, society, innovators and private sectors). In that sense, we argue that extending responsibility to the different R&I actors would contribute to creating a culture of anti-fragility where R&I actors can develop skills to address elusive and varied understandings of RRI; discuss asymmetry of power, different roles, expectations and logics; and navigate uncertainties and doubts inherent to anticipating future impacts or opting for adequate implementation options for RRI.

The tensions explored in this chapter are useful for giving a contour to these anti-fragile, hybrid partnerships around teaching RRI. Several themes, stemming from these tensions, could indeed help structure partnerships between researchers, innovators, and research policy institutions by asking:

1. How to foster a common understanding of RRI, its theoretical and practical contours, and its aims and temporalities, when it is specifically aiming at addressing grand societal challenges?

2. How to concretely implement RRI for grand societal challenges, in a context of changing policy landscapes that do not prioritise RRI in the same way; uncertain outcomes and impacts of R&I projects; tensions between various social, research, and policy temporalities and logics; conflicting expectations and demands from the different R&I actors; and tacit assumptions inherent to different socio-technical visions and futures?

3. How to navigate early career researchers' challenges and tensions, in particular by encouraging an anti-fragile culture of trial and error that allows real risk-taking initiatives, that is open to discussing failures, and that learns to report them, knowing that there is strength in showing weaknesses and vulnerabilities? Indeed, such hybrid partnerships also depend on recognising and thriving on failures (Gulbrandsen, 2019).

In aiming towards hybrid, anti-fragile partnerships around RRI, the value of complexity needs to be remembered, such that the outcomes of such partnerships are not represented through one consensual indicator or metric (Viseu, 2015). Inclusive, entangled, and co-existing voices must be welcomed so that the research and innovation processes are response-able (Haraway, 2016), that is, "cultivating collective knowing and doing" (p.34).

Giving careful attention to the tensions expressed by post-graduates around RRI means that teaching RRI deploys a potential for creating shared spaces and anti-fragile, hybrid partnerships. Teaching RRI moves from being a matter of finding the optimal ways of conveying ideas, values, and practices of RRI to extending the responsibility to other R&I actors, for a profound rethinking of the value of failures, the different temporalities, aims, logics, and how to navigate them in a response-able and anti-fragile way.

References

Arnold, E., Aström, T., Andréasson, H., Nielsen, K., Wain, M., Tofteng, M., & Røtnes, R. (2019). Raising the Ambition Level in Norwegian Innovation Policy. *Final Report*: *technopolis group*.

Borch, A., & Throne-Holst, H. (2021). Does proof of concept trump all? RRI dilemmas in research practices. *Science and engineering ethics*, *27*(1), 7.

Broerse, J. (2016). Integrating RRI in universities: Training future students in transdisciplinary research. In *Keynote session at the 1st HEIRRI conference "Teaching Responsible Research and Innovation at University"*, Barcelona, Spain.

Cilli, E., Ranieri, J., Guerra, F., & Di Giacomo, D. (2023). Early career researchers and mental health: Observational study of challenge and wellbeing. *Health Science Reports*, *6*(11), e1649.

Delaney, N., Iagher, R. & Tornasi, Z. (2020). *Institutional changes towards responsible research and innovation – Achievements in Horizon 2020 and recommendations on the way forward*, European Commission, Directorate-General for Research and Innovation Publications Office, https://data.eu ropa.eu/doi/10.2777/682661

Felt, U., Igelsböck, J., Schikowitz, A., & Völker, T. (2015). Transdisciplinary Sustainability Research in Practice: Between imaginaries of collective experimentation and entrenched academic value orders. Pre-print; Published by the Department of Science and Technology Studies, University of Vienna, May 2015. Available at http://sts.univie.ac.at/publications

Felt, U. (2021). In Conclusion: The Temporal Fabric of Academic Lives: Of Weaving, Repairing, and Resisting. In *Inquiring into Academic Timescapes* (pp. 267–280). Emerald Publishing Limited.

Fisher, E., & Rip, A. (2013). Responsible innovation: multi-level dynamics and soft intervention practices. *Responsible innovation: Managing the responsible emergence of science and innovation in society*, 165–183.

Gulbrandsen, E. (2019); PhD titled "The New Production of Politics – Between the No Longer and the Not Yet"; Blekinge Institute of Technology, Department of Technology and Aesthetics; Doctoral Dissertation Series No. 2019:13

Gulbrandsen, M., & Sivertsen, G. (2018). Impact i anvendt forskning: Begrepsavklaring og praksis. I *52 [Working paper]. Nordisk institutt for studier av innovasjon, forskning og utdanning NIFU.* https://nifu.brage.unit.no/nifu-xmlui/handle/11250/2570526

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Haraway, D. J. (2016). *Staying with the trouble: Making kin in the Chthulucene*. Durham: Duke University Press.

Kismihók, G., McCashin, D., Mol, S. T., & Cahill, B. (2022). The well-being and mental health of doctoral candidates. *European Journal of Education*, *57*(3), 410–423.

Larrue, P., & Santos, R. (2022). "Towards a new stage in Norway's science, technology and innovation system: Improving the long-term plan for research and higher education", *OECD Science, Technology and Industry Policy Papers*, No. 133, OECD Publishing, Paris,https://doi.org/10.1787/ce07b7c3-en.

Learning, S. (2020). What researchers think about the culture they work in. *Wellcome Trust*, *London*.

Macnaghten, P., Owen, R., Stilgoe, J., Wynne, B., Azevedo, A., de Campos, A., . . . & Velho, L. (2014). Responsible innovation across borders: tensions, paradoxes and possibilities. *Journal of Responsible Innovation*, *1*(2), 191–199.

Mejlgaard, N. Buljan, I., Saladie E.N., et al. (2016). HEIRRI – Higher Education Institutions & Responsible Research and Innovation. Stock Taking/Inventorying (WP2) D2. 2 State of the Art Review. *Horizon 2020 EU Project Report*.

Mejlgaard, N., Christensen, M. V., Strand, R., Buljan, I., Carrió, M., Cayetano i Giralt, M., . . . & Wuketich, M. (2019). Teaching responsible research and innovation: A phronetic perspective. *Science and engineering ethics*, *25*, 597–615.

Nulli, M., & Stahl, B. (2018). RRI in Higher Education. *The ORBIT Journal*, *1(4)*, *1–8.* https://doi.org/10.29297/orbit.v1i4.78

Musselin, C. (2022). Evaluation and Merit-Based Increase in Academia: A Case Study in the First Person. *Valuation Studies*, *8*(2), 73–88.

OECD STI Outlook (2021), *OECD Science, Technology and Innovation Outlook 2021: Times of Crisis and Opportunity*, OECD Publishing, Paris, https://doi.org/10.1787/75f79015-en.

Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013). A framework for responsible innovation. *Responsible innovation: managing the responsible emergence of science and innovation in society*, 27–50.

RCN (2015). Research for innovation and sustainability. Strategy for the Research Council of Norway 2015–2020.

RCN (2019). A Framework for Responsible Innovation – under BIOTEK2021, IKTPLUSS, NANO2021 and SAMANSVAR

RCN (2020). Empowering ideas for a better world. Strategy for the Research Council of Norway 2020–2024.

Ribeiro, B. E., Smith, R. D., & Millar, K. (2016). A mobilising concept? Unpacking academic representations of responsible research and innovation. *Science and engineering ethics*, *23*, 81–103.

Rip, A., & Voß, J.-P. (2013). Umbrella terms as a conduit in the governance of emerging science and technology. Science, Technology & Innovation Studies, 9(2), 39–59.

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences*, *4*(2), 155–169.

Stahl, B. C., Aicardi, C., Brooks, L., Craigon, P. J., Cunden, M., Burton, S. D., . . . & Helena, W. E. B. B. (2023). Assessing Responsible Innovation Training. *Journal of Responsible Technology*, 100063.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a Framework for Responsible Innovation. *Research Policy* 42 (9): 1568–80. https://doi.org/10.1016/j.respol.2013.05.008.

Stilgoe, J. (2019). Shared space and slow science in geoengineering research. *International Handbook on Responsible Innovation: A Global Resource*, 259(17).

Tassone, V. C., O'Mahony, C., McKenna, E., Eppink, H. J., & Wals, A. E. (2018). (Re-)designing higher education curricula in times of systemic dysfunction: A responsible research and innovation perspective. *Higher Education*, *76*, 337–352.

Viseu, A. (2015). Caring for nanotechnology? Being an integrated social scientist. *Social Studies of Science*, *45*(5), 642–664.

Völker, T., Slaattelid, R., & Strand, R. (2024). Translations of Responsibility: Innovation Governance in Three European Regions. Routledge, NY and London.

Von Schomberg, R. (2013). "A Vision of Responsible Research and Innovation." In *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, edited by Richard Owen, John Bessant and Maggy Heintz, 51–74. John Wiley & Sons. https://doi.org/10.1002/9781118551424.ch3.

Wickson, F., & Carew, A. L. (2014). Quality criteria and indicators for responsible research and innovation: Learning from transdisciplinarity. *Journal of Responsible Innovation*, *1*(3), 254–273.

Wickson, F., Strand, R., & Kjølberg, K. L. (2015). The walkshop approach to science and technology ethics. *Science and engineering ethics*, *21*, 241–264.

Åm, H. (2019). Ethics as ritual: Smoothing over moments of dislocation in biomedicine. Sociology of Health and Illness 41 (3): 455–469.

Atle Midttun

Chapter 6
Companies squeezed between autocratic and democratic regimes

For decades, multinational companies have stepped up their efforts to embrace corporate responsibility. They have done so under the Western-led global agenda based on market liberalism and liberal-democratic values. The vision has been that globalization of markets will stimulate globalization of liberal values and that Western-style corporate responsibility will follow, energized by civic engagement and public debate.

However, the rise of China as a major economic powerhouse, in alliance with an increasingly aggressive and dictatorial Russia, has marked an authoritarian counterpoint, not only to the Western dominance of global commerce but also to Western liberal democracy and its civic-driven corporate responsibility. The global political economy has thereby become increasingly marked by a new, bipolar rivalry between democratic and autocratic states.

As the world moves towards a bipolar contestation between democratic and autocratic regimes, aggravated by the Russian war in Ukraine, this chapter argues that it is time to adapt corporate responsibility (CR) to new bipolar realities. It contends that the shift from a neoliberal, Western-dominated model to a polarized globalization requires a transformation in corporate responsibility – from a heroic Western multinational championed model to a pragmatic, negotiated, and government-partnered approach.

Our research investigates CR across this divide through studies of affected companies, including analyses of new strategies to counter aggressive Russian energy policies and the corresponding Western financial sanctions.

The analysis combines a conceptual approach with explorative case studies, supplemented with a review of relevant literature.

1 Introduction: toward a clash of ideologies

Following on from the era of neoliberal globalization that emerged at the end of the 20th century, the global economy has moved towards a new bipolar confrontation between democratic and autocratic states. The new bipolar rivalry differs from the previous Cold War between the Western and the Communist blocs, where political controversy was paralleled by economic segmentation. In the 21st century the authoritarian regimes have embraced global capitalism and joined the 'Western' market arena. We are, in other words, faced with *a clash of ideologies* and *an agon* between

authoritarian and liberal spheres, but this time in the context of common market engagement.

The conflict is now not just over politics and governance but also cultural values. Western authorities and civil societies embrace democratic government, human rights, and liberal freedoms. Authoritarian and theocratic powers, such as China and Russia, not only demand patriotic nationalism and obedience to the autocratic ruler but also repudiate liberal norms they view as Western centric. This echoes Samuel Huntington's 1993 'Clash of Civilizations' thesis, which suggests that cultural and religious identities will be the primary sources of conflict in the post–Cold War world.

In this bipolar world,[1] multinational business risks being caught in the middle with complementary commercial interests but conflicting expectations with respect to values, governance, and accountability. The clash of ideologies is especially salient with regards to the agenda of corporate responsibility/sustainability, where values are most pronounced. Typical topics such as human rights and workers' social conditions, as well as the right of civic organizations to engage in open stakeholder dialogue, easily put global business in a crossfire of contradictory points of view. Corporate environmental sustainability has also stoked controversy across the democratic-autocratic divide.

How can multinational business actors tackle this situation? Will they have to reduce corporate responsibility/sustainability to a limited common minimum? Will the ideological resentment across the authoritarian-democratic divide lead to stronger market segmentation? Or, to consider a third option, are there opportunities for strategic reconfiguration that allow companies to operate more flexibly on both sides of the divide?

This chapter discusses these and related questions, based on three short explorative case studies of companies that have been exposed to controversies over their sustainability agenda. They range from disputes over human rights in Chinese supply chains to transparency and security issues in Western telecoms infrastructure and on to engagement in Russian markets during Putin's aggression in Ukraine. On the basis of the case analyses, our study outlines strategic options for corporate responsibility going forward, including alternative business models and market segmentation.

The chapter starts out with a brief review of core aspects of the Western/liberal tradition of corporate responsibility, as a conceptual basis for the discussion.

---

1 The bipolar divide between authoritarian and democratic states is, of course a simplification, and covers a gradual scale between the two as indicated for instance in the democracy index, with many states attempting to occupy some space in the middle (EIU, nd). Nevertheless, the cases presented indicate that this divide is central to modern globalization and defines significant conflicts between major economic, political and military power-holders.

1.1 Liberal globalization and corporate responsibility

The basis for the present clash of ideologies stems from the neoliberal shift and the extensive deregulation of Western economies starting in the 1980s. This period marked the beginning of a pronounced transformation, culminating in the fall of the Soviet Union and the transition of Eastern and Central European nations to market economies in the 1990s. The post–Cold War period heralded what was seen as a golden era of liberalism (Fukuyama, 1992). The expansion of economic liberalization into the fast globalizing world facilitated the dramatic scaling up of corporate enterprises, extensive transnational networks, and the widespread outsourcing of production to emerging and developing economies.

Traditional regulatory theory advocates oversight to ensure markets are not distorted, and in a liberal democracy, such oversight should be legally established by the democratically elected legislature. However, as markets underwent rapid expansion under neoliberal globalization, regulatory mechanisms lagged behind. The result was an under-socialized global economy with a governance void at the global level where political consensus on regulation was hard to achieve (Ruggie, 2014, Midttun, 2022).

The doctrine of Corporate Social Responsibility (now referred to as CR)[2] emerged as an attempt to fill this void, postulating *the business case* for prosocial and environmentally benign corporate behaviour with such labels as 'creating shared value' (Porter and Kramer, 2011) or 'Purpose driven organization' (Quinn & Thakor, 2018). This approach might seem overly optimistic, especially as businesses increasingly manage global supply chains that incorporate production in developing nations with lax economic, social, and environmental standards. However, the argument for CR gains plausibility when civic engagement is factored in, with active community groups pressuring corporations to adopt social and environmental concerns through stakeholder engagement (Freeman, 1984) and leveraging the potential repercussions of media scrutiny (Fombrun, 1995).

Consequently, since the early 2000s, multinational corporations have faced what Australian political scientist John Keane (2013) describes as a "monitory democracy" – a form of direct-democratic activism that operates alongside traditional parliamentary democracy. Fuelled by social media and multiple information channels, this form of activism directly challenges businesses while also exerting influence on parliamentary politics to further corporate responsibility objectives. Many civil society organizations (CSOs)/non-governmental organizations (NGOs),[3] which themselves operate on a

---

2 There are many terms used to denote such behavior: *corporate social responsibility* - CSR, *corporate responsibility* – CR, *corporate sustainability*, and environmental, social and governance oriented investment – ESG, just to mention a few. We shall use CR as a generic term throughout this chapter.
3 We shall use both terms: 'Civil Society Organizations (CSOs) and 'Non Governmental Organization' intermittently to refer to groups or associations that operate in the public interest, independently from the government. The primary goals of such organizations are to advocate for social causes, con-

multinational scale and are unconstrained by geographical limits, have played a crucial role in extending governance to the global sphere.

These developments have yielded a flourishing array of novel governance initiatives intertwining with parliamentary democracy. They also led to court rulings in a mix of soft and hard power across national, regional, and global realms (Figure 1). Witness the United Nations' Sustainable Development Goals, which have little implementation power behind them but possess a normative power articulated through civic mobilization, communication, and monitory democracy. Although the limited mandate of the UN cannot mobilize strong implementation power behind the Goals, civic mobilization, communication, and monitory democracy may provide substantive help. This is especially the case if civic mobilization affects consumer sentiments and companies incur reputational setbacks and diminished sales.[4]
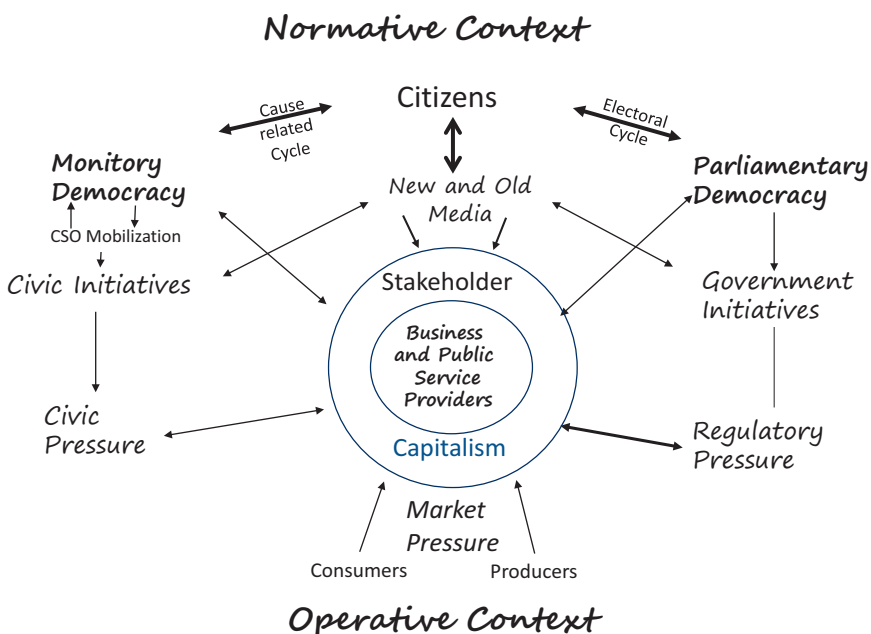
Figure 1: Novel Governance Horizons (Source: Midttun 2022).

tribute to the development of public policies, promote civic engagement. While the term CSO stresses the character of the organization, the term NGO stresses their formal independence of governments.
4 Corporate responsibility has been analyzed in a massive literature, including Freeman et. AL. (2010), Porter (2011); Kotler et. AL. (2010); Fombrun (1995); Schaltegger et al (2016). Midttun (2022) chapter7 provides an overview of some of it. This paper only draws selectively on it to highlight our discussion of the corporate challenges arising from exposure to the authoritarian-democratic divide.

Western-style CR has, in other words, been anchored in a liberal political context, with free media, democratic elections, and the right to civic mobilization – characteristics of the major Western economies that dominated the global markets throughout the late 1990s and early 2000s.

In this context, business has had to engage both with traditional parliamentary-regulatory and monitory democratic processes, typically adopting an outreach towards stakeholders in addition to the engagement with shareholders, customers, and regulatory authorities (Freeman et. al., 2010, Midttun, 2022).

Exposed to pressure from domestic stakeholders, multinational corporations were subsequently expected to practice and propagate responsible social and environmental behaviour throughout their supply chains across the world, turning them into carriers of not only neoliberal commerce but also *heroic* promoters of social and environmental upgrading through CR across the world.

Recent initiatives to strengthen pressure on human rights initiatives in multinational supply chains follow in this tradition (EU Council, 2023; see Chapter 10; see also Chapter 9 for a discussion of the role of whistle-blowing in institutionalizing responsibility).

1.2 Corporate responsibility under emerging bipolar rivalry

Navigating between the complexities of Corporate Responsibility (CR) and monitory democracy in the late 1990s and early 2000s, the corporate realm, alongside the Western model of neoliberal globalization, faced a new wave of challenges as the 21st century progressed. The period witnessed remarkable economic growth outside the traditional 'Western' sphere, propelling nations and regions from diverse corners of the world to the forefront of global capitalism. These entities have increasingly opted to diverge from the strictures of the Western neoliberal framework, striving to assert their distinct values and interests on the global stage.

The rise of Asian economic powers has been a significant game changer. Initially spurred by Japan's growth, this wave was further propelled by the 'Asian Tigers' – Hong Kong, Singapore, South Korea, and Taiwan – and subsequently by the BRICS countries: Brazil, Russia, India, China, and South Africa. Among these, China's rapid economic growth in the early 21st century has become part of the landscape of contested globalization.

Khuong Vu's 2020 analysis highlights this shift, examining the emergent division of economic powers between the E7 – seven large, non-Western emerging economies – and the traditionally dominant, primarily Western, G7. In 2000, the G7 commanded 43.8% of the global economy, with the E7's share of 22.3%. By 2017, the scenario had significantly transformed, with the E7's share surging to 36.5%, surpassing the G7's ratio of 31.0%. This trajectory persisted, leading to the G7's share in global GDP falling to 26.6% by 2023, marking a profound realignment in global economic power.

1.2.1 Values

Although the E7 and G7 divide does not neatly replicate the democratic-authoritarian divide, the dominant role of authoritarian-leaning China within the E7, juxtaposed with the democratic leadership of the United States[5] in the G7, indicates an ideological divide. While the majority of G7 nations are classified as 'full democracies' according to the Economist Democracy Index, the E7 is characterized by a predominance of authoritarian and hybrid regimes, with several 'flawed democracies' and one full democracy. The global economy's shifting epicentre is therefore accompanied by a transformation in values and political culture.

Authoritarian powers, led by China and Russia (the latter often included in an expanded E7 or E8 grouping), advocate for patriotic nationalism and allegiance to autocratic leaders while challenging liberal norms often perceived as Western centric. Hybrid regimes share similar objectives but tend to exert a more moderate level of coercion to enforce their value-charged principles. In this increasingly bipolar world, multinational corporations find themselves navigating a complex terrain of concurrent commercial interests and divergent expectations regarding values, governance, and accountability.

The divide between liberal and authoritarian regimes is particularly evident in their approaches to Corporate Social Responsibility (CSR) and human rights. During the 1990s, under the umbrella of neoliberal globalization, Western-driven CSR initiatives were introduced to China and other developing/emerging economies primarily through global supply chains, which were crucial for their export-led growth and rapid industrialization. However, as these nations developed more robust domestic economies, their bargaining power vis-a-vis Western counterparts increased, leading them to challenge Western-dominated international norms. Consequently, influential emerging economies like China began to establish their own standards, reflecting their unique political economies and cultural values. This shift means that China, for instance, maintains more restrictive stances on human rights and free speech, including assembly rights. Yet, it has allowed for relatively greater civic engagement and critique of environmental issues (Midttun, Wei, and Wu, 2014).

---

5 The United States has been categorized as a "flawed democracy" in recent evaluations, including the 2023 Democracy Index published by the Economist Intelligence Unit, (EIU, nd). This classification reflects concerns over several factors such as social cohesion, consensus on fundamental issues including election outcomes, and ongoing political polarization. These issues have been exacerbated by events like the Capitol riot on January 6, 2021, and contentious election processes. The presence of strong civil liberties and high voter turnout in the 2020 elections has nevertheless helped prevent an even steeper decline in the democracy rating.

1.2.2 Governance under bipolar rivalry

The competition for regulatory dominance between democratic and autocratic states introduces an additional layer of complexity to the already intricate governance of global business operations. In the new bipolar setting, achieving regulatory governance is challenging not only due to the limited sovereignty of governments in paralleling the scope of global corporate activities but also because of the political and value-based diversity among the involved governments. Similar conflicts of values are also prevalent among civil society actors and in debates over global norms, such as the Sustainable Development Goals (SDGs) advocated by the United Nations.

To reflect the increasing complexity, we have expanded the governance model shown in Figure 1 by adding an authoritarian 'layer' atop the liberal-democratic base, as illustrated in Figure 2: authoritarian/illiberal democracy competing with parliamentary democracy; monitory autocracy contesting monitory democracy; and authoritarian norms vying against democratic norms (Figure 2).
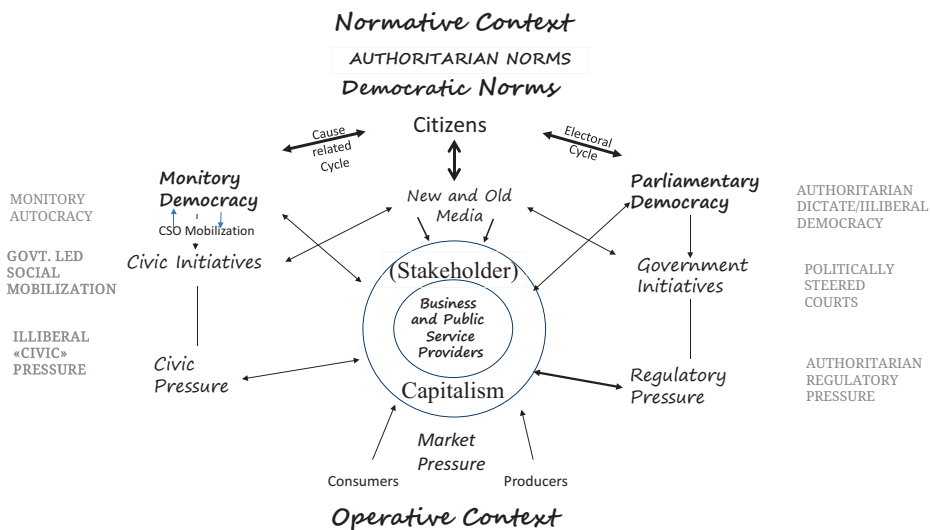
Figure 2: Novel Governance Horizons Under Bipolar Rivalry (Source: Midttun 2022).

Consequently, businesses operating on a global scale may find themselves exposed to multiple pressures from conflicting legislations and civic expectations which complicate their operations across the democratic-authoritarian divide.

Corporate Responsibility (CR), deeply rooted in ethical and societal values, emerges as a critical arena where the clash between democratic and authoritarian outlooks is most pronounced, reflecting the fundamental discord in governance and societal expectations.

As the economy is evolving from neoliberal towards contested globalization, so must CR. This paper argues that with the loss of neoliberal economic hegemony the West needs to transcend *heroic CR* and move towards a *pragmatic negotiated CR* that takes the evolving complexity and value confrontations into consideration.

1.3 Methodology: three cases

We have brought in three mini-cases to illustrate the dilemmas faced by multinational industry affected by the clash of democratic-authoritarian ideologies in an explorative research design (Yin, 2018). The first is the apparel company H&M's squeeze between Western human rights stakeholders and Chinese authorities over social conditions in the Uyghur region. The region furnishes the clothing industry with much of its cotton. The second is the tech company Huawei, squeezed between ambitions to serve global 5G markets and security concerns from democratic societies about authoritarian control of vital infrastructure. The third case highlights commercial hurdles arising from the Russian invasion of Ukraine and the economic sanctions that followed. The cases are based on available information through the press, research reports, and blogs. It draws on information and analysis from both democratic and authoritarian camps.

The first two cases have been chosen to exemplify critical challenges under bipolar confrontations for companies with respectively democratic and authoritarian home bases. The last case illustrates how challenges intensify when confrontations escalate into war.

2 The H&M case

H&M Hennes & Mauritz AB, or the H&M Group, is a multinational clothing company based in Sweden that focuses on fast-fashion clothing for men, women, teenagers, and children. As of 23 June 2022, the group operates in 75 geographical markets with 4,801 stores under various company brands (H&M a, nd). It is the second-largest international clothing retailer, behind Inditex (Tharawat Magazine, 2018)

As the world's largest producer of cotton, China has been an attractive supplier of both raw material and processed products to the Western apparel industry. This includes many of the world's best-known fashion retailers, and mainland China has traditionally been H&M's largest supplier/sub-contractor for cotton products with 366 factories, almost 3 times as many as its second-largest, Bangladesh (H&M b, nd). More recently, mainland China has also become one of H&M's major consumer markets, with 445 stores, only surpassed by the United States (548), and boasting the seventh largest regional revenue (more than 7 billion SEK). This surpasses major European markets such as Italy, Spain, and the Netherlands (H&M, 2022).

2.1 Human rights in Xinjiang

Around 85% of the Chinese cotton production comes from the Xinjiang province, with a dominant Turk Uyghur population. Chinese authorities have been accused of exercising strong repression against this population and of compelling the Uyghurs to participate in cotton harvesting and industrial production under forced labour conditions. A recent UN report corroborates this impression, as indicated in a strongly worded assessment at the end where the Office of the UN High Commissioner for Human Rights (UN OHCHR) says that the extent of arbitrary detentions against Uyghurs and others, in the context of "restrictions and deprivation more generally of fundamental rights, enjoyed individually and collectively, may constitute international crimes, in particular crimes against humanity" (UN OHCHR August 2022).

The Chinese government, on its side, has rejected the UN assessment. In the words of the Chinese Ministry of Foreign Affairs spokesperson Wang Webin: "This so-called assessment is orchestrated and produced by the US and some Western forces and is completely illegal, null and void. It is a patchwork of disinformation that serves as a political tool for the US and some Western forces to strategically use Xinjiang to contain China" (Chinese Embassy of New Zealand, 2022).

2.2 The pressure from Western CSOs

Following information from numerous refugees and their own analysis that pointed in the direction of the later UN report, a broad coalition of civil society organizations, trade unions, and investor groups mobilized to call for an end to forced labour in the Uyghur region.

In classical monitory democracy style, the Coalition launched a Call to Action (Freedom United, nd), seeking commitments from brands and retailers to stop sourcing cotton and other raw materials, yarn, textiles, and finished products from the Uyghur Region; cut ties with companies implicated in forced labour; and prohibit any supplier factories located outside of the Uyghur Region from using Uyghurs and other Turkic or Muslim-majority peoples supplied through the Chinese government's forced labour transfer scheme.

The call to action, in July 2020, directed at brand companies in the apparel industry, referred to their responsibility under *the UN Guiding Principles of Business and Human Rights*, according to which all businesses have a responsibility to respect human rights in their supply chains, regardless of where they operate. The Call to Action, voiced by more than 190 organizations, demanded that brands meet this responsibility by fully exiting the Uyghur region and ensuring they do not work with any suppliers complicit in the abuses of Uyghurs.

2.3 First company responses

The first reaction from the apparel industry was to express concern about the discrimination of the Uyghur minority but to assure everyone that they were not involved in any wrongdoing. To take the H&M company as an example, the company denied that it had any problems in its own supply chain (HM group c nd):

> *H&M Group is deeply concerned by reports from civil society organisations and media that include accusations of forced labour and discrimination of ethnoreligious minorities in Xinjiang Uyghur Autonomous Region (XUAR). We strictly prohibit any type of forced labour in our supply chain, regardless of the country or region . . .*
>
> *. . . We are committed to respecting human rights and our approach is guided by the UN Guiding Principle on Business and Human Rights and OECD Guidelines for responsible business conduct*

Nevertheless, H&M later admitted (16 September 2020) that it had cut ties with a Chinese supplier over accusations of "forced labour" (Business & Human rights, nd). Furthermore, in October 2020, five auditing firms refused to inspect labour abuses in Xinjiang because of lack of access to realities on the ground (Human Rights Watch, 2020), thereby weakening the credibility of the control of suppliers to the apparel industry. Indicative of deep problems, the apparel industry's own *Better cotton initiative* pulled out of certification. These moves revealed a lack of access to information on the human rights situation in Xinjiang, which made it pretty much impossible for global clothing brands to figure out if their Chinese suppliers were using forced labour.

2.4 Chinese reactions

The Chinese reaction has been a combination of rejection, partial accommodation, and counter-mobilization in formal political, industrial strategic, and monitory democratic modes.

One of the Chinese measures to regain control of its cotton industry work-life agenda has been to develop its own standards. On February 1, 2021, the Xinjiang region released a social responsibility report on its cotton textile industry (Textalks, 2021). The report was published by the Xinjiang Textile Industry Association "to clarify facts and build a communication bridge between Chinese and international stakeholders based on shared values and common interests". In March 2022, China launched its *sustainable cotton standard* as part of the country's backlash against Western criticism over the alleged use of forced labour (Ecotextile, 2022).

The controversy flared up again in 2021 as European countries imposed sanctions on officials in China with ties to the Xinjiang Uyghur repression. China retaliated

through official political channels but also through an orchestrated consumer rebellion with strong nationalist overtones.

A social media post by the Communist Youth League triggered Chinese mobilization on the Weibo microblogging platform (Global Times, 2021):

> *Spreading rumours to boycott Xinjiang cotton, while also wanting to make money in China? Wishful thinking!*

There was a huge backlash against the company, with numerous calls for people to boycott its products. The hashtag "I support Xinjiang cotton" was the top trending topic on Weibo, with more than 1.8bn views.

The State broadcaster CCTV said that H&M – a multinational firm with the headquarters in Sweden – had "miscalculated" in trying to be a "righteous hero" and that it "must pay a heavy price for its wrong actions".

In addition, various celebrities such as Wang Yibo, Huang Xuan, and Victoria Song released statements that they were severing ties with the Western brands, noting that "the country's interests are above all".

The Chinese counter-mobilization had economic consequences. At least three major Chinese e-commerce platforms – Pinduoduo, JD.com, and Tmall – withdrew H&M products from sale. China went from being ranked No. 5 on H&M's sales list in 2020 to being outside the company's top 10 for the third quarter of 2021 (CGTN, 2021).

2.5 The H&M response

Facing the grim consequences of its Chinese market operations, H&M, along with other Western apparel colleagues, took a very humble position. In March 2021 the company issued a plea, asking Chinese consumers to return. "We are working together with our colleagues in China to do everything we can to manage the current challenges," said the statement, which did not mention Xinjiang. "China is a very important market to us" (*New York Times,* 2021).

The caution also shines through in the 2021 annual report, in which H&M does not mention a single word about the Xinjiang scandal and the financial consequences of the Chinese boycott.

CNBC/Reuters (2021) points out that the Human Rights section of H&M's website, hmgroup.com, no longer carried the link to its 2020 statement on Xinjiang. This is in line with disappearing statements expressing concerns about Xinjiang previously seen on the websites of several Western fashion retailers, including Zara's owner, Inditex. Instead, H&M started communicating reconciliatory messages, such as (H&M, 2021):

> *We are working together with our colleagues in China to do everything we can to manage the current challenges and find a way forward . . . .*

Being at the forefront of innovation and technology, China will clearly continue to play an important role in further developing the entire industry. We are proud our suppliers are being part of that development and we want to continue contributing to driving progress together with our partners and stakeholders in the country . . . .

We are dedicated to regaining the trust and confidence of our customers, colleagues, and business partners in China. By working together with stakeholders and partners, we believe we can take steps in our joint efforts to develop the fashion industry, as well as serve our customers and act in a respectful way.

Other brands, including Hugo Boss and Asics, went even further and vowed to continue buying Xinjiang cotton, even after reports of alleged human rights abuses had led Western countries to impose sanctions on China: "Xinjiang's long-stapled cotton is one of the best in the world. We believe top quality raw materials will definitely show its value," German luxury fashion house Hugo Boss said in a statement posted on its official Weibo account (La Prensa, 2021).

2.6 H&M in a true dilemma

The human rights abuses in the heart of their Chinese cotton sourcing areas have put Western apparel brands like H&M in a real dilemma. If they fail to purge Xinjiang cotton from their supply chains, the apparel companies are confronted by Western-style monitory democracy, possible customer punishment in Western markets, plus political reactions both in the United States and the EU.

On the other hand, even the suggestion of a pullout from Xinjiang has triggered massive Chinese reactions, with severe consequences for sales in the world's largest growth market. China's evolution as an industrial nation has made it far more difficult for Western brand companies to dictate the terms of trade along human rights dimensions. As long as China was mainly a supplier to Western brands for products sold to Western consumers, they could credibly threaten to exit by sourcing elsewhere. However, as China has developed into a major consumer market in its own right, the exit strategy has become far more costly.

Furthermore, the growth of China's self-esteem as it starts competing with the United States for market power has increased the Chinese government's willingness to weaponize the country's consumer market against interference from Western multinationals. And given China and India's role as the countries that produce the most cotton in the world, replacing China alone in companies' supply chains would be a massive task.

2.7 Preliminary solutions

H&M's preliminary strategy appears to be cautious, low-key adjustments that refrain from principled positions on Western values and maintain a low profile. Another feature of the H&M approach has been to retreat to *classic philanthropy, more in line with Chinese expectations.* H&M is therefore looking to increase the proportion of donations and CR events in the country. The first donation after the Xinjiang episode went to support flood relief in Henan province.

2.8 Profiling sub-brands

The H&M Group also consists of several other brands that are independent of H&M. They have recently launched "Arket" and "Other Story" in China. The fact that they are not associated with the mother company's conflict with China seems to be a solution and an alternative way for multinationals to gain access to markets in authoritarian regimes they have been previously squeezed out of. As long as Chinese consumers do not see the link, the authorities can appear 'strong' by boycotting large global companies such as H&M, but the multinationals can still gain access through subsidiary/ sister companies.

3 The Huawei case

After spectacular growth in the Chinese home market following its foundation in 1987, Huawei evolved into a leading global information and communications technology (ICT) provider. The company's revenue grew from around 48 billion CNY in 2005 to over 891 billion CNY in 2020 (Huawei, 2005–2021) as it positioned itself globally and established long-term and stable cooperative relationships with world-leading operators such as Vodafone, BT, Telefonica, Orange and China Telecom (AR., 2006). In addition, the company scaled up to be a core competitor with Samsung and Apple in the mobile telephony market.

   With China's rapid economic growth and a massive, protected home market, the country and its leading companies became a major power factor in the international economy. Huawei became integral to this transformation, not only because of the company's success but also because its field of operation lies at the centre of communication and infrastructure control. Having been crucial in Western 4G telecoms markets, Huawei positioned itself for a leading role in the rollout of 5G.

3.1 Security issues across the democratic-autocratic divide

The Huawei case highlights the security and policy issues involved when critical communication infrastructure in democratic societies is auctioned out to a company with an authoritarian home base. Strong state control over industry and a culture of massive censorship of its domestic population in China make it difficult for Huawei to profile itself as a trustworthy communication provider in democratic countries. In addition, there is industrial rivalry as the United States, EU, and China compete for hegemony in lucrative communications markets.

With the tightening of political control under Xi Jinping's leadership, it became clear that the naive Western belief in globalization as a road to liberal democracy had partially failed. And as Chinese multinationals started conquering Western markets, strategic responses gradually tweaked global free trade towards mercantilism.

This shift was most forcefully fronted by the Trump administration, which singled out China as a rival authoritarian power and focused on preserving US control with core infrastructure and technology. And Huawei became a central victim. In May 2019, Washington put the company on its "Entity List" over national security concerns, barring the telecoms giant from doing business with US firms without a license.

Huawei's smartphone business – once on its way to challenging Apple and Samsung in Europe – suffered due to US sanctions that cut Huawei's devices off from Android, the Google-owned operating system. As Huawei lost music, maps, and other services from Google that handset buyers expect to see, pre-loaded smartphone sales outside China collapsed. The US subsequently lobbied its Western allies to cut out Huawei as a 5G telecoms supplier.

Throughout 2020 and 2021, European governments including France, Sweden, Romania, the Baltic countries, Belgium, and Denmark either banned Huawei equipment in key parts of the country's 5G network or required its operators to wean themselves off the company's kit. While Huawei built a legal case against excessive punitive measures, the EU Commission dismissed concerns about the measures being in violation of EU law (Politico, 2020).

At the beginning of 2021, Huawei's Brussels lobbyists were still optimistic that Europe's hunger for cheap, speedy 5G installation would win out over security concerns. They even had meetings lined up in the European Parliament to make their case. Those meetings got cancelled on February 24, the day Putin launched his all-out invasion of Ukraine. For many in Europe, the risk-benefit calculation regarding Huawei had changed overnight (Politico, 2022).

Under President Biden, pressure on Huawei only increased and spread to Europe. In October 2021, the European Commission issued a fresh warning against using Huawei technology to underpin 5G networks, which has been followed by many EU countries, and the UK government reaffirmed its requirement to strip Huawei equipment from British telecoms infrastructure.

Other democratic countries – like Canada, Japan, and Australia – also banned Huawei on 5G on the grounds of security. Ditto Poland, Slovakia, and the Czech Republic from the Visegrad Group (Observer Research Foundation, 2020; Pleschova, G. 2022). An exception is Hungary, which stands out with broad agreement for Huawei to develop 5G and smart city networks in the country (Developing Telecoms, 2021)

3.2 Controversy over semiconductors

The United States also sought to block Chinese ICT capability by barring its access to semiconductors, both from the United States and its supply chains. The restrictions require a hard-to-get license for the sale of advanced semiconductors to entities within China, largely depriving the country of the computing power it needs to train artificial intelligence (AI) at scale.

The first layer of restrictions involved the direct sale of American-fabricated semiconductors to specific Chinese companies, initially ZTE and then Huawei. This happened when each of the firms was added to the Commerce Department's Entity List, which bans companies from buying controlled US exports without a license (ZTE was later removed). But that move proved ineffective at cutting off Huawei's access to the advanced semiconductors used in 5G equipment, because the company could still purchase advanced logic chips from foreign fabrication facilities, known as fabs, in Taiwan or South Korea. American firms were also sometimes granted licenses to sell low-end chips to Huawei.

To plug some of these holes, controls on Huawei were then expanded to include a new Entity List Foreign Direct Product Rule (FDPR), which requires companies that use controlled US technology in their production processes to comply with US export restrictions. Over the next few years, Washington placed China's leading fabrication facility on the Entity List, further restricting its access to crucial SMEs.

The impacts of these restrictions over the next decade are deeply uncertain. If China's SME firms, fabs, and chip design industries are forced to work together and manage to survive the initial onslaught of restrictions, they may emerge both strong and fully untethered from US controls, as some analysts predict.

3.3 The Huawei solution

In November 2020, Huawei sold its budget brand, Honor, to a consortium of buyers including the government of Shenzhen, the city where its headquarters is located (CNBC, 2020). This was to ensure Honor survived because US sanctions on the Chinese technology giant had cut off supplies to key components and crippled its smartphone business, which at that time included Honor. Selling the brand allowed Honor to gain access to key components like semiconductors again (CNBC, 2020).

One of the ways for Huawei to move forward in Europe, in spite of the 5G setback in communication infrastructure, is to engage in ICT for logistics markets. Starting with a close EU ally, Hungary, Huawei has recently deployed its 5G private network at the East-West Gate (EWG) Intermodal Terminal located in Fényeslitke, Hungary. This is Europe's first intermodal logistics terminal that utilizes 5G technology in order to handle rail, air, and road freight transport logistics. The project is part of China's Belt and Road Initiative. Huawei partnered with Vodafone Hungary to provide the hardware for the 5G network, while Vodafone supplied its network support and platform. The 5G services provided by both communication service providers (CSPs) are presently being used for the terminal's internal communications in addition to the operation of cranes and self-driving vehicles.

Huawei is also drawing on experience from the 'smart terminal' at the Tianjin Port, east of Beijing. This is a data network built by Huawei which is reinventing itself as a supplier for self-driving cars, factories, and other industries; the company hopes that it will be less vulnerable to Washington's worsening feud with Beijing over technology and security. Although American tech giants have traditionally dominated the AI research rankings, in 2021 the Chinese firms – Tencent, Alibaba, Huawei, and State Grid Corp – took 4 of the top 10 spots in both volume and citations (Nikkei, Asia 2023).

4 Putin's war in Ukraine as a challenge to Western companies

The ultimate challenge of the democratic-authoritarian divide comes with war, as illustrated by the current Russian invasion of Ukraine. Businesses operating across the warring parties and their allies incur direct risks of war damage, indirect economic risk as a consequence of the political weaponizing of the economy, and customer/public responses to business strategies.

4.1 Pulling out of Russia

The Russian invasion in February 2022 triggered strong pressure on Western companies to exit Russia: firstly, through monitory democracy, as public sentiment against the Russian aggression grew, and companies saw themselves incurring large reputational costs; secondly, through pressure from investors; and thirdly, through sanctions from Western governments affecting their Russian operations.

The exit from Russia spans most sectors of the economy (NY Times, 2022): *Consumer goods and retail* – where Danone wrote off 1 billion EUR while exiting its dairy business in Russia. H&M is also in the process of shutting down its business in Russia; *Energy* – where most Western companies, such as BP, Exxon Mobil, Shell, and Equi-

nor, have sold, or are in the process of selling, their stakes in Russian energy; and *Finance* – with companies such as American Express, Deutsche Bank, Société Générale, and JPMorgan Chase in the process of pulling out. It's a similar story in other sectors such as *food, media, professional services, travel and logistics, and manufacturing.*

4.2 Weaponizing Russian energy

One of the moves made by Russia to counter Western weapons support Ukraine's self-defence was to weaponize energy. Having positioned itself as the major gas supplier to Europe, Russia had control of a vital resource for undermining the European economy and social welfare. Throughout 2022, Russia reduced gas supplies through a main pipeline, Nord Stream 1, for a number of months. In June, it cut deliveries through the pipeline by 75%, and in July Russia shut it down for 10 days, citing the need for maintenance. When it reopened, the flow was halved to 20 m cubic meters a day. And in late August, it shut down Nord Stream 1 entirely. In late September, both the Nord Stream 1 and Nord Stream 2 pipelines in the Baltic Sea were blown up in an act of sabotage, seen by most observers as likely to have been executed by Russia.

Figure 3: Natural Gas Price Europe (Source: Tradingeconomics, 2023).

Companies like Gazprom and Rosneft, which had taken on board Western-style corporate responsibility and published sustainability reports (Gazprom, 2021 and Rosneft, 2021), suddenly found themselves weaponized to undermine European energy supply. The high price of gas that resulted throughout 2022 – but particularly in August of that year and for much of the late autumn/early spring – has affected household budgets across Europe and driven up costs for manufacturing firms (Figure 3). Yet the EU

has managed in record time to wean itself off Russian supplies and re-establish more 'normal' gas prices.

4.3 Weaponizing the Western financial economy

The West has countered Russia's energy strategy by weaponizing the financial economy. The United States has barred Russia from making debt payments using foreign currency held in US banks. Major Russian banks have been removed from the international financial messaging system Swift. This has delayed payments to Russia for its oil and gas exports. The UK has excluded key Russian banks from the UK financial system, frozen the assets of all Russian banks, barred Russian firms from borrowing money, and placed limits on deposits Russians can make at UK banks (BBC, 2022).

The European Union has prohibited all transactions with the National Central Bank of Russia related to the management of its reserves and assets. Due to the ban on transactions from the EU and other countries, it is estimated that more than half of Russian reserves are frozen. The ban was also imposed by other countries (such as the United States, Canada, and the UK) which also store a share of Russia's foreign reserves (EU Concilium, nd).

4.4 Blockade and other sanctions

Restrictive measures against Russia were introduced with the Crimean and first Donbas invasions in 2014. After 24 February 2022, in response to Russia's military aggression against Ukraine, the EU, and the United States massively expanded the sanctions with the aim of significantly weakening Russia's economic base, depriving it of critical technologies and markets, and significantly curtailing its ability to wage war (EU Consilium, nd, U.S. Home Treasury, nd).

Both the United States and the EU have excluded Russian airlines and vessels from their airspace and ports with some selective exemptions. In addition, there are restrictions on numerous products, ranging from weapons to war-relevant technology, as well as sanctions against individual actors and entities, including the Russian leadership, members of the Duma and the National Security Council, oligarchs linked to the Kremlin, and other prominent business people, in addition to so-called "propagandists and disinformation" actors.

To hit Russia's economy, which is highly dependent on the import of services from European companies, the EU has prohibited the provision of certain business-relevant services to the government of Russia or to any legal actors, such as companies and other entities or bodies, established in Russia.

4.5 Business implications

The business implications are massive when the controversy between authoritarian and democratic regimes escalates into war and heavy sanctions. The Russian gas embargo has surely been a serious blow to European energy-intensive industries, as well as a contributor to inflation. However, a rapid capacity for transition to alternative energy and gas supplies, and energy saving, has softened the effect. Western oil and gas companies that have exited their Russian assets have incurred losses but have been more than compensated by the higher hydrocarbon prices provoked by the war. The price rise for hydrocarbons has likewise recompensed the Russian energy industry from some of the volume losses due to Russia's energy blockade of Europe.

The energy embargo has been particularly challenging for East European countries with traditional infrastructure ties to Russia under the Soviet era. The EU embargo on Russian oil has therefore exempted deliveries through the Druzhba pipeline. However, under strong pressure from the Polish government to end Russian supplies, Poland's energy group PKN Orlen in January 2022 stated that it will reduce Russian feedstock for its refineries to 10% in February (Nasdaq/ Reuters, 2023).

The Hungarian government has taken a very different position and is continuing its dependency on Russian oil and gas. The Hungarian refiner, MOL, is therefore continuing its reliance on Russian oil, and the Hungarian energy minister is planning extension of Russian oil supplies to Serbia (Upstream, 2022).

The massive financial sanctions directed at Russia have obviously had extensive negative effects on the functioning of the central bank as well as the bottom line of large parts of the finance industry. However, the IMF forecasts for the Russian economy have been adjusted upwards, indicating that the country has found compensatory solutions, and – as a major energy exporter – it benefits from the massive price increase for hydrocarbons.

Furthermore, many of the companies that have been pressured to leave Russia, both because of political sanctions and because of monitory democratic mobilization, have been compensated by the stock market (Market Watch, 2022).

5 Concluding discussion

As shown in the three cases presented in this paper, the age of liberal globalization is now being substituted by a divisive globalization marked by ideological confrontations.

In the H&M case a traditional compliance response to civic human rights abuses in China backfired and was met with a counter-attack from host authorities and monitory engagement by 'netizens' in defence of nationalist values.

In the Huawei case, commercial transactions under global free trade were set aside and substituted by security-based concerns. The full-scale Russian war attack on

Ukraine and the Chinese sabre-rattling every time Taiwan speaks to a democratic leader diminished democratic countries' enthusiasm for auctioning out critical communication infrastructure to firms with an authoritarian home base.

The Russian case shows how business may be turned into a coercion vehicle by the weaponizing of energy production and how the West weaponizes finance in response.

Stranded in the maelstrom of ideological contradictions, global business is now being pushed to develop a new repertoire of commercial bargaining and strategic configurations to cater to rising security concerns in the bipolar world.

5.1 Corporate responsibility strategies in a polarized world

In the process of developing a novel approach to CR one must consider that at the core of the authoritarian-democratic discords lie not only diversity of values but also collisions of governance structures. In terms of our governance model (Figures 1 and 2), we have shown how authoritarian and democratic governance have structural similarities but work in opposite ways. While the democratic organization model typically reveals a bottom-up bias, autocratic governance has a clear top-down structure.

This is illustrated in the operation of monitory democracy, which – in the Western case – relies on bottom-up civic initiatives, based on freedom of association and open society with media access to mobilize public engagement for environmental and social rights. Here CSOs typically play an important role in coordinating and organizing the public interest – as well as advocating a critical angle – in such processes. Western CSOs activist repertoire has included organizing boycotts and taking legal action to hold corporations accountable for environmental damage, human rights abuses, and breaches of labour law. CSOs' leveraging power lies in their capacity to use both social and mainstream media to expose corporate misconduct and promote corporate social responsibility.

Monitory engagement under autocracy, on the other hand, operates under a strong government leadership, where freedom of civic association is restricted unless it is carefully aligned with autocratically controlled institutions. CSOs or NGOs in authoritarian states are therefore typically instruments used by the state to control public discourse and maintain social stability while projecting an image of responsibility and legitimacy. Strictly speaking, the concept of NGOs is a misnomer in this case, because the organizations in question represent a species of incongruous, self-contradictory GONGOs (government-organized non-governmental organizations).

In a classical model of monitory democracy Western style, the Nordic company H&M was singled out as a prime target for the Western NGOs. The latter used the risk of bad publicity and brand damage in H&M's home market, to pressure the company to advocate the human rights agenda among their Chinese cotton suppliers. What was interesting in this case was that H&M was met not only by monitory democratic chal-

lenge from Western NGOs but found itself under an eastern GONGO attack as well. The effective internet action by the Communist Youth League pushed H&M and other Western apparel brands to back down from the critique of Uyghur work conditions in the cotton industry, praise the high quality of Chinese cotton, and more or less accept the Uighur work conditions.

Just like monitory engagement, political governance and regulation diverge across the democratic-autocratic divide. The Western democratic model implies an open competition between political parties and novel civic initiatives vying for popular support in transparent and fair elections. In contrast, the autocratic model operates under censored media and communication with candidates vetted by the authoritarian incumbent. In addition, the voting process is itself often rigged so as to produce pre-determined results.

Orchestrating CR across such divides is obviously a challenge, especially when the focus is on human rights, which include many features that authoritarian states find problematic. In line with the early neoliberal world order, such divides were often overcome through the power imbalance in favour of Western multinationals and liberal ideology. Under *heroic CR,* they could implement their own CR standards, if necessary, supervised by private certifiers and heroically impose them on their suppliers in emerging economies.

However, the dynamics are shifting in the world of bipolar globalization, where emerging economies are gaining strength. This shift challenges the previously dominant 'heroic' approach to CR, and Western multinationals operating across the geopolitical fault lines find it increasingly difficult to maintain their old strategies. Caught in the middle, they are compelled to develop adaptive strategies to navigate the new global landscape.

5.2 Emerging CR strategies

Our cases indicate how *several adaptive strategies* have emerged as corporations have struggled to meet the conflicting expectations of authoritarian and democratic regimes.

The first corporate strategy has been to bow down to authoritarian demands and reduce social responsibility to a *'harmless' philanthropy*. As indicated in the apparel industry case, this has largely been H&M's and the Western clothing industry's response to oppressive work conditions for Uighurs in Xinjiang.

Another corporate strategy has been to *bifurcate the global economy* in sensitive areas so that companies are able to choose between respective authoritarian and democratic regions. As indicated in the case of the Russian occupation of Ukraine, companies with democratic home bases have in many cases had to leave their Russian assets and build up supply chains in democracies, while companies with authoritarian home base have been pushed to rely on supply chains in autocracies or countries with hybrid governance.

The third strategy has been to establish *specialized certification schemes* to seek to uphold human rights standards in the Western company's supply chains from authoritarian regimes. Such special industrial regulation on top of the host state's general practice is very demanding and requires considerable bargaining power. The example of the Nordic garment industry shows that even the largest multinational companies failed in their efforts of acquiring relevant certification and gave in when faced with a shutdown in one of the world's largest markets.

Finally, the Huawei case underscores the challenges posed when companies with home base in authoritarian regimes lose trust in sensitive democratic telecommunication markets. The company ended up having to *retreat to less ideologically sensitive areas*, like port management in a semi-authoritarian state such as Hungary.

Common to all these strategies is that they weaken or completely undermine the effect of *heroic corporate responsibility* as conceived under Western-dominated globalization. With the strengthened position of authoritarian countries in the global economy, the Western multinationals' control of their supply chains and their CR is progressively undermined by host country values. Consequently, democratic and liberal societies must rethink their governance strategies and substantially revise their policies.

5.3 Redefining corporate responsibility in a bipolar world

Re-imagining CR for future development must face the fact that the shift of power has changed and that the G7 group – including mostly democratic countries – no longer rules the world. With emerging economies – many of them authoritarian or hybrid – there is a need for a shift in both economic, political, and socio-cultural orientation. A global economic environment that continues to be characterized by a mix of lingering neoliberal policies and increasing nationalistic tendencies needs redefining the CR agenda. Our cases indicate that when authoritarian economies mature so as to become attractive consumer markets for Western companies, they acquire stronger capacity to shape the social and environmental terms of trade. Practicing the heroic CR, where the Western brand company – instigated by international CSOs – exclusively sets the terms for their suppliers, does often not work any longer.

5.3.1 The case for pragmatic CR

To meet the complex, socio-economic and political challenges, the advocates of Western CR need to abandon their heroic modus and invest in what can be called a *pragmatic CR*. Such CR needs to develop capacity and resources to build a more complex strategic approach, often in partnership with governments. It must be capable of negotiating reciprocal gives and takes, such as market access, in return for social and

environmental compliance. The European carbon border adjustment mechanism is a recent example. One could imagine a similar mechanism for human rights. The idea is to be able to differentiate CR and related regulatory intervention in order to deal adequately with sector-specific confrontations and to meet reactions from authoritarian quarters with relevant counter-measures. For example, Chinese restrictions on communication across democratic-autocratic fault lines are not matched by adequate democratic responses. The so-called Chinese 'information wall' that bans Western internet, Google, Facebook, and others needs a clearer Western response. Ditto the Russian 'Foreign agent' legislation that bans international organizations and free media from engaging in the country has to be met with firm opposition from the democratic world.

Ultimately, it is possible to restrict the trading circle in cases where values collide too much, or where one is concerned with dependencies on the trading partner. CR may in such cases indicate reversal from 'offshoring' to 'friendshoring', where trade is restricted to partners with common human rights and environmental standards.

5.3.2 The dynamics of pragmatic CR

Globalization is an evolving process that significantly shifts the relative advantages among bargaining parties and impacts the dynamics of Corporate Responsibility from phase to phase. Initially, Western governments, businesses, and civil society organizations (CSOs) set the global economic agenda under neoliberal principles, extending CR initiatives to emerging markets through the supply chains of Western businesses. As emerging economies develop affluent middle classes and substantial consumer markets, they gain leverage against Western companies, which are increasingly drawn to these burgeoning markets. This shift allows the emerging economies to exert pressure for modifications in CR demands.

That said, a new era may be in the making. Emerging economies establish themselves not only as significant global consumers but also as globalizing producers, aiming to penetrate Western consumer markets. This development could potentially restore some bargaining power to Western civil society and politicians, allowing them to integrate CR stipulations into market entry conditions.

It is impossible to predict the precise outcomes of these dynamics with any certainty. What is clear, however, is that governments, businesses, and CSOs need to move beyond heroic CR by adopting a pragmatic CR approach. This approach may still uphold heroic CR ideals, such as human rights and environmental values, but it must maintain symmetrically negotiated trade relations built on a realistic evaluation of each side's bargaining power.

References

Theory references

EIU, nd, https://www.eiu.com/n/campaigns/democracy-index-2023/;https://en.wikipedia.org/wiki/The_Economist_Democracy_Index.

EU. Council (2023) Press release on agreement on: Corporate sustainability due diligence: https://www.consilium.europa.eu/en/press/press-releases/2023/12/14/corporate-sustainability-due-diligence-council-and-parliament-strike-deal-to-protect-environment-and-human-rights/ 14. Dec 2023. Retrieved 9 April 2024.

Freeman, E.R. (1984) *Strategic Management: A Stakeholder Approach*. Boston: Pitman.

Freeman, E., Harrison, J., Wicks, A., Parmar, B. (2010) Stakeholder Theory: The State of the Art. Cambridge; Cambridge University Press.

Fombrun, C. (1995) Reputation: Realizing Value from the Corporate Image. Boston; Harvard Business School Press.

Fukuyama, F. (1992) *The end of History and the Last Man*. New York; Free Press.

Global Compact, our participant. (nd) https://unglobalcompact.org/what-is-gc/participants Retrieved 20 February 2023.

Huntington, S. (2002) Clash of Civilizations And the Remaking of World Order. New York. Simon Schuster.

IMF (2023) Fragmenting Foreign Direct Investment Hits Emerging Economies Hardest https://www.imf.org/en/Blogs/Articles/2023/04/05/fragmenting-foreign-direct-investment-hits-emerging-economies-hardest Retrieved 9 April 2024.

International Trade Union Confederation ITUC (nd): 180+ Orgs Demand Apparel Brands End Complicity in Uyghur Forced Labour.https://www.ituc-csi.org/apparel-brands-Uyghur-forced-labour Retrieved 20 February.

Keane, J. (2013) 'Civil Society in the Era of Monitory Democracy' in Trägårdh, L., Witoszek, N and Taylor, B. (eds) *Civil Society in the Age of Monitory Democracy*. Oxford. Berghahn Books.

Kotler, P., Kartajaya, H. and Setiawan, I. (2010) Marketing 3.0: From Products to Customers to the Human Spirit. Hoboken. Wiley.

Midttun, A. (2022) *Governance and Business Models for Sustainable Capitalism*. New York. Routledge.

Midttun, A. Wei, C. and Wu, X (2014). 'CSR Across Developmental and Political Divides: A Comparative.

Analysis of China and the Nordic *Countries' Working Paper Centre for Corporate Responsibility*. BI Norwegian Business School. Oslo.

Porter, M.E, and Kramer, M.R. (2011) 'Creating Shared Value'. *Harvard Business Review*. January-February issue.

Quinn, R.E. & Thakor, A. V. (2018) "Creating a Purpose-Driven Organization». *Harvard Business Review.* July-August.

Ruggie, J.G. (2014) Global governance and 'New Governance Theory': Lessons from Business and Human Rights. *Global Governance 20*.

Schaltegger, S.,Lüdeke-Freund, F. and Hansen, E.G (2016) 'Business Models for Sustainability'. *Organization & Environment*. Vol. 29 Issue 3, p264–289. 26p.

Yin Robert K. (2018) Case Study Research and Applications: Design and Methods SAGE Publications. New York.

Khuong M. Vu (2020) Chapter 4: "Sources of growth in the world economy: a comparison of G7 and E7 economies" in: Barbara M. Fraumeni (ed): *Measuring Economic Growth and Productivity*. Elsevier, Beijing. Measuring Economic Growth and Productivity.

The H&M case

Business & Human Rights (nd) *H&M ends relationship with Chinese yarn producer over accusations of forced labour.* https://www.business-humanrights.org/en/latest-news/hm-ends-relationship-with-chinese-yarn-producer-over-accusations-of-forced-labour/. Retrieved 20 February 2023.

CGTN (2021) *China falls off H&M's top ten sales markets in Q3 2021.* https://news.cgtn.com/news/2021-10-02/China-falls-off-H-M-s-top-ten-sales-markets-in-Q3-2021-142edkdiT4I/index.html.

CNBC/Reuters (2021) *H&M vows to rebuild trust in China after Xinjiang backlash* https://www.cnbc.com/2021/03/31/hm-vows-to-rebuild-trust-in-china-after-xinjiang-backlash.html.

Ecotextile (2022) *China launches sustainable cotton standard*.

Freedom United (nd) *Call to Action: Coalition to End Forced Labour in the Uyghur Region* https://www.freedomunited.org/coalition-to-end-forced-labour-in-the-uyghur-region-faq/ Retr 20 Feb. 2023.

Embassy in New Zealand on the so-called "Assessment" on Xinjiang issued by the OHCHR. http://nz.china-embassy.gov.cn/eng/gdxw/202209/t20220902_10759496.html Retrieved 20 February 2023.

Freedom United nd Call to Action https://www.freedomunited.org/?s=Call+to+Action&x=0&y=0 Retrieved 20 February 2023.

Global Times (2021) *Chinese public voices support for Xinjiang cotton, turns to domestic brands over H&M, Nike*. https://www.globaltimes.cn/page/202103/1219414.shtml Retrieved 20 February 2023.

H&M Group, a (nd) Annual Sustainability Report. https://hmgroup.com/wp-content/uploads/2022/03/HM-Group-Annual-and-Sustainability-Report-2021.pdf Retrieved 20 February 2023.

H&M Group, b (nd) Supply Chain. https://hmgroup.com/sustainability/leading-the-change/transparency/supply-chain/ Retrieved 20 February 2023.

https://www.ecotextile.com/2022032129108/social-compliance-csr-news/china-launches-sustainable-cotton-standard.html Retrieved 20 February 2023.

H&M Group c (nd) *Statement on due diligence*. https://hmgroup.com/sustainability/fair-and-equal/human-rights/h-m-group-statement-on-due-diligence/ Retrieved 20 February 2023.

H&M Group d (2021) *Statement on H&M in China*. https://hmgroup.com/news/statement_hm_china/.

H&M Group (2022) *Annual Report*. https://www.annualreports.com/Company/hm-group.

Human Rights Watch (2020) Social Audit Reforms and the Labor Rights Ruse", https://www.hrw.org/news/2020/10/07/social-audit-reforms-and-labor-rights-ruse Retrieved 20 Feb. 2023.

La Prensa (2021) *Hugo Boss, Asics will continue buying Xinjiang cotton*. https://www.laprensalatina.com/hugo-boss-asics-will-continue-buying-xinjiang-cotton/ Retrieved 20 February 2023.

New York Times (2021) *Global Brands Find It Hard to Untangle Themselves From Xinjiang Cotton*. https://www.nytimes.com/2021/04/06/business/xinjiang-china-cotton-brands.html Retrieved 20 Feb. 2023.

Statement by the Chinese Embassy in New Zealand (2022) *Statement by the Spokesperson of the Chinese Embassy*.

Textalks (2021) *The Xinjiang cotton textile industry releases its Social Responsibility Report*. https://textalks.com/the-xinjiang-cotton-textile-industry-releases-its-social-responsibility-report/ Retr20Feb.23.

Tharawat Magazine (2018) Top Five Largest Fashion Retailers in the World. https://www.tharawat-magazine.com/facts/top-5-largest-fashion-clothing-retailers-world/ Retrieved 20 February 2023.

UN OHCHR (2022) *Assessment of human rights concerns in the Xinjiang Uyghur Autonomous Region, People's Republic of China*. https://www.ohchr.org/sites/default/files/documents/countries/2022-08-31/22-08-31-final-assesment.pdf Retrieved 20 February 2023.

The Huawei case

CNBC (2020) *Huawei to sell smartphone unit for $15 billion to Shenzhen government, Digital China, others, sources say*. https://www.cnbc.com/2020/11/10/huawei-to-sell-smartphone-unit-to-shenzhen-government-others-sources-tell-reuters.html Retrieved 20 February 2023.

CNBC (2021) *The smartphone brand that Huawei sold off is regaining market share in China.*https://www.cnbc.com/2021/09/29/honor-the-smartphone-brand-huawei-sold-off-regains-share-in-china.html Retrieved 20 February 2023.

Developing Telecoms (2021) *Hungary, Huawei sign MoU on long-term cooperation*. https://developingtelecoms.com/telecom-technology/enterprise-ecosystems/12205-hungary-huawei-sign-mou-on-long-term-cooperation.html Retrieved 20 February 2023.

Huawei (2005-2021) Annual Reoports. https://www.annualreports.com/Company/huawei Retrieved 20 February 2023.

Nikkei Asia (2023) *China trounces U.S. in AI research output and quality*. https://asia.nikkei.com/Business/China-tech/China-trounces-U.S.-in-AI-research-output-and-quality Retrieved 20 February 2023.

Observer Research Foundation (2020) *Czech Republic's CETIN does what the government couldn't: reject Huawei*. https://www.orfonline.org/expert-speak/czech-republics-cetin-does-what-the-government-couldnt-reject-huawei/ Retrieved 20 February 2023.

Pleschova, G. (2022) *Significant conversations on Slovakias cybersecurity: China, Huawei and the struggle for the countrys political orientation*. https://www.elgaronline.com/display/book/9781800371859/book-part-9781800371859-11.xml Retrieved 20 February 2023.

Politico (2020) *Huawei challenges legality of 5G bans in Poland, Romania*. https://www.politico.eu/article/huawei-hints-at-legal-action-against-5g-bans-in-poland-romania/.

Politico (2022) *How Washington chased Huawei out of Europe*. https://www.politico.eu/article/us-china-huawei-europe-market/ Retrieved 20 February 2023.

Putin's war case

BBC (2022) *What are the sanctions on Russia and are they hurting its economy?* https://www.bbc.com/news/world-europe-60125659 Retrieved 20 February 2023.

EU Concilium (nd) *EU sanctions against Russia explained*. https://www.consilium.europa.eu/en/policies/sanctions/restrictive-measures-against-russia-over-ukraine/sanctions-against-russia-explained/ Retrieved 20 February 2023.

Freightwaves (2023) *Welcome to the dark side: The rise of tanker shipping's 'shadow fleet'* . https://www.freightwaves.com/news/welcome-to-the-dark-side-the-rise-of-tanker-shippings-shadow-fleet Retrieved 20 February 2023.

Gazprom (2021) Sustainability Report. https://sustainability.gazpromreport.ru/en/2021/ Retrieved 20 Feb 2023.

Market Watch (2022) *Companies that exited Russia after its invasion of Ukraine are being rewarded with outsize stock-market returns, Yale study finds – and those that stayed are not* https://www.marketwatch.com/story/companies-that-exited-russia-after-its-invasion-of-ukraine-are-being-rewarded-with-outsize-stock-market-returns-and-those-that-stayed-are-not-11654258241 Retrieved 20 February 2023.

Nasdaq/ Reuters (2023) *Refiner PKN Orlen to cut Russian oil imports to 10% as of February*. https://www.nasdaq.com/articles/refiner-pkn-orlen-to-cut-russian-oil-imports-to-10-as-of-february Retrieved 20 February 2023.

NY Times (2022) *Companies Are Getting Out of Russia, Sometimes at a Cost*. https://www.nytimes.com/article/russia-invasion-companies.html Retrieved 20 February 2023.

Oil Price (2023) *The UK Loophole That Keeps Russian Oil Coming In*. https://oilprice.com/Energy/Crude-Oil
/The-UK-Loophole-That-Keeps-Russian-Oil-Coming-In.html Retrieved 20 February 2023.

Reuters (2022) *Poland's LPP sells its Russian company to a Chinese consortium*. https://www.reuters.com/mar
kets/deals/polands-lpp-decides-sell-its-russian-company-chinese-consortium-2022-05-19/ Retrieved
20 February 2023.

Rosneft (2021) Sustainability report. https://www.rosneft.com/upload/site2/document_file/Rosneft_
CSR2021_ENG.pdf Retrieved 20 Feb. 2023.

Trading economics (2023) Natural Gas EU. https://tradingeconomics.com/commodity/eu-natural-
gas?ref=vc.ru Retrieved 20 February 2023.

U.S. Home Treasury (nd) *Russian Harmful Foreign Activities Sanctions*. https://home.treasury.gov/policy-
issues/financial-sanctions/sanctions-programs-and-country-information/russian-harmful-foreign-
activities-sanctions Retrieved 20 February 2023.

Upstream (2022) *Hungary strengthens energy ties with Russia and Serbia*. https://www.upstreamonline.com/
politics/hungary-strengthens-energy-ties-with-russia-and-serbia/2-1-1334324 Retrieved
20 February 2023.

Section 3: **Practices: fragile or robust?**

Harald Throne-Holst

Chapter 7
Including societal actors in R&D – Different expectations, different responsibilities

The outcomes of research and development (R&D) have profoundly influenced modern societies. Not all outcomes have been good or beneficial. It is against this backdrop one can understand the call for governance of R&D. Because R&D is viewed to have great powers and potential to solve current or future challenges but also for the outcomes of R&D to turn out to have significant negative side-effects. This chapter investigates the role of societal actors, citizens, and stakeholders in such governance processes. It will discuss why societal actors are and should be involved, what are the drivers and barriers against the inclusion of them, and how inclusion can be done 'better' or more productively. Norway has an egalitarian and democratic culture, and there are strong norms surrounding social equality. This implies that there are expectations and anticipations from society in general to be included in matters of decisive importance. These expectations are also reflected in the guidelines for grant applications on emerging technologies from the Research Council of Norway, where the inclusion of stakeholders and members of the public is stressed.

1 Introduction

A central question on the governance of research and development (R&D) is who could be considered as legitimate actors that should be involved? Traditionally, the government (policy-makers) would turn to scientific experts. They knew the inside functioning of R&D. The scientific experts were considered able to, objectively as it were, assess what questions R&D should attend to, and how these questions could be addressed properly (Fisher, Mahajan & Mitcham, 2006, pp. 486–487).

However, over the last decades, the legitimacy of experts has waned (Beck, 1992; Owen, von Schomberg & Macnaghten, 2021). This has probably happened both gradually, grinding down their legitimacy over the years, and at other times more dramatically, in more pronounced shifts (in connection to various scientific 'scandals', e.g., BSE [mad cow disease], GMO [genetically modified organisms], Chernobyl). The expertise was primarily reliant on instrumental rationality, and broader social issues and concerns had little room in the deliberation between experts (Einsiedel, Jelsøe & Breck, 2001). As it became clearer that there were certain risks and uncertainties that were irreducible and that decisions still had to be taken in the face of them, it became apparent that the notion of experts needed to be updated. In the words of Nowotny

(2003, p. 152): *"(. . .), but the assertion that in modern societies there can be no safe way of making decisions".* In their search for more legitimate voices, policy-makers turned to the public and to stakeholders. A case in point is radioactive waste management. Over the last decades of the 20th century, the framing of it changed from being understood as a technical matter that somehow could be solved "to a controversial social problem in need of acceptance and legitimacy" (Bergmans, Sundqvist, Kos, & Simmons, 2015, p. 349).

Further, in modern societies, the grand societal challenges tend to be complex, and increasingly so. They cannot be addressed through traditional linear thinking and decision-making processes but rather through collaborative efforts involving a range of state and non-state actors (Voegtlin, Scherer, Stahl, & Hawn, 2022). Examples of such processes are climate change, the aging society, and global security. They are sometimes referred to as wicked problems, as the definition and location of the problems and their impacts on society, economic structures, and the environment are highly uncertain, and identifying appropriate actions is challenging (Rittel & Webber, 1973). Solving complex and systemic challenges stresses the need for a capacity to work together across traditional institutional and disciplinary boundaries. Wicked problems typically also affect a broad range of societal actors. This, in turn, necessitates that societal actors, governments, and businesses must work together, negotiate, and eventually agree on new and novel ways of doing things. And appreciate that uncertainty is an endemic property and that several types of knowledge are needed to create a common understanding and framing of the problems at hand (Kangas, Kujala, Lönnqvist et al., 2019).

1.1 Alignment

A central ambition of Responsible Research and Innovation is to make researchers and innovators work together with stakeholders and citizens over the research and innovation process to better align the outcomes of research and innovation processes with the needs, values, and expectations of society (Rome Declaration, 2014). It is this process of alignment that is the point of departure for this contribution: What does the venture of 'working together' look like in practice, and how could it be implemented in a meaningful way to tackle the grand societal challenges constructively?

Another frequently mentioned definition of this ambition uses a slightly different wording (von Schomberg, 2013, p. 63):

> Responsible Research and Innovation is a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products.

The phrase 'mutually responsive' can be considered a synonym of alignment, as 'mutually' points to a common, reciprocal arrangement where both or several parties commit. And 'responsive' has its roots in the noun response, so being responsive is about adjusting courses of action based on inputs from others or the outside. Using the frame of RRI, one would claim that the input R&D need is "recognising the insufficiency of knowledge and control" (Stilgoe, Owen & Macnaghten, 2013, p. 1572). However, for both these slightly different definitions, the emphasis is on interactions between different societal actors and technology enactors (innovators and researchers), as well as the changing responsibilities and relationship between them in the innovation process, from a traditional government process (top-down) to mutual responsibility (Jansma, Dijkstra & de Jong, 2022).

Alignment implies that it is not just one 'party' (here: science or society) that should make concessions but that it is rather an ongoing negotiation. And over such negotiations, both parties would have to make concessions if they were to take the call for alignment seriously. This, however, brings representativeness to the fore: Who can represent each party and claim to have legitimacy for negotiating and make concessions? Power and information asymmetries could be further complicating factors. R&D actors are more powerful as they have positions within the hierarchy of science and tend to use jargon that may alienate, frustrate, and push lay people into a passive position. Information asymmetry includes that R&D actors master the scientific background, relevant methodology, and an understanding of the limitations of current research and development (Borch & Throne-Holst, 2021).

Legitimate representatives of R&D are often 'given', as it would be the researchers that are involved in a particular project or innovation process. The researcher can be employed in businesses or at public universities. They can be funded through public and private funding. When it comes to identifying legitimate representatives of 'society', the selection criteria are often more nebulous and opaque.

1.2 Inclusion

In their seminal paper, Stilgoe, Owen, and Macnaghten (2013) suggest a framework for responsible innovation, the so-called AIRR framework: Anticipation, Inclusion, Reflexivity, and Responsiveness. These dimensions emerged from the authors' own experiences with public debates on new and novel science and technology (e.g., nanotechnologies). The authors indicated that the four dimensions would not happen sequentially but rather in an iterative process, going back and forth between the different elements. Still, I would argue that inclusion has primacy (Oftedal, Foss & Iakovleva, 2019). Not just as a condition to increase the legitimacy of the innovation process (Jansma, Dijkstra & de Jong, 2022), but even in the absence of such input from other parts of society through involvement and inclusion, the quality of the other three dimensions tends to be poorer. Outsiders' perspective represents an impetus to the responsibility agenda, as

those that are included at least would be perceived by researchers and innovators to have expectations that they keep up their efforts. So, whereas the three other dimensions, in principle, could be facilitated limited to the inner circle of a project, inclusion entails opening up R&D processes to people not directly associated with the project. Inclusions nurture the other dimensions.

Accordingly, in this chapter 'inclusion' will be used throughout for practices of organising public engagement, especially connected to governance of R&D. Although inclusion is often associated with concerns related to marginalised groups in society, the use of the term adopted in this chapter is wider and includes concerns for all societal actors. However, who are we really thinking about when we use the term "societal actors"?

1.3 Societal actors

The main groups of societal actors are members of the public and stakeholders. Before we go into broad categories for when the two are called upon, we turn our attention to how we can understand and define the two. A more nuanced approach to the rights and expectations of different groups is crucial when we are considering who to include and plan for the actual process.

'Members of the public' is necessarily a broad category, as 'we all' are included in this category. However, to make this meaningful in this context, we should limit the category to when we are not acting in our professional capacity. Put differently, it is when we are not representing our workplace but rather ourselves as private persons. Sometimes it could be difficult to take our 'hat off', in different situations, and forget in what role we are (expected) to be in the given moment. It is when we act as private persons or as members of our household that one can claim one is member of the public.

On the other hand, there are sometimes also calls for, or even a need for, inclusion of other, more specific groups of society. Some might call them synonyms to 'members of the public' and may use them interchangeably or even arbitrarily. Examples that are often used in calls for inclusion include consumers and citizens. There are different expectations, rights, and obligations to each of these roles. This is exactly why it matters to have more precise definitions of societal actor groups. They are different, and the inclusion of one over the other makes a difference. What we should reflect on is who it is that is important for a particular process.

Traditionally, consumers and citizens have been described as opposites (Närvänen, Mattila & Mesiranta, 2019): Consumers maximise their self-interest in the market and pursue more and more consumption to satisfy needs that are created by other market actors (producers, retail, marketers). At least according to neo-classical economy, consumers should not make political choices in the market but rather be rational and well informed on the range of available consumer goods, prices, and quality

and make their choices based on them. Consumers have certain rights and are pro-tected against reckless behaviour by retailers, marketing, and producers through laws and regulations.

Citizens, on the other hand, have traditionally been understood as members of a nation or political community and ones that participate in public life. Further they can be viewed as people who take collective responsibility to social and ecological commons (Johnston, 2008). Citizens participate in political elections, hearings, and local or national referendums (e.g., casting votes to legalise marijuana, municipal mergers, or membership in the EU). In a representative democracy the citizens vote for persons on the ballot that are supposed to represent the voters' (i.e., citizens') in-terests. Based on this, it has been argued that citizens are more altruistic, whereas consumers are considered more selfish (Turow, 2010).

Some have also argued for a hybrid notion of the 'citizen-consumers' to address political consumerism. Frustrated by the slow working and the compromises of the political system, consumers act on political beliefs through boycotting or buycotting[1], by carefully selecting products according to these beliefs. Examples include selecting organic produce and avoiding products with certain chemicals or products where child or slave labour are part of the value chain.

For the three roles or categories of societal actors discussed thus far, it is time to emphasise one crucial aspect: they can be active in societal questions and responsibil-ity-taking. Even on complex and complicated matters, citizens, consumers, or citizen-consumers are willing *and* able to make up their minds, form opinions, and even take action. Too often, they are dismissed by so-called experts as being ignorant, unknowl-edgeable, and not possessing the 'right' knowledge or education to contribute to the discussions. Such out-of-hand dismissal is often referred to as the 'deficit model' of public understanding (Sturgis & Allum, 2004).

Stakeholders, on the other hand, are traditionally contrasted with shareholders. Shareholders are those that own stocks in a business. Some typical stakeholders, how-ever, include governments, customers, owners, local community organisations, and employees of local or central administration. Rather than focusing on creating 'just' profit for shareholders, stakeholder theory would claim that businesses also would create value even for other, broader range of stakeholders (Strand & Freeman, 2015). Of particular interest for the topic of this contribution is the authors' use of the con-cept democratisation in business (referring to the title of Eric Rhenman's book (1968), *Industrial Democracy*), where businesses would actively engage and negotiate to ad-vance stakeholders' interest. The more recent concept of 'creating shared value' is, according to Strand and Freeman (2015), in essence, identical to the 'jointness of inter-est' concept, which they claim to be similar to the particular Scandinavian way of un-derstanding advantages of a co-operative approach between the business and its stakeholders. Strand and Freeman (2015) claim that Rhenman's understanding of stakeholders was a product and reflection of Scandinavian cultural norms and institu-tional structures, and that these norms and structures are more or less the same

today, which implies that Scandinavian businesses have adopted the same understanding of stakeholders even today .

1.4 Different roles for different societal actors

We could imagine that the three main groups of actors (stakeholders, citizens, and consumers) would be included for different means and ends. It appears more fitting to name them as 'triggers', in the sense that they prompt a call to include certain societal actors in certain roles. Trigger 1: to resolve a particular ongoing conflict, e.g., conflicts between fishermen and industry on quotas or pollution. This would typically involve *stakeholders* (fishermen's associations, environmental NGOs, industry associations, and government representatives). Trigger 2: to develop a policy for governing the effects of R&D on society, e.g., AI and deep fakes. In this case one would expect *citizens* to be included, as this to a limited degree is a 'consumer product', but rather, it is a political question on how to regulate productively and efficiently. Trigger 3: consumer products from a new and novel technology are already on the market (or on their way). In this case, one would expect *consumers* to be included, as they are or will be the first to experience this firsthand. Trigger 4: researchers would like to collect data from several discrete sources (e.g., geographical distribution of garden birds) or from various social environments where they necessarily do not have access themselves (e.g., minority culture). This would fall under the citizens science umbrella, and as the name implies, one would anticipate calls for *citizens* in this case.

Inclusive and participatory processes, however, are not straightforward. Central considerations include the number of participants that should be included to optimally nurture constructive discussions and the format of the interactions, and these should depend on the issues that are to be discussed. There have been both supporters and critics of the widespread use of inclusive processes, and next, we will delve into what can be viewed as the strengths and fragilities of inclusion.

2 The strengths of inclusion

In the introduction we saw that the trust in experts has dwindled over the last decades. This has prompted the policy-makers to include the public in the governance of controversial and novel innovations and technologies.

Scientific expertise is still vital for factual and descriptive information. However, when it comes to evaluating how we should act or judge based on this information and make decisions, it is something that basically belongs to democratic deliberations. And the more people are affected by a decision, the more they should have a say in the formulation of a policy to address how alternatives could be identified and imple-

mented (Kaiser, 2015). In other words, the more powerful a technological option is, the more essential it is to ensure that its governance isn't left solely to the scientists.

Broadly speaking, there are two ideal types of the knowledge status of members of the public as they are invited to inclusion processes: 1) they should not be specifically prepared or prepped on the subject matter beforehand, the rationale for which is (probably) that the participants are expected to reflect and reason in their personal capacity, as truly non-expert citizens, and 2) the members of the public should be prepared and receive 'objective' information in advance, in the form of leaflets, carefully crafted notes or short reports, or web resources. The rationale here is that this may increase the quality of the discussions. The participants are empowered to weigh different options and subsequently the output of them (e.g., decisions or a set of recommendations). In the GoNano project, information material in the form of brochures on the three topics (food, health, and energy) was made as background material for the citizen workshops[2].

However, these are ideal types, and one could imagine variations of this. In the RCN-financed BioZEment project the ambition was to investigate the viability of making a climate-friendly cement by using bacteria (Myhr, Røyne, Brandtsegg et al., 2019; Røyne, Phua, Balzer Le et al., 2019). In a series of focus groups, the participants (consumers[3]) only knew that building materials would be the topic of the focus groups. However, as the focus group progressed, the participants subsequently got to know more about the new technological option (the climate-friendly cement) in three blocks. They answered a survey after each block before they deliberated over the new and increasingly detailed information.

Inclusion could even be viewed as a contribution to the democratisation of R&D. As citizens and stakeholders are included in R&D one could argue that it serves three purposes (Throne-Holst, 2021). 1) Access ('Innsyn'): Through participation, the societal actors would get at least partial access to how R&D takes place and how researchers and others actually proceed. This could contribute to a (soft) governance of scientific conduct. As a result, of both this access, and even further through their interaction with researchers, societal actors would gain 2) Insight ('Innsikt') into (scientific) knowledge and become more scientific literate[4]. Then, finally, through the preceding steps (Access and Insight), the societal actors have acquainted themselves with the science, but not least with the scientists. The inclusion would open for the publics' 3) Input ('Innspill') to how R&D could be governed, and that scientists (and policy-makers) would listen carefully and implement appropriate measures (i.e., Responsiveness).

As critical voices for inclusion emerged, proponents pointed out that the perfect should not be the enemy of the good. We may not get perfect results from inclusion exercises or participatory processes, but it can still be good. If the processes manage to open up entrenched assumptions and challenge them, it can be a benefit in itself. Contributing to scientists' questioning and reflection over their more or less tacit assumptions can assist them in reviewing and changing them.

Inclusion has the potential to increase the diversity of voices in the governance of R&D. This is not to suggest that researchers do not have the capacity to see their project or innovation from a different viewpoint, that they somehow are not members of the public. However, as all of us probably have experienced, we are all inclined to group-think and possess limited capacities to view our own actions from the outside. So are researchers and scientists. Outsiders bring in different values and viewpoints from other parts of society. Including their voices will increase the social robustness of any research projects or innovation that at some point will be implemented by society or prevent innovations with undesirable side effects. This can be a good thing in itself, as it may bring new perspectives and approaches to the subject matter at hand. Greater inclusion and greater diversity of participants have the potential to give new perspectives and approaches. Engaging different populations enhances the democratic legitimacy of the process (Quick and Feldman, 2011).

In an article on how inclusion can be implemented in innovation processes based on findings from the GoNano project[5], the authors investigated how values deduced from citizen consultations were taken up by professional stakeholders (Jansma, Dijkstra & de Jong, 2022). The authors focused on one of the areas of the project, namely potential nanotechnology applications for health. They found that through focused, guided stakeholder workshops, innovative suggestions on how to include broader considerations emerging from the consultations with citizens can lead to innovative suggestions: A producer of an artificial pancreas for diabetes patients discovered that ownership of data was something they had not considered, and this was something that citizens were concerned about. However, in a different workshop on sensor technologies, it turned out to be difficult to include any input from the citizens' consultation, as it was hard to anticipate the potential impacts of this technology-oriented workshop compared to the application-oriented workshop on the artificial pancreas. The authors conclude that the design of a co-creation methodology influences which of the four dimensions of RRI it supports. Accordingly, they stress the need to define the aim of the exercise, that full transparency is required about the engagement (inclusion) process from recruitment all the way to the expected impact of the process, and that the design of the event should speak to these.

Finally, there appears to be a certain 'myth' or idea among certain actors in R&D, and even some policy-makers, that members of the public or stakeholders do not possess 'enough' knowledge of science or innovations to come up with valid and relevant viewpoints or opinions, the so-called "deficit model" (Rowe & Frewer, 2000). This implies that members of the public reportedly are not capable of formulating good and interesting arguments. This raises questions of how much scientific knowledge is actually necessary on the side of the public, and what valuable and relevant knowledges are needed to make decisions even in the context of uncertainties. Arguably, you do not need to know the inner workings of an internal combustion engine to come up with relevant viewpoints of traffic. Similar examples can be found in other scientific or technological arenas (Kaiser, 2015).

Another issue is indigenous knowledge. Traditional or local knowledges has historically been excluded, in parallel to the knowledges of members of the public and stakeholders. However, indigenous knowledge has risen in prominence and status as it has proven its value and usefulness (Wynne, 1996). Scientists and innovators tend to exaggerate the 'formal' scientific insights needed to come up with relevant viewpoints.

3 The fragilities of inclusion

A concern that came up rather early in the experimentations with various inclusive and deliberative processes on health care systems, environmental issues, and emerging technologies (like biotechnology) was to what extent the outcome of such engagement exercises had any impact on policy-making (Rowe & Frewer, 2000; Stirling, 2008). What could prevent policy-makers from just shelving reports from such participatory processes or exercises? If this happened on a larger scale, it would pose a substantive barrier to further inclusion processes. A comprehensive literature review (Abelson, Forest, Eyles et al., 2003) found that citizens' juries have information flows that tended to be unidirectional, from the organisers to the participants, rather than bidirectional, as the deliberative ideal suggests.

As the popularity and proliferation of participatory approaches continued to rise, questions emerged on the appropriate methods of participation (Stilgoe, Owen & Macnaghten, 2013), the appropriate purposes for which they could be used, and what criteria they could be evaluated to judge if it had been appropriate or not (Abelson, Forest, Eyles et al., 2003; Stilgoe, Owen & Macnaghten, 2013). Such criticisms developed further with more procedural issues: so-called 'framing effects' where the dialogue of the participatory processes reinforced certain ideas about professional power or the role of the public rather than opening up on such roles, relations, and power.

In the review by Abelson, Forest, Eyles et al. (2003) the authors find that the dominating, narrow theoretical frame for designing and evaluating a deliberative, inclusive process is based on two principles: Fairness (equal opportunities to participate) and competence (the quality of the information provided to the participants). However, the authors claim that this narrow frame fails to acknowledge the role of power. The idea that power can be somehow excluded in deliberative and inclusion processes is both wrong and potentially dangerous. Power can have various guises and exert influence through what information is presented or made available and through the organisers' or sponsors' organisational agendas and purposes.

If the goal is to democratise R&D, there is a need to increase and extend structured participatory processes. However, we cannot, and should not, automatically presume that this will only have beneficial and positive effects, let alone be sufficient (Stirling, 2008). Power and privilege are 'always' potentially at play when it comes to closing down assessment and evaluation or commitments. Closing down wider policy

discourses on technological choice is not always negative, but there is a need to critically assess who benefits from such decisions (Stirling, 2008).

There are also other issues related to inclusion: The call to include hard-to-reach groups. Many, if not most, inclusion processes strive to be 'representative' (Rowe & Frewer, 2000, pp. 12–13). As a starting point, this may appear to be an overshoot, as inclusion events usually will not include more than 30 participants. Anyhow, it can also be considered a good thing to go outside of the usual suspects and invite other groups. The usual suspects in this context are majority groups, typically white, middle-class, and retired participants. The hard-to-reach groups are minority groups, typically immigrants of low socio-economic status, working two jobs, with language barriers. Their voices tend to be overlooked, or at least not heard, on issues like governance of R&D. However, you could also end up amplifying your recruitment base through the so-called 'participation paradox' (Kern & Hooghe, 2018), where more available options for political participation/inclusion will mainly be used by more privileged groups (white, middle-aged, middle class).

A further challenge is inclusion fatigue or burnout, particularly in the following instances. a) It is hard to recruit someone outside of the majority groups, so you end up asking the same group of societal representatives to join your event or process. This is definitely the case even for stakeholders, as some are more popular than others, and they may be overwhelmed, and have to prioritise. They cannot answer and attend 'all' calls for inclusion. b) Potential participants decline to take part as they may not feel that their participation will make any difference, or c) decline after having participated, but end up feeling that the process was unsatisfactory or inauthentic.

If one were to take societal actors' lack of trust in science seriously, critics point out it will require a significant institutional shift in the policies for governing research and development (science and innovation). That is, if R&D actors should listen to societal actors on this point, it would entail true self-reflexivity on their own imaginations and assumptions on the part of their institutions. They need assistance from science policy, but that should not relieve them of (part of) responsibility. Public engagement activities risk falling short due to the "dominance of technocratic frames" (Kerr, Cunningham-Burley & Tutton, 2007) rather than more substantial engagement. Science and technical experts are arbiters and the authors cite (Collins & Evans, 2002, p. 271): "romantic and reckless extension of expertise has many well-known dangers –the public can be wrong". However, this may appear somewhat contradictory, as the dominance of technocratic frames ostensibly limits societal actors' sway over R&D.

Others point to the fact that the public views are more valid on certain issues or certain phases of innovation, but not so on others (welcome on debates on the future direction of medicine, but less so on production and application of medical knowledge) (Kerr, Cunningham-Burley & Tutton, 2007). And that citizens and consumers are set in a reactive rather than active modus in most inclusive processes. That is, they are presented with issues, dilemmas, or visions to which they only are anticipated to react instinctively, parroting previously held assumptions and prejudices. And they are, to a

lesser extent, as these critics say, actively taking part or being included in the situation or the process and co-creating, developing, and exchanging ideas or thoughts. And that such a "reactive mode" in contrast to an active mode would limit the potential for the democratisation of science (Kerr, Cunningham-Burley & Tutton, 2007).

There appears to be a tacit assumption that inclusive or participatory processes should produce outcomes on which the citizens and stakeholders agree, that the result would always be a consensus. This may not always be the case. Modern societies are indeed pluralistic societies, with large variations in values, cultural backgrounds, and knowledge systems. This implies that dissensus may be more common than consensus. How dissensus between societal actors could be managed and properly conveyed constitutes something of an Achilles' heel of these processes. It could undermine the exercise and leave the outcomes at the discretion of the participating R&D members or policy-makers. This is similar to the findings of Kerr, Cunningham-Burley, and Tutton (2007): for the public consultations these authors studied, the format implicitly privileged consensus and optimism.

In addition, Kerr, Cunningham-Burley, and Tutton (2007) found that criticism that questioned or challenged the purpose of the inclusion processes, or the technologies that were discussed, even forcefully, tended to be bracketed by other participants. They conclude that the lay positions were rather fragile, that they were easily compromised, so effortlessly subsumed and aligned with expert positions.

Another feature is that one often assumes that when societal actors come up with ideas, visions, or suggestions, these would immediately be actionable. In the GoNano project with alternate workshops with societal actors and professionals, this was found to be a major bottleneck. When the citizens and consumers who participated in these workshops formulated that all products should be sustainable, it did not represent applicable input to professionals who were working on rather specific innovations. The various inputs from citizens and consumers were either not considered relevant or too unspecific, as the example just given (Jansma, Dijkstra & de Jong, 2022).

4 Some suggestions on how inclusion can be done "better"

As we now have carefully considered the fragilities and the strengths of inclusion, we have found that there are complications surrounding inclusion processes. What are possible roads going forward? Below we sketch out four main issues to consider when devising a road ahead for inclusion: "Expectation management", "What's in it for participants?", "The design of the process", and "Accountability". These align rather well with four research questions suggested for the assessments of deliberative and participatory experiments and their contribution "to democratic and effective decision making" (Papadopoulos & Warin, 2007, p. 453):

> questions of openness and access (input-legitimacy) [i.e., The design of the process]; questions regarding the quality of deliberation (throughput) [i.e., What's in it for participants?], questions of efficiency and effectiveness (out-put legitimacy) [i.e., Expectation management], and ultimately their insertion in the public sphere (issues of transparency and accountability) [i.e., Accountability] (Papadopoulos & Warin, 2007, p. 453)

Designing inclusion processes to achieve desirable outcomes is ideal (Bryson, Quick, Slotterback, & Crosby, 2013). However, it may not be clear what such desirable outcomes are. They are somewhere on the range from achievable to ideal. And both in the design, the implementation, and the interpretations of outcomes, there are sensitivities to potential framing by enactors (Stirling, 2008).

4.1 Expectation management

Here, expectation management entails that promises made to citizens and stakeholders about what effects could be expected from their participation should be moderate and not excessively hyped to the extent outcomes would shift, i.e., national or regional policy. Of course, sometimes it might be the case, but organisers or initiators of such processes should be careful with regards to such expectations. Alternatively, it should be made sure that they have policy-makers or businesses on board and invested in the inclusion processes so that there is a reasonable possibility that they will include the outcome in their strategies. If not, the sponsors should explicitly explain to the participants why their concerns, hopes, or dreams are not considered relevant. Dismissing the outcomes from participation and inclusion out of hand and closing down informing and forming technological choices appear to be widespread (Stirling, 2008). Explicit accountability towards the participants may be a possible and effective antidote. Hyping the effects of inclusiveness risk leading to loss of trust in such processes, just as many feared that hyping of biotechnology could result in loss of trust and subsequently loss of public backing and support for biotechnology. However, there appears to be limited empirical evidence for this effect on biotechnology (Master & Resnik, 2013). And possibly, this may to some extent be valid even for promises made for the effects of inclusive processes in the governance of R&D.

4.2 What's in it for participants?

Particularly for the facilitators of the process that are discussed here, they should be attentive to what is in it for the citizens or stakeholders for their participation. Is it 'just' to be able to make their voice heard on their desires, dreams, concerns, or fears? Is the purpose to be better informed? Are their interests affected (losing their local hospital), or should they rather imagine or come up with further tangible benefits for their involvement (Abelson, Forest, Eyles et al., 2003)? A focus on what's in it for par-

ticipants as they sign up for inclusion processes is an important aspect to consider when designing the process. Explicating the positive aspect of being included will have a knock-on effect on the recruitment of relevant participants, including hard-to-reach groups. Potential participants live hectic everyday lives, and efforts to motivate them to participate need to be convincing.

4.3 The design of the process

The design of inclusive or participatory processes is somewhat of a challenge. One should be vigilant to avoid framing effects or avoid, unwittingly or not, steering or nudging the outcomes of the process towards specific, 'desired' ends. On this note, it is pertinent to remind the involved parties to be reflexive over the design process to strive to limit such framing effects. A literature review of the previously mentioned GoNano project found that a shared goal and mutual trust were key requirements to enable beneficial collaboration between stakeholders. Further, it suggested that the collaboration should be designed as a protected space to nurture mutual trust and facilitate experimentation. The participants should feel safe to express their meanings and share their thoughts. All participants need to have a stake in the issue at hand, and if potential participants do not appreciate that, the facilitators should make concerted efforts to elucidate why potential participants are considered to actually have a stake. Participants should be assisted to have a genuine influence in the deliberations (GoNano D1.1., 2018). Although the GoNano project focused on co-creation, the literature review had a broader scope on 'mutual learning'.

As has come up several times in this text, representativity is a key issue for inclusion. How can we understand, think, and reflect on what that would mean in practice? Participatory exercises have, by some, been viewed as a potential threat to numerical and representative democracy (Abelson, Forest, Eyles et al., 2003; Papadopoulos & Warin, 2007). In political elections, citizens are invited to cast their vote, and on the basis of that vote, political representatives are elected to parliaments. Whereas deliberative, participatory processes invite a limited number of citizens and stakeholders to identify what 'the public' has to say on a particular issue. One thing is to what extent this short circuits the idea of numerical democracy, with "one man[6], one vote". On the other hand, numerical democracy is connected to the idea of representatives of the voters in parliament. Could we interpret participatory and deliberative processes in that way, i.e., those who participate would be in some way representatives of the population at large? Or is it a more productive view that participatory, deliberative processes rather are extensions of the existing democratic structures and not alternatives per se (Einsiedel, Jelsøe & Breck, 2001)?

To further develop the framework to design and evaluate public participation processes, Abelson, Forest, Eyles, et al. (2003) suggest expanding the principles beyond fairness and competence. They suggest four principles: Representations, Procedural

rules, Information, and Outcomes/decisions. Each of these has several associated specific evaluation points. Under Procedural rules, there is an evaluation point that is relevant to this context (Abelson, Forest, Eyles et al., 2003, p. 244): "What point in the decision-making process is public input being sought (i.e., is the public involved in significant aspects of decision-making such as agenda setting or in minor decisions only?)". To put this differently, we could equate it with a spectrum of inclusion: from very little, and merely symbolic, to significant and substantiative. However, one should keep in mind that achieving substantiative inclusion on this point would be both quite demanding for the involved organisation, in terms of financial funds and time, and equally challenging for participants.

4.4 Accountability

A salient question is, who is or can be responsible for the outcomes? And who can be held accountable for the eventual implementation of the outcomes at the laboratory or in the market? One way of describing the public participants in deliberative processes is that they are experts in (their) everyday life in households. If one indeed chooses to label the public so, does that imply they should be held accountable for the outcomes of the participatory process, in parallel to the expectations we may hold other 'professional' experts (Langvatn & Holst, 2022)? This is, however, dependent on how we understand accountability.

There are different understandings in the literature, and their article on expert-reliant governance (Langvatn & Holst, 2022, p. 3) suggests the following understanding (p.3):

> Accountability thus understood refers then to a specific social relation between an actor, and a forum, where the actor is under an obligation, formally or informally, to explain and to justify his or her conduct, and the forum has authority to request information and explanation of the actor's actions in a domain, and subsequently to sanction the actions.

In our context this would then imply that participants would be expected to justify and further explain potential negative or detrimental effects, following the implementation of the outcomes of the participatory process by an organisation, like a business or local or national policy. This appears to be quite unreasonable. The participants would hold no sway over either how an organisation translates the outcomes to actions or how such an implementation would occur (if at all) (Stirling, 2008). Further, such obligations would likely constitute a serious and significant impediment to recruiting (future) participants.

Those that should be held accountable are those that sponsor the inclusive processes: policy-makers, businesses, or researchers. The hope is that they will take lessons from the inclusion of citizens and stakeholders. And as they do implement these measures, accountability follows.

The responsibility of participants from society, be they citizens or stakeholders, is threefold. First, they should have a responsibility to participate in the exercises described here. When asked or an opportunity arises, societal actors should carefully consider joining. Second, when they join such processes, they should be responsible for their conduct and interactions over the event. They should be expected to be sincere in the conversations, questions, and contributions and behave respectfully with everyone involved in the process, be it peers or researchers, policy-makers, or business representatives. Third, they are responsible as to try to reach a consensus. This will not always be possible; however, the responsibility is then to articulate the various sides and arguments of a dissensus.

Inclusion has been a feature of the governance of R&D for a number of years now. Further, it appears that experts have not yet regained the trust to govern these areas themselves. This reservation is likely a good thing, as has been argued here. So, as a final sentiment to this chapter, and perhaps even as a polemic against the criticism of Collins and Evans (2002) on the romanisation of the expertise of the public: The public is not always right, but nor are the experts. This means that societal actors and experts are dependent on each other for an efficient and effective governance of R&D. This insight provides fertile ground for inclusion and cooperation towards alignment.

References

Abelson, J., Forest, P.-G., Eyles, J., Smith, P., Martin, E., & Gauvin, F.-P. (2003). Deliberations about deliberative methods: issues in the design and evaluation of public participation processes. *Social Science & Medicine*, *57*(2), 239–251. https://doi.org/https://doi.org/10.1016/S0277-9536(02)00343-X

Beck, U. (1992). *Risk society: Towards a new modernity* (Vol. 17). sage.

Bergmans, A., Sundqvist, G., Kos, D., & Simmons, P. (2015). The participatory turn in radioactive waste management: Deliberation and the social–technical divide. *Journal of Risk Research*, *18*(3), 347–363.

Borch, A., & Throne-Holst, H. (2021). Does Proof of Concept Trump All? RRI Dilemmas in Research Practices. *Science and Engineering Ethics*, *27*(1), 7. https://doi.org/10.1007/s11948-021-00288-8

Bryson, J. M., Quick, K. S., Slotterback, C. S., & Crosby, B. C. (2013). Designing public participation processes. *Public administration review*, *73*(1), 23–34.

Collins, H. M., & Evans, R. (2002). The third wave of science studies: Studies of expertise and experience. *Social Studies of Science*, *32*(2), 235–296.

Einsiedel, E. F., Jelsøe, E., & Breck, T. (2001). Publics at the technology table: The consensus conference in Denmark, Canada, and Australia. *Public Understanding of Science*, *10*(1), 83-98. https://doi.org/10.3109/a036857

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

GoNano D1.1. (2018). *Building on the state-of-the-art: ex-post evaluation on mutual learning* (GoNano-deliverable, Issue. GoNano-project.

Jansma, S. R., Dijkstra, A. M., & de Jong, M. D. (2022). Co-creation in support of responsible research and innovation: An analysis of three stakeholder workshops on nanotechnology for health. *Journal of Responsible Innovation*, *9*(1), 28–48.

Johnston, J. (2008). The citizen-consumer hybrid: ideological tensions and the case of Whole Foods Market. *Theory and Society*, *37*(3), 229–270. https://doi.org/10.1007/s11186-007-9058-5

Kaiser, M. (2015). Public Participation as Opportunity and Challenge. In C. F. Gethmann, M. Carrier, G. Hanekamp, M. Kaiser, G. Kamp, S. Lingner, M. Quante, & F. Thiele (Eds.), *Interdisciplinary research and trans-disciplinary validity claims* (Vol. 43, pp. 123–141). Springer.

Kangas, A., Kujala, J., Lönnqvist, A., Heikkinen, A., & Laihonen, H. (2019). Introduction: Leadership for dealing with complex changes. *Leading change in a complex world: Transdisciplinary perspectives*.

Kern, A., & Hooghe, M. (2018). The effect of direct democracy on the social stratification of political participation: Inequality in democratic fatigue? *Comparative European Politics*, *16*(4), 724–744. https://doi.org/10.1057/s41295-017-0093-y

Kerr, A., Cunningham-Burley, S., & Tutton, R. (2007). Shifting subject positions: Experts and lay people in public dialogue. *Social studies of science*, *37*(3), 385–411.

Langvatn, S. A., & Holst, C. (2022). Expert accountability: What does it mean, why is it challenging – and is it what we need? [Early online view]. *Constellations*, 16. https://doi.org/https://doi.org/10.1111/1467-8675.12649

Master, Z., & Resnik, D. B. (2013). Hype and Public Trust in Science. *Science and Engineering Ethics*, *19*(2), 321–335. https://doi.org/10.1007/s11948-011-9327-6

Myhr, A., Røyne, F., Brandtsegg, A. S., Bjerkseter, C., Throne-Holst, H., Borch, A., Wentzel, A., & Røyne, A. (2019). Towards a low CO2 emission building material employing bacterial metabolism (2/2): Prospects for global warming potential reduction in the concrete industry. *PLOS ONE*, *14*(4), e0208643.

Nowotny, H. (2003). Democratising expertise and socially robust knowledge. *Science and public policy*, *30*(3), 151–156. https://doi.org/10.3152/147154303781780461

Närvänen, E., Mattila, M., & Mesiranta, N. (2019). Consumer-citizens as leaders of change: Case food waste. *Leading change in a complex world: Transdisciplinary perspectives*.

Oftedal, E. M., Foss, L., & Iakovleva, T. (2019). Responsible for responsibility? A study of digital e-health startups. *Sustainability*, *11*(19), 5433.

Owen, R., von Schomberg, R., & Macnaghten, P. (2021). An unfinished journey? Reflections on a decade of responsible research and innovation. *Journal of Responsible Innovation*, *8*(2), 217–233.

Papadopoulos, Y., & Warin, P. (2007). Are innovative, participatory and deliberative procedures in policy making democratic and effective? *European journal of political research*, *46*(4), 445–472.

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169.

Rome Declaration. (2014). *Rome Declaration on Responsible Research and Innovation in Europe*. https://digital-strategy.ec.europa.eu/en/library/rome-declaration-responsible-research-and-innovation-europe

Rowe, G., & Frewer, L. J. (2000). Public Participation Methods: A Framework for Evaluation. *Science, Technology, & Human Values*, *25*(1), 3–29. https://doi.org/10.1177/016224390002500101

Røyne, A., Phua, Y. J., Balzer Le, S., Eikjeland, I. G., Josefsen, K. D., Markussen, S., Myhr, A., Throne-Holst, H., Sikorski, P., & Wentzel, A. (2019). Towards a low CO2 emission building material employing bacterial metabolism (1/2): The bacterial system and prototype production. *PLOS ONE*, *14*(4), e0212990.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580.

Stirling, A. (2008). "Opening up" and "closing down" power, participation, and pluralism in the social appraisal of technology. *Science, Technology, & Human Values*, *33*(2), 262–294.

Strand, R., & Freeman, R. E. (2015). Scandinavian cooperative advantage: The theory and practice of stakeholder engagement in Scandinavia. *Journal of Business Ethics*, *127*, 65–85.

Sturgis, P., & Allum, N. (2004). Science in Society: Re-Evaluating the Deficit Model of Public Attitudes. *Public Understanding of Science*, *13*(1), 55–74. https://doi.org/10.1177/0963662504042690

Throne-Holst, H. (2021, 7.12.2021). Tillit og medvirkning. *Forskningspolitikk*, *2021*(4), 26–27. https://www.fpol.no/wp-content/uploads/2021/12/Fp_4_2021_RGB.pdf

Turow, J. (2010). Segment-Making and Society-Making Media: What Is a Good Balance?

Voegtlin, C., Scherer, A. G., Stahl, G. K., & Hawn, O. (2022). Grand Societal Challenges and Responsible Innovation. *Journal of Management Studies*, *59*(1), 1–28. https://doi.org/https://doi.org/10.1111/joms.12785

von Schomberg, R. (2013). A vision of responsible research and innovation. *Responsible innovation: Managing the responsible emergence of science and innovation in society*, 51–74.

Wynne, B. (1996). May the sheep safely graze? A reflexive view of the expert-lay knowledge divide. *Risk, environment and modernity: Towards a new ecology*, *40*, 44.

Elin M. Oftedal, Matthias Kaiser, Tatiana Iakovleva
Chapter 8
Do you value responsible innovation?

> When change is easy, the need for it cannot be foreseen;
> when the need for change is apparent, change has become expensive,
> difficult, and time-consuming. (Collingridge, 1980)

Responsible Innovation (RI) represents a transformative shift in the approach to innovation, prioritizing societal needs and ethical considerations over purely market-driven objectives. This paper explores the challenges and opportunities inherent in this paradigm shift, focusing on the integration of diverse stakeholder values into the innovation process, particularly in the context of e-health technology. While the literature has focused on the importance of inclusion, there is less focus on the challenges of inclusion. Further, the different stakeholders in an innovation process may have very different values that drive their thinking and action, something that may hamper the inclusion process. Drawing on qualitative research methods, including interviews and participatory workshops, we investigate the value landscape of stakeholders ranging from end users to industry representatives and municipalities. Our analysis reveals the complexities of stakeholder engagement and highlights the importance of understanding the value landscape in shaping communication and collaboration within the innovation ecosystem. We propose a nuanced approach to responsible innovation, grounded in an understanding of the cognitive, affective, and conative dimensions of stakeholder values. By integrating these insights into innovation governance frameworks, we aim to foster a more inclusive, equitable, and impactful innovation ecosystem. Our study contributes to a deeper understanding of the role of values in responsible innovation and offers practical insights for policymakers, industry stakeholders, and researchers seeking to navigate the complex socio-ethical landscape of innovation in the digital health sector.

1 Introduction

Responsible Innovation (RI) is shifting the way we view innovation, from solely catering to customers and markets to addressing broader societal needs and concerns (Blok and Lemmens, 2015; Von Schomberg, 2019). This evolution in our approach to innovation also brings forth some challenges. First, RI demands that innovators question how new ideas may shape current practices towards the systemic transformation and impactful change required in society. Further, firms are asked to assess the potential for unintended consequences and negative externalities in the innovation pro-

cesses (Barben et al, 2008; Collingridge, 1980; Rip, 2018, Sveiby, 2017). For instance, an innovation that is highly valued by end users for its convenience might raise concerns among other stakeholders (i.e. policymakers, public) regarding privacy and data protection. The economic and social implications of these potential misalignments are profound, leading not only to financial losses but also to a delay in achieving potential advances in public health (Bradonjic, P., Franke, N., & Lüthje, 2019). The opportunity cost of investing in misaligned technologies is significant, potentially diverting resources away from solutions that could offer more substantial benefits to society (Bradonjic, P., Franke, N., & Lüthje, 2019).

The RI framework (discussed in the next section) suggests a solution to these challenges: to initiate a dialogue among various stakeholders to collaboratively navigate the path towards impactful societal change and to avoid misalignments of the innovation (Von Schomberg and Hankins, 2019).

However, industry must optimize the usage of their limited resources, and the inclusion of diverse stakeholders in the innovation process may be challenging as groups of stakeholders – users, customers, bureaucrats, NGOs, and laypersons – come from different contexts, have different interests and drivers to their actions, and do not necessarily understand the importance of inclusion into the innovation process (Schroder; 2020; Thapa et al., 2019; Eiken and Oftedal, 2024, Skeie, Leersum and Oftedal, 2024). Further, a common denominator among the different stakeholders may be differences in values, which could lead to poor communication among them. Understanding the stakeholder value landscape may therefore be an important first step in the inclusion process.

Value landscapes are described by Kaiser (2022) as an intricate and growing field focusing on how overlapping and even conflicting values may impact stakeholders' communication and collaboration. Values are held by all individuals in society but are often buried deep within individuals' psyche, so we are often not aware of each other's value. We know that some values change over time and are context dependent, but that some are stable based on a person's personality, experience, and education. We also know that people's value drives their behaviour (Schwartz, 1992; Erikson, 1959). Understanding the value landscape may facilitate more effective and empathetic stakeholder engagement by revealing underlying value-driven motivations. Moreover, by understanding stakeholders value, one can also identify potential areas of conflict or synergy and thus have a better starting point for good collaboration that further can contribute to the development of more inclusive and sustainable practices, communication, policies, and innovations that resonate with the diverse values of all actors involved. Values may therefore be essential in guiding responsible innovation, underpinning the ethical and societal considerations that lead to the creation and application of innovations.

Acknowledging the intricate and diverse viewpoints of different stakeholders, our objective is to delve deeper into the values of a unique set of stakeholders. These are important to the innovation processes in e-health. The stakeholders range from end users of e-health technology, meaning digitalized technology used for health purposes,

to the municipalities which purchase and also use the technology, to the industry which innovates such technologies, each bringing their distinct values to the ecosystem. Our purpose in mapping this landscape is to identify the similarities and differences in values among stakeholders and consider how these differences might affect their interactions.

Thus, our research question is centred *on mapping the value landscape among diverse stakeholders in the innovation process around e-health, highlighting the significance of values in shaping communication and collaboration within the innovation ecosystem.*

To achieve this, our methodology integrates qualitative research techniques, drawing on tools such as interviews and participatory workshops to gather deep insights into the value systems of the actors involved using a form of interpretative phenomenological analysis (IPA). Our context is the healthcare sector, and we aim to look at the inclusion of different stakeholders in the innovation process as well as in the debate about the future of digital health within the living lab (Norwegian Smart Care Lab), which works to create an arena where different stakeholders can meet and it is possible to include users early on.

1.1 From user to stakeholder perspective

Responsible Innovation can be described as an offshoot from the broader concept of RRI (Responsible Research and Innovation) and focuses on the realm of industry, looking more narrowly at innovation processes (Koops,2015; Owen et al. 2013; Stilgoe et al. 2013). The frameworks of Responsible Research and Innovation (RRI) may be described as guiding principles for integrating scientific endeavours and innovation (Owen and Pansera, 2019). However, although intertwined and sharing common objectives to enhance societal outcomes and tackle ethical issues, RRI and RI have developed in distinct ways (Owen and Pansera, 2019).

For example, RI is grounded in academic discourse which contrasts with RRI's policy-centric origins (Owen and Pansera, 2019). Further RI delves into innovation within business contexts and therefore focuses less on research and more on the process itself. RI is defined as a collective commitment to future care through present innovation stewardship and as a novel approach that balances social, ethical, economic, and environmental considerations (Grinbaum and Groves, 2013) and, as such, includes an element of the firm's social responsibility (Owen and Pansera, 2019; Barben et al., 2008)

User inclusion, defined as the active involvement of end users in the innovation process (Von Hippel, 1995), is central to both RRI and RI. The concept of user inclusion is, however, not new, but it has for a long time been a fundamental step in the innovation process (Von Hippel, 2005; Chesbrough, 2003). By incorporating users' feedback and insights, innovators can create products and services that are more aligned with user needs, increasing satisfaction and fostering a sense of community and belonging

among the target audience. This method, typically realized through user-centred design and co-creation methodologies, involves end users directly in the development process, helping to ensure the resulting innovations are usable, appealing, and effective (Von Hippel, 2005).

However, inclusion in the RI paradigm suggests not only end users or customers but also a broader group of stakeholders, including everyone who might influence or be affected by the innovation. Their focus is to understand not only the innovation-in-use but also the societal implications of the innovation. Each of the stakeholders holds a different position relating to the innovation and therefore brings unique values, expectations, and concerns to the table, which can significantly influence the direction and impact of innovation (Freeman, 1984).

The inclusion can therefore be performed through a variation of activities from small-group processes of invited public dialogue in the form of focus groups, consensus conferences, deliberative mapping, and citizen assemblies – what Goodin and Dryzek (2006) usefully call mini-publics – of more official governance arrangements in the form of multistakeholder partnerships, citizen forums, the inclusion of lay members on scientific advisory committees, user-centred design, and other hybrid mechanisms. This comprehensive approach may help innovators anticipate and navigate ethical dilemmas, societal impacts, and potential conflicts of interest that might (Oftedal, Iakovleva, and Foss, 2019).

However, stakeholders cannot just be seated in the same room and expected to participate in innovation processes. Stakeholders may come from very different contexts, may have different values, and thus may not readily understand each other (Chapter 7 extensively looks at the challenges of stakeholder inclusion). Understanding stakeholders value landscape may be an important step for understanding how to design good processes, and thus we will take a deep dive into this area.

1.2 Values for stakeholders

A deep understanding of these values fostering public trust and acceptance is important for the adoption and successful implementation of new technologies (Stilgoe, Owen, & Macnaghten, 2013).

Stakeholders' values include the beliefs, priorities, and ethical considerations of all parties involved in or affected by innovation processes, guiding innovation towards broader societal goals and ethical standards. Kaiser promotes understanding group dynamics through the lens of a value landscape, challenging the fixed and universal values assumption by highlighting the need for empirical validation. Drawing on classical philosophy, particularly Leibniz and Kant (Hilgard, 1980), he emphasizes the cognitive, affective, and conative dimensions as essential for comprehending human actions and forming attitudes. Kaiser (2022) argues that values are central to these dimensions, outlining that values are interconnected, vary in intensity, change

meaning based on context, and evolve over time alongside beliefs (see also Chapter 2 by the same author). We employ this framework to understand the different values of the different stakeholder groups.

In the realm of responsible innovation, the cognitive aspect involves analyzing the knowledge and perceptions that shape technological advancements. This means weighing the risks, benefits, and moral considerations of new technologies based on our understanding and beliefs. The affective aspect looks at the emotional responses innovations trigger in stakeholders, including trust, fear, or excitement. Meanwhile, the conative aspect focuses on acting in a way that is aligned with one's values. It encompasses the motivational aspects of values that drive behaviour. This component is crucial because it turns values from abstract concepts into actionable drivers that can influence decision-making and behaviour in tangible ways.
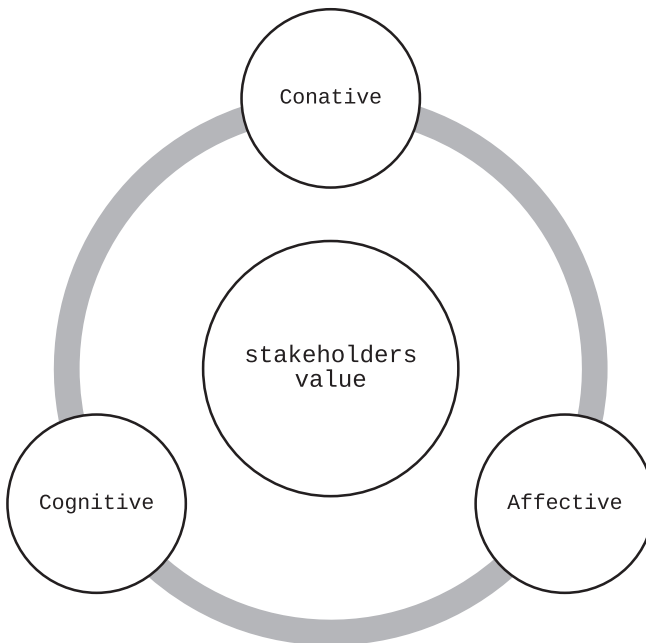
Figure 1: Stakeholders Value Landscape.

The deep-rooted relationship between values, action, and communication underscores the significance of understanding and articulating values clearly, as they ultimately shape the landscape of human interaction and societal progression (see Chapter 4 for a further discussion on values).

2 Methods

The landscape of values contributing to RI is shaped by a wide array of stakeholders, each bringing their unique perspectives, priorities, and ethical considerations to the table. Given this multifaceted nature of values among stakeholders, we adopt a qualitative methodology that emphasizes the exploration of values from multiple stakeholder perspectives.

We begin by examining the context of the study, which revolves around e-health. This context is intriguing due to the fundamental value-laden nature of health, encompassing ethical traditions, political dimensions, and technological developments. Additionally, it involves a diverse array of stakeholders, spanning from vulnerable individuals such as patients and their families to influential entities like the public sector and including innovators in the field. To fully grasp the implications of these developments, it is essential to comprehend the organizational structure of Norway's healthcare system, particularly in the context of assistive care.

We further discuss the type of data we have retrieved from the different stakeholders identified before we provide an analysis of the data through interpretative phenomenological analysis (IPA), a methodology focusing on how respondents make sense of their experiences with health technology (Smith, 2009; Finlay, 2011). Finally, we discuss the findings of this study.

2.1 Context

This study situates itself amidst substantial shifts in the Norwegian healthcare sector, caused by notable demographic shifts across the nation through an aging population. The prevalence of these health issues signals a crisis, and thus this situation reflects a pivotal moment in addressing the evolving healthcare needs within the country (Helgesen & Herlofson, 2017). This trend escalates the demand for health and care services, which currently account for one-third of a municipality's gross operating expenses (SSB, 2019). These challenges necessitate new assistive care technologies and the adoption of novel organizational frameworks. The municipalities in Norway right now are grappling with choosing and implementing these technologies. In addition, new technology is being developed by the industry and made available to the users. E-health technologies are a vast landscape, but one can group them as follows: Data Management and Access, which includes systems like EHRs and patient portals; Telehealth Services, offering remote medical services and patient monitoring; Mobile Health (mHealth), featuring apps and wearable devices for health management; and Clinical Support Tools, such as decision support systems and AI-enhanced diagnostics. In this plethora of challenges and technology available, the Norwegian Smart Care Lab attempts to create arenas where different stakeholders can come together to understand

each other's need with regards to this technology. Especially, they aim to include the users. However, in doing that, it becomes clear that different values are steering the different stakeholder groups.

2.1.1 The Norwegian public sector

In Norway, the primary healthcare service aims to deliver fundamental medical services and care to the populace. It is structured to ensure accessibility for all residents, acting as the initial interface with the healthcare system. This section of the municipality focuses on purchasing and implementing Data Management and Access in addition to Telehealth technology.

Central to the primary healthcare service, the Assistive Technology Center focuses on aiding individuals with both temporary and permanent disabilities. These can purchase and distribute other types of technology such as Mobile Health technology (i.e. medicine dispensers, safety alarms, and communication technology). They also distribute tools and technology that might not be high tech such as walking aids and pillow systems. Its primary mission is to foster active and autonomous living for people with disabilities through the provision of technical aids. The Assistive Technology Center, under the auspices of NAV (the Norwegian Labour and Welfare Administration), plays a crucial role in bridging the gap between individual users, local health and care services, and the national welfare state's objectives of inclusivity and accessibility.

The distribution of assistive aids in Norway is a collaborative effort involving multiple stakeholders, including the primary and specialist healthcare services, NAV, the industry, users, and their families. Each party plays a distinct role, ensuring that individuals receive the assistive aids they require following a comprehensive evaluation of their unique needs.

2.1.2 The industry

The Norwegian health technology sector includes diversity of companies, ranging from agile small- and medium-sized enterprises (SMEs) to large, well-established firms with extensive product portfolios and strong financial resources. The sector is distinguished by an environment for innovation and public-private partnership cooperation. What sets this industry apart is its startup landscape, which is often deeply user driven (Oftedal et al, 2019). Many startups within this sector are founded by individuals with personal experiences of healthcare challenges, either as patients themselves or as next of kin. This firsthand insight into the healthcare system's complexities fuels the development of solutions that directly address real-world needs, ensuring that innovations are both relevant and impactful and align with responsible innovation principles.

Alongside the startups, there are several established companies that have been operating in the sector for a long time. These companies have developed significant expertise and a good reputation in specific areas of health technology, from medical equipment to digital health solutions. These companies have more tools and resources available in their innovation process and are often concerned about ethical questions, though being directly aware of the concept of responsible innovation. The companies have a broad focus area when it comes to technology, as several companies within Norwegian health technology focus on developing specialized products that address niche needs within the healthcare system, but at the same time, there are also companies that offer broader health technology solutions designed to integrate with existing healthcare systems.

2.1.3 The innovation ecosystem

The innovation ecosystem within the e-health sector comprises the university sector, governmental funding programs, and hybrid organizations. The system plays a crucial role in fostering research, development, and the commercialization of new technologies and solutions by supporting entrepreneurs, startups, and established companies at various stages of innovation. The innovation ecosystem is aimed at providing help and support for the innovation activities going on.

Within this ecosystem, the Norwegian Smart Care Cluster (NSCC), funded through public grants, focuses on e-health organizations and includes 500 members nationwide. It serves as a collaborative community for stakeholders, including businesses, municipalities, hospitals, academia, and investors, aiming to leverage health technology for change and foster the growth of Norway's health industry. The Norwegian Smart Care Lab (NSCL), a pivotal component of NSCC, offers testing resources and support to companies integrating innovative welfare technologies. It bridges the gap between technology development and application, catering to healthcare providers, municipal services, and the private sector. NSCL facilitates product verification to expedite market readiness, ensuring user-centric, legally compliant solutions that meet industry standards, thereby promoting healthcare innovation.

Academia also plays a vital role in the innovation ecosystem, not only through education but also by laying the foundations for new technologies through research. Beyond traditional observation, academic researchers have actively participated in the innovation process especially focusing on embedding principles of responsible innovation to influence operational philosophies and practices. The dynamic interaction between researchers and the innovation environment, particularly in labs, exemplifies a "living laboratory" where responsible innovation principles are explored in real time. Reflective practice ensures a critical understanding of how research activities co-construct knowledge and influence practices within the innovation ecosystem, highlighting the ethical dimensions of research engagement.
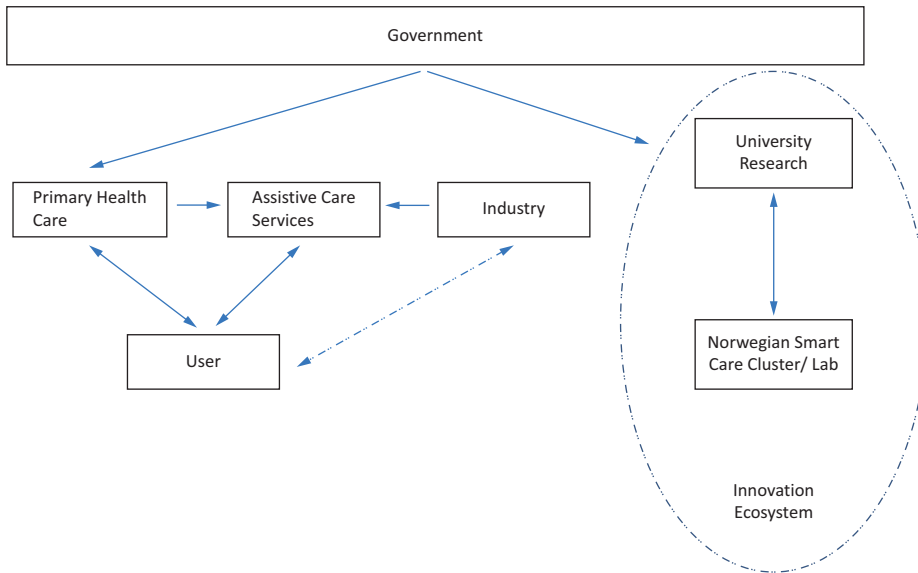
Figure 2: The organization of health care, assistive care technologies, and technology development in Norway.

2.2 Methods to gather our data

In the exploration of how technological solutions can better cater to the needs of the elderly, a variety of data collection methods have been employed to ensure a comprehensive understanding of the challenges faced by this demographic, as well as the potential solutions that could be developed. The following table, Table 1: Sources of Data, outlines the diverse approaches taken to gather insights directly from the elderly, as well as from stakeholders in the industry and public sector who play a crucial role in the development and implementation of these technological solutions.

The data are primary data, but not gathered just for this study. Instead, they are interviews and observations that have been gathered throughout a 4-year NFR-funded project on the topic of user inclusion called "Releasing the Power of the User". Throughout the project, the interviews we performed were semi-structured and most lasted about 1 hour.

User interviews have been a fundamental source for this study: 20 interviews conducted with elderly individuals in the southwest of Norway and in east Norway. The interviews lasted for an hour and participants were recruited through a snowballing process. The discussions with the elderly respondents have provided in-depth insights into the everyday challenges and specific needs of the elderly population, offering a direct perspective that is invaluable for the development of user-centred technologies.

The User Cafe serves as an innovative platform for engagement (Thomas, Skeie and Huang, 2022), where elderly individuals, representatives from the industry, and public sector officials come together in an informal setting to discuss technological solutions over coffee and cake. This initiative has been conducted in three sessions, each attracting between 15 and 25 participants. The casual atmosphere of these workshops encourages open dialogue and the sharing of ideas, making it an effective method for gathering diverse perspectives. The technology tested covered all the main categories of e-health technology.

Interviews with municipal representatives were conducted throughout the study. We focused on the needs of the municipality and the opportunities and challenges that were presented to them. The interviews selected for inclusion in this study yielded particularly relevant insights into the value landscape. Similarly, numerous interviews were conducted with representatives from the industry, where we focused on the innovation process and the opportunity and challenge of user and stakeholder involvement and inclusion. Again, we chose the interviews that were relevant for understanding the value landscape. Moreover, the selected interviews encompassed a range of perspectives, including those from established companies as well as startups.

Collectively, these methods represent a multi-layered approach to understanding the intersection of technology user inclusion.

Table 1: Sources of Data.

	Description	Numbers
User interviews	Interviews with elderly people about challenges and needs.	20
User Cafe	An arena where elderly, industry and public sector come together to talk about technical solution around a coffee and cake. Between 15 25 participants on each workshop.	3
Interview with municipality	interviews with municipal representatives to discuss local government perspectives and needs in implementing technical solutions for the elderly.	3
Interview with industry	Interviews with industry stakeholders to understand their perspectives, challenges, and contributions towards developing and implementing technical solutions for the elderly.	5

2.3 Analytical method

Data analysis would follow the IPA methodology, focusing on how respondents make sense of their experiences with health technology (Smith, 2009; Finlay, 2011). First of all, the IPA methods require researchers to immerse themselves in the data, and as

such, we have read and discussed the interviews and other data several times to gain a deep understanding of participants' experiences.

Further, in line with the IPA methodology, we generated codes. First, we coded inductively in a bottom-up approach from the data. We discussed the codes and emergent themes within our team. However, to understand our results in line with our theoretical framework, we then deductively linked the themes to our dimensions of values, including cognitive, affective, and conative. Here *cognitive values* refer to values understood and processed at an intellectual level. Information concerning knowledge, understanding, and interpretation in relation to values was thus categorized within the cognitive dimension. Secondly, *affective values* pertain to the emotional connection and attachment to values. As such, information related to emotions and feelings regarding values was coded within the affective dimension. Finally, *conative values* encompass values associated with actions and behaviours. Information in the data linked to motivation and the intention to act were therefore categorized within the conative dimension.

These themes capture the essence of participants' experiences and perspectives. Through this analytical endeavour, we sought to provide a comprehensive understanding of the respondents' values and perspectives (Smith, 2009; Finlay, 2011).

3 Main (preliminary) results: the value landscape

The analysis results revealed a value landscape characterized by both distinct differences and similarities among the stakeholder groups. While we didn't quantify the intensity of values (Kaiser, 2022), our analysis successfully highlighted the unique characteristics of each stakeholder group's values.

3.1 Values of individual users

Users embrace a wide array of technologies to address their diverse health and wellness needs. This includes wearable devices that track various metrics such as steps taken, sleep patterns, and pulse rate, providing valuable insights into daily physical activity and overall well-being. On the other hand, specialized technologies are employed, designed to address specific health challenges, such as safety alarms for emergencies, hearing aids to improve auditory function, and medicine dispensers to ensure timely and accurate medication intake. Further, we also didn't discriminate low-tech technology (i.e. walking aids, etc.), as this type of technology can also be important to understand the user's value landscape. Finally, sometimes users talk about non–health-related digital tools, and we have also not discriminated against that. Our data revealed certain values around these technologies.

3.2 Cognitive values

Cognitive values among users highlight an engagement with technology, balancing curiosity and practicality combined with an awareness of the limitations and implications of digital advancements. Very importantly, there's an evident openness to embracing digital tools for enhancing information access and managing daily tasks to improve and simplify everyday life and, not least, to keep living as before as opposed to being vulnerable and sick: *"For me, the most important thing is to use technology to live like before. The internet can help with that. To be able to be online and have access to mail and that kind of thing".* At the same time, the openness towards technology is tempered by an acknowledgment of the realities most face, captured in the reflection, *"Most people do not live in smart homes, most live in ordinary homes."* This statement underscores a gap between the idealized potential of technology-infused living spaces and the more commonplace, traditional dwelling environments. Also, most respondents show an openness to adapting technology for personal health and safety. This forward-thinking approach to technology adoption reveals an understanding of its value and a readiness to incorporate useful innovations into their life as needed.

The conversation around technology also brings to light concerns about the role and limits of digital solutions. A poignant observation shared by a user, *"I believe in human contact, I must say that technology will never be able to take over,"* speaks about the important value placed on human interactions. This perspective emphasizes a collective understanding that, despite the convenience and efficiency technology offers, it should not and cannot supplant the fundamental human need for connection and support. Concerns about the rapid pace of technological advancement and its potential to overshadow personal connections resonate through the discussion. This anxiety about the loss of the personal touch in an increasingly digital world underlines a widespread apprehension about technology. The insistence on the importance of human interaction in navigating this landscape is a powerful reminder that technology, for all its benefits, should enhance rather than diminish the human experience.

Central to the cognitive values is the significance of family and social networks. Many users underscore the indispensable support and guidance people give each other. *"It is important for me to help people who have difficulties with their health and with technology".* It's a call for empathy and understanding towards those less acquainted with digital tools, advocating for community engagement to raise awareness and accessibility of technological resources. This approach not only fosters inclusivity but also strengthens communal bonds by bridging the digital divide. The cognitive values show that users embrace new technology, especially to live the way that they are used and to help others do the same; however, they have a strong desire to not use human contact.

3.3 Conative values

Conative values in this area concentrate on the factors that drive and inspire action. The central motivations seem to be that of steering individuals towards self-reliance and independence. For the end users of e-health technology, this is based on underlying desires for autonomy and privacy, with users valuing technology that addresses their challenges while fostering independence.

Technologies that enhance mobility and sociability are particularly valued, as they counteract isolation and enable continued social engagement. This transcends mere physical or financial ability to perform tasks without assistance, advocating for technologies that enhance an individual's capability to live. *"That's exactly what it does, you know, you get a completely different freedom. You don't have to sit at home waiting for someone."* It's about empowering individuals with tools that support daily living, prioritizing devices and aids that bolster personal decision-making autonomy and enhance the quality of life without fostering dependency: *"We usually buy all tickets online. It's been a long time since we've been to the concert hall physically to buy tickets or at the theatre. It's all online shopping."* The same is seen when using services at the doctor's office: *"It is so fantastic that we do not have to go all the way to the Dr. office to get a prescription"*.

The value of self-reliance and freedom promotes a positive but cautious attitude towards welfare technologies, emphasizing the importance of not becoming unnecessarily reliant on external support systems. It promotes the cautious use of technology to supplement, not substitute, one's capabilities, encouraging the selection of technologies that empower users.

3.4 Affective values

Affective values delve into the emotional and relational dimensions of technology interaction, especially within communities. For users, affective values are very strong. This dimension underscores the importance of a sense of belonging, active participation in communal life, and the value placed on social and family support. However, there is also a palpable concern over mastering technology, with a preference for solutions that include human support to mitigate anxiety and ensure safety. *"So it's one of the nightmare scenarios you have, that you can not keep up"; "When something doesn't work, I get sick to my stomach, because I have no one who can help me"* . . . Affective values reflect the emotional ties and the importance of human connections in the adoption and use of technology. For example, fear when the technology doesn't work is present throughout the data and family often feels like an important support for the different types of technologies. The significance of family and social support emerges as a crucial affective value, with a deep appreciation for the advice and support received from family members. This support network plays a vital role in navi-

gating the technological landscape, offering a safety net that encourages exploration and adaptation to new technologies. *"I probably would have had the help of a son-in-law who is very technical and installed the things that I am used to today".* These technologies help respondents live the life they appreciate and are non-invasive, *"We get pictures from family every day. Facetime, is that what it's called*?"

There is also a narrative around image and identity that may highlight the struggle between maintaining one's established self-image and adapting to new health circumstances. This struggle often contrasts with the stigma associated with signs of aging or dependence. This dichotomy underscores a societal challenge where technology adoption can evoke feelings of pride or shame, influenced by one's proficiency and its impact on self-image. For example, on one side, elderly users prefer using technologies that affirm youthfulness and vitality, such as modern gadgets and automated home devices. However, when it comes to medical aids, they often feel they are stigmatized, *"Why do medicine dispensers have to look so grey and ugly*". It became apparent that medical equipment that signalled age or sickness were shunned, unless the equipment became unavoidable due to health situation. Another example is associated with walking aids such as sticks: *"When I meet cabin neighbours, they comment on me walking with sticks.' Then I say that it is not forever, it is only out for a short while ".* Another comment was, *"I will never be seen using walking sticks, I will only use walking sticks in my grave".* Similar comments were made about other types of technology such as hearing aids or safety alarms. This means that users do not wish to use technology that is stigmatizing until their health requires that they do. This can sometimes be an issue as it is easier to train people with technology when they are not sick and vulnerable.

As such, in navigating the technological landscape, elderly adults face a complex interplay of cognitive, conative, and affective values. They seek to uphold their identities and autonomy while embracing technologies that offer practical, non-stigmatizing solutions and foster social connections. This intersection reveals the nuanced considerations elderly adults make in choosing technologies, balancing the desire for independence and self-sufficiency with the need for social support and safety.

3.5 Public sector

The municipality integrates e-health technologies into healthcare delivery against a backdrop of the significant challenges of scarce funding and human resources – a condition that is only strengthened by an aging population that places increasing demands on the system. Understanding the interplay of cognitive, conative, and affective values in this context highlights may be central to promoting a better dialogue between the different stakeholders.

3.5.1 Cognitive values: understanding and evaluation

The cognitive approach to e-health technologies in the municipality is shaped by a recognition of the current healthcare landscape. The technological differentiation and integration process is informed by an awareness of the dual challenges of an aging population and resource scarcity, *". . . the amount of resources are very limited."* This understanding guides the municipality's exploration of digital health monitoring and other technological solutions that promise to enhance care while optimizing the use of scarce resources.

With limited financial and human resources, the municipality faces a pressing need to find innovative solutions that can alleviate the strain on healthcare services and still cater to a rising demand for healthcare services. Municipalities also value e-health solutions that empower individuals to take an active role in managing their health and well-being. After implementing a solution: *"People who receive services from environmental services, such as individuals with developmental disabilities, instead of having home visits in the evening, sometimes receive a video visit. There has been very positive feedback from staff. Very positive feedback from users on that, and it has undoubtedly been an efficiency improvement."* The municipality prioritizes initiatives where they can provide broad and identical solutions to a wide target group. Through evaluating technological investments, the municipality critically assesses the potential of digital solutions to streamline healthcare delivery and improve patient outcomes. The point is to know what types of care can be solved with which type of technology.

3.5.2 Conative values: action and implementation

Faced with financial and staffing limitations, the municipality's actions towards deploying digital technologies reflect a pragmatic yet innovative approach to healthcare service improvement. *"We are very clear in our priorities: Technology should save money, not create more work for health workers"*. The implementation of digital solutions, such as electronic health records and home monitoring tools, is aimed at addressing the immediate and future needs of an aging population. The pursuit of digital technologies is driven by a conative commitment not just to enhance healthcare delivery but also to do so in a manner that respects the municipality's resource limitations. *"We must prioritize scalable, flexible, and interoperable solutions because our world changes so fast";* the municipalities are aware of the fact that they will live with their current choices for a long time and that they have a huge challenge in their capacity to provide care, despite the ongoing challenges of funding and manpower.

The use of public resources to provide accessibility to healthcare services for all residents, regardless of their location or socioeconomic status, is an important aware-

ness in the municipalities. They prioritize initiatives that ensure fair and easy access to healthcare and that improve the efficiency in the sector.

3.5.3 Affective values: emotional and attitudinal aspects

Amidst these challenges, the municipality's approach to e-health technologies is also characterized by an emotional depth that acknowledges the gravity of the situation. The cautious optimism about the potential of these technologies to improve health-care services is tempered by a realistic understanding of the limitations imposed by financial and human resource constraints. *"We really have to be very sober with the introduction of technology. It must clearly contribute to operating these services further and not just expand them. That's my worst fear."*

This balanced perspective underscores a deeper affective commitment to finding solutions that are not only technologically advanced but also accessible and practical for both healthcare workers and residents. *"For us, it's very important to use technology where it actually makes sense to use it, where it really has a purpose for the users, un-doubtedly a purpose for some, but at the same time, someone has to follow up on it in the next round".* Further, while users are occupied with what type of technology can im-prove their lives, the municipality is focused on how the technology can help them be self-sufficient. For example, one respondent made an example of users that wanted chairs so that they did not have to move very often but pointed out that the users would actually benefit from moving more often and therefore *"one thing is what they want but another is what they need."* The municipality also valued solving its responsi-bilities in health challenges by not overburdening its employees. Therefore, they are not focused on tailoring technology or if the technology is stigmatizing. The municipal-ity wished to help as many as possible while not exhausting its resources.

In conclusion, the municipality's approach to integrating e-health technologies into healthcare services amidst financial and staffing challenges exemplifies a com-mitment to new initiative- and values-driven solutions. Through a blend of cognitive understanding, conative action, and affective empathy, the municipality strives to ad-dress the needs of its aging population and overcome the systemic challenges of re-source scarcity, highlighting a path forward that is both pragmatic and hopeful.

3.6 The industry

The industry is multi-faceted with small and large businesses, young and more estab-lished. However, the values we have captured, we have discussed among ourselves and also with industry members. Therefore, we believe that these may be recognized across many companies within the sector.

3.6.1 Cognitive values

At the cognitive level, the industry focuses on its commitment to innovation and research orientation. The industry often has a keen interest in resolving a certain challenge for the user. This commitment is not a mere pursuit of technological solutions for its own sake but a deeply ingrained understanding of the complexities inherent in healthcare needs. *"It's a whole field of study, so you have to have a lot of respect for reaching out, tests and quantity training and getting sort of your finger on the pulse yourself to be able to know how to make things better."* The industry values its ability to understand and solve challenges. They relate this ability to them being flexible and knowledgeable.

The industry also values that the public sector is trustworthy and predictable. As trust is integral to the functioning of any society, firms also initially value trust between the public and private sectors. Recognizing the importance of predictable and stable markets reflects a knowledge-based appreciation of how trust impacts regulatory compliance, tax collection, property rights, and overall economic functioning. Firms value the predictability of the public sector because it allows for strategic planning and risk assessment based on established norms and expectations . . . However, this trust can sometimes be compromised when the municipality develops solutions that the industry looks upon as its focus. *"We think the industry should make the solution and the municipality should make the framework condition."*

3.6.2 Conative value dimension

The health technology industry comprises a wide range of companies, from the well-established to the emerging startups. The driver for these entities is to balance the necessity to achieve financial sustainability and profitability. The way they can achieve this is by solving real problems in the sector. *"We're dedicating substantial resources not only because it's the right thing to do but because it's foundational to our financial sustainability and our ability to bring impactful solutions to market."* The industry is unique in that the 'customer' encompasses two distinct stakeholders: municipalities, who purchase solutions, and end users, who directly benefit from the technology. These stakeholders have divergent interests and capabilities, presenting the industry with the dual challenge of devising solutions that adequately serve both parties. *"We constantly have to ask ourselves who the customer is because we aim to make user involvement becomes part of the company's culture."* We also see signs that as the firms grow, there are challenges to staying as user oriented as they were in the startup phase: *"I think we were more user centric in the startup phase. As we start to grow, we have to be realistic in who is paying for the products and we have to focus the use of our resources."*

As such, the industry may have idealistic visions to help the end users, but their focus on survival can make this challenging. The primary challenge for these evolving companies is to sustain their user-centric ethos amidst expanding into broader market territories and satisfying the demands of the public sector, *"In the end, we have to survive, and align with what our customer want."*

3.6.3 Affective value dimension

Many startups and established companies in the health sector are distinctive for their origins; they are frequently founded by individuals directly affected by the health conditions their technology seeks to address. Termed "patient-innovators" (Zejnilovic; Oliveira and Canhao, 2016), these founders, along with their close associates or "next of kin," bring a unique perspective to the challenges faced by users.

Consequently, these companies often have a deep understanding of user needs and place a high value on maintaining close relationships with their end users. However, as these companies grow and become more established, it often comes with a higher prioritization of the customer (public sector) as opposed to the end user. Further, the companies that put great emphasis on involving users in the development process do it with a degree of frustration as involving users can be challenging: *"Our reflections can be negatively charged in the process, but the experience afterwards is always positive".*

The initial motivation for many of these health sector ventures is deeply rooted in personal experiences and the earnest aspiration to address real-life issues: *"I made this technology because my mother was sick, I know what the issues are . . . ."*

This grounding in personal connection and dedication to solving user challenges is indicative of affective values, where emotional commitment and a sense of purpose shape the company's mission and innovation approach.

3.7 Proximity and distance

In examining the relationships between users, municipalities, and the industry within the context of technology deployment and development, several interactions unfold, and they are displayed in Table 1. The relationship between users and municipalities is characterized by a notable distance in perspectives across cognitive and affective dimensions, elucidating a fundamental discord in the perception and application of technology, since users are focused on individual experience, including worrying about stigma and focusing on community support while the municipality focuses on system improvements. However, there is a greater alignment in the conative dimension, underscoring the need for user autonomy.

The interaction between users and the industry presents a more nuanced picture, where a moderate level of proximity emerges, particularly in cognitive and initial affec-

tive realms. This alignment suggests a shared recognition of technology's potential to enhance daily life and address specific needs. However, as industries grow and their market orientations become more pronounced, a moderate distance begins to materialize, especially in terms of conative and, over time, affective values. This evolution points to the complexities of maintaining user-centric approaches amidst expanding market demands and the necessity of balancing innovation with practical application.

The dynamic between municipalities and the industry reveals a high degree of cognitive proximity, driven by mutual interests in systemic improvements and the broad application of technology for public benefit. Nevertheless, this shared vision encounters challenges in the conative and affective dimensions, where divergences in implementation strategies and motivations for technology adoption appear. These differences highlight the nuanced negotiations required to reconcile the industry's drive for innovation to have a unique and effective offering and to serve their multiple stakeholders with municipal priorities for efficiency and scalability against a backdrop of resource constraints and public service obligations.

Table 2: Proximity and Distance of values among user groups.

Dimension	Users ↔ Municipalities	Users ↔ Industry	Municipalities ↔ Industry
Cognitive	**High Distance:** Users and municipalities differ in their view of technology's role and accessibility. Users focus on practical enhancements, while municipalities focus on systemic improvements and broad applicability.	**Moderate Proximity:** Both show a keen interest in technology's potential for problem-solving, though users emphasize personal enhancements, and industry looks at broader innovations.	**High Proximity:** Both prioritize technology for systemic improvements, though industry's drive for innovation slightly diverges from municipalities' focus on service delivery.
Conative	**Moderate Proximity:** Users prioritize autonomy and personalization. Municipalities also focus on autonomy, but diverging in emphasis on efficiency and scalability without much customization.	**Moderate Proximity:** Users desire technology that supports independence aligns somewhat with the industry's goal to create functional and adaptable solutions, though market demands can introduce distance.	**Moderate Distance:** Municipalities focus on cost-effective, broad solutions, while the industry must balance this with having unique technology.
Affective	**High Distance:** The value users place on human connection, social stigma and community support contrasts with municipalities' operational focus, highlighting significant differences in emotional engagement.	**Moderate Proximity:** Initial motivations of the industry (especially startups) resonate with users' affective values, but as companies scale, a shift towards balancing economic with user-centric goals can widen the distance.	**Moderate Distance:** Municipalities' pragmatic approach to technology implementation shows less alignment with the industry's affective drive originating from personal experiences and a desire to address specific health challenges.

4 Discussion: bridging distance and proximity in values

Aligning technological advancements with societal needs is crucial in Responsible Innovation (RI), especially when considering the diverse values of users, municipalities, and the industry within digital health technology. The interaction among these groups reveals a complex mix of shared and conflicting values, emphasizing technology's role in improving healthcare and life quality.

A key insight from this study is the contextual and evolving nature of values, showing that stakeholder societal role may affect their value landscape. However, understanding the unique differences in stakeholder value might be an important first step in an inclusion process for responsible innovation. Therefore, it might be useful to perform an analysis among stakeholder to understand their viewpoints.

Another important insight from this study is that the differences in value across the stakeholder landscape may challenge the idea of inclusion and engagement around innovation development.

To further help stakeholders navigate the value landscape, effective communication strategies can navigate value proximity and common ground. For instance, in the Cognitive Dimension, stakeholders might unite over goals like enhancing healthcare technology's accessibility and effectiveness. In the Conative Dimension, balancing personal autonomy with municipal efficiency or finding the middle ground between industry innovation and broad, cost-effective solutions becomes essential. Meanwhile, the Affective Dimension suggests focusing on the emotional aspects users associate with technology, such as relationship improvement and stigma avoidance, and integrating these into technological strategies to bridge differences between economic and user-centric goals.

As such, one might be developing a shared language of Responsible Innovation (RI), weaving inclusivity, accessibility, and sustainability into the fabric of technology innovation. From there on, organized dialogues, educational workshops, and participatory design practices, stakeholders can explore and integrate the principles of RI. Finally, by viewing technology as an expansion of human work rather than a replacement, this approach may champion a forward-looking perspective on innovation.

In summary, adopting a shared language of responsible innovation, enriched by the diverse values of stakeholders, is meaningful for reconciling priorities within the healthcare technology sector. This approach facilitates dialogue, consensus-building, and collective action, ensuring that innovations meet immediate needs while adhering to broader principles of responsibility, equity, and sustainability.

5 Conclusion

In summary, this study navigates the crucial confluence of Responsible Innovation (RI), user inclusion, and the exploration of value landscapes in advancing healthcare technology. It emphasizes the critical role of understanding and integrating the multi-faceted values of users, municipalities, and the industry to drive technology forward in a manner that meets societal needs and upholds ethical standards. By manoeuvring through these complex value landscapes via collaborative efforts, stakeholders are better positioned to develop technologies that not only solve healthcare challenges but also align with broader societal and ethical expectations.

The concept of a shared RI language, highlighting inclusivity, accessibility, and sustainability, stands out as a key strategy for bridging value differences and enhancing mutual understanding. This participatory approach to innovation ensures a broad range of stakeholder perspectives are considered, fostering the development of healthcare technologies that are both impactful and ethically sound.

However, the study also acknowledges limitations, including the challenge of capturing the dynamic value landscape and the specificity of its Norwegian healthcare context, which may not fully translate to other settings. Also, values are contextual and evolving and perhaps difficult to truly capture. The rapid evolution of healthcare technologies further underscores the need for ongoing, adaptable research to keep pace with emerging ethical considerations and stakeholder expectations.

Future research focusing on effective communication strategies among stakeholders with diverse values is identified as essential for further advancing RI in healthcare technology. By developing methods that facilitate consensus-building and inclusive dialogue, future efforts can significantly contribute to creating healthcare innovations that are not only technologically advanced but also deeply resonant with societal values, thus embodying the essence of responsible innovation in a truly collaborative, inclusive manner.

References

Barben D, Fisher E, Selin C, and Guston D.H. (2008). Anticipatory governance of nanotechnology: Foresight, engagement, and integration. In *The Handbook of Science and Technology Studies*, 3rd Edition, ed. Hackett EJ, Amsterdamska O, Lynch M, and Wajcman J, 979–1000. Cambridge, MA: The MIT Press.

Bradonjic, P., Franke, N., & Lüthje, C. (2019). Decision-makers' underestimation of user innovation. *Research Policy, 48*(6), 1354-1361. https://doi.org/10.1016/j.respol.2019.01.020

Blok, V., & Lemmens, P. (2015). The emerging concept of responsible innovation: Three reasons why it is questionable and calls for a radical transformation of the concept of innovation. In B.-J. Koops, I. Oosterlaken, H. Romijn, T. Swierstra, & J. van den Hoven (Eds.), Responsible innovation 2: Concepts, approaches, and applications (pp. 19–35). Springer International Publishing. https://doi.org/10.1007/978-3-319-17308-5_2

Callon, M., Lascoumes, P., & Barthe, Y. (2009). Acting in an uncertain world: An essay on technical democracy (G. Burchell, Trans.). The MIT Press.

Chesbrough, H. (2003). *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business School Press.

Collingridge, D. (1980). *The social control of technology*. St. Martin's Press; F. Pinter.

Erikson, E. H. (1959). "Identity and the Life Cycle." W. W. Norton & Company.

Finlay, L. (2008). Reflecting on reflective practice. PBPL Paper 52, The Open University.

Finlay, L. (2011). Phenomenology for Therapists: Researching the Lived World. Oxford: Wiley-Blackwell.

Freeman, R.E. (1984). Strategic Management: A Stakeholder Approach. Boston: Pitman.

Goodin, R.E., & Dryzek, J. (2006). Deliberative Impacts: The Macro-Political Uptake of Mini-Publics. *Politics & Society*, *34*, 219–244.

Grinbaum, A. and Groves, C. (2013). What is 'responsible' about responsible innovation? Understanding the ethical issues. In R. Owen, J. Bessant and M. Heintz (eds), Responsible Innovation (pp. 119–142). Chichester: John Wiley & Sons

Iakovleva, T., Bessant, J., Oftedal, E.M., & da Silva, L.M. (2021). Innovating Responsibly – Challenges and Future Research Agendas. *Sustainability*, *13*, 3215.

Jarmai, K. (2020). Responsible innovation: Business opportunities and strategies for implementation. Dordrecht: Springer Netherlands.

Helgesen, M. K. & Herlofson, K. (2017). *Kommunenes planlegging og tiltak for en aldrende befolkning.* (NIBR-rapport 2017:16). By- og regionforskningsinstituttet NIBR.

Kaiser, M.C. (2022). 5. Taking value-landscapes seriously. Transforming food systems: ethics, innovation and responsibility.

Koops, B., Oosterlaken, I., Romijn, H.H., Swierstra, T., & Hoven, J.V. (2015). Responsible Innovation 2: Concepts, Approaches, and Applications.

Mierlo, B.V., Beers, P.J., & Hoes, A. (2020). Inclusion in responsible innovation: revisiting the desirability of opening up. *Journal of Responsible Innovation*, *7*, 361-383.

Oftedal, E.M., Foss, L., & Iakovleva, T. (2019). Responsible for Responsibility? A Study of Digital E-health Startups. *Sustainability*.

Owen, R., & Pansera, M. (2019). Responsible innovation and responsible research and innovation. In D. Simon, S. Kuhlmann, J. Stamm, & W. Canzler (Eds.), *Handbook on Science and Public Policy* (Chapter 10). Edward Elgar Publishing. DOI: 10.4337/9781784715946.00010 Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. Research.

Remenyi, D., Williams, B., Money, A., & Swartz, E.M. (1998). Doing Research in Business and Management: An Introduction to Process and Method.

Rip, A. (2018). The past and future of RRI. In *Futures of Science and Technology in Society* (pp. 115–133). Part of the book series: Technikzukünfte, Wissenschaft und Gesellschaft / Futures of Technology, Science and Society (TEWG).

Smith, J.A. (2011). Evaluating the contribution of interpretative phenomenological analysis. Health Psychology Review, 5(1), 9-27.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. Research Policy, 42(9), 1568-1580.

Stirling, A. (2005). Opening Up Or Closing Down? Analysis, Participation And Power In The Social Appraisal Of Technology.van Mierlo et al. 2012).

Schwartz, S. H. (1992). "Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries." In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 25, pp. 1–65). Academic Press.

Thapa, R., Iakovleva, T., & Foss, L. (2019). Responsible research and innovation: a systematic review of the literature. *European Planning Studies*, *27*, 2470–2490.

Thomas, E., Skeie, K. B., & Huang, H. (2022). Inclusion of elderly users via virtual spaces in the early stages of the innovation process. R&D Management. 10.1111/radm.12551

Von Hippel, E. (2005). Democratizing Innovation. Cambridge, MA: MIT Press.

Von Schomberg, René (2019).'Why Responsible Innovation' in: The International Handbook on Responsible Innovation. A Global Resource. Von Schomberg, René and J. Hankins (Eds.). Cheltenham and Northampton: Edward Elgar Publishing. pp. 12–32

Voss, J., Bauknecht, D., & Kemp, R. (2006). Reflexive governance for sustainable development.

Welman, J.C., & Kruger, S. (2002). Research Methodology for the Business and Administrative Sciences.

Yin, R.K. (1984). Case Study Research: Design and Methods.

Zejnilovic, L., Oliveira, P., & Canhão, H. (2016). Innovations by and for Patients, and Their Place in the Future Health Care System.

Zhang, Y., & Wildemuth, B.M. (2005). Qualitative Analysis of Content

Kristian Alm and Heidi Leclaire Karlsen

Chapter 9
Has law and social science trivialised the concept and practice of whistleblowing in Norway 2007–2023?

This chapter examines the modern evolution of whistleblowing legislation and practice in Norway and its implications for Corporate Social Responsibility (CSR). Our discourse analysis spans legal provisions, scholarly research, and national surveys and reports on whistleblowing. A finding is that the whistleblowing institution is tied to a power interest in encouraging employees to speak out, predominantly to prevent severe public interest damage to the population and economy. Yet, we have found that the law's broad whistleblowing concept has led to the predominance of commonplace personal-related cases within organisations' whistleblowing channels. Our analysis further aimed to discern whether this prevailing broad understanding of whistleblowing serves the public interest or rather prioritises individual grievances at its expense.

Both organisations' whistleblowing routines and research surveys exploring the landscape of whistleblowing tend to be based on or are bound by the broad whistleblowing concept in the legislation. Consequently, we have limited tools for assessing whether whistleblowing serves the public interest and is an efficient mechanism for detecting when CSR fails. We examine recent research indicating an increase in retaliation against whistleblowers. However, the broad whistleblowing concept makes it difficult to discern whether this increase targets public interest reporting, individual grievance reporting, or both. Consequently, our ability to assess the effectiveness of whistleblowing to restore CSR is impaired. The Norwegian Supreme Court recently interpreted the whistleblowing provisions, solidifying a low threshold for what constitutes whistleblowing. A dissenting minority highlighted the provisions' intent: whistleblowing in the public interest.

Against this background, we conclude by providing suggestions for improving the whistleblowing institute and its potential as a CSR mechanism.

1 Introduction

Whistleblowing legislation has been in force in Norway since 2007, institutionalising employees' right to report objectionable conditions in working life. Prior to this, there are several examples of whistleblowing in practice, early research on whistleblowing, and a political and legal discourse advocating for the enactment of whistleblowing legislation. After its enactment, the whistleblowing provisions within the Working Environment Act have gone through several amendments, with the most recent becoming effective in

2020. The Norwegian Supreme Court recently (21 December 2023) interpreted the current provisions, establishing a precedent that solidifies a considerably low threshold for what should be considered objectionable conditions and whistleblowing.

Our initial hypothesis posited that the concept and practice of whistleblowing in Norway from 2007 to 2023 has been established and further developed in a trivialised manner. This assumption stemmed from the broadly defined concept of whistleblowing in the law, official reports, and research and organisations' documents, as well as from insights from journalists and lawyers specialising in labour law. We have investigated this hypothesis through a Michel Foucault–inspired analysis of discourse in political and legislative documents, as well as in works in social science, with a focus on how the tradition of critical CSR might be influenced by this development.[1]

We show that the emergence and institutionalisation of whistleblowing is partly tied to a power interest in protecting the population and the economy. We analyse this inspired by Foucault's works on biopolitics and population politics, underscoring the co-dependency of the state of the economy and the quality of the population. In a technologically advanced society, insiders are increasingly crucial for uncovering severe irregularities or harm to public interest. This aligns well with prominent early definitions of whistleblowing. One such definition, articulated by American lawyer and activist Ralph Nader, describes whistleblowing as: "An act of a man or a woman who, believing that the public interest overrides the interest of the organization he serves, blows the whistle that the organization is involved in corrupt, illegal, fraudulent or harmful activity" (Nader, 1972, p. 1). However, we are concerned about the extent to which whistleblowing helps the common good in this sense.

By the term "trivialised" in this context, we mean that whistleblowing expands considerably, becoming closely associated with rather commonplace cases of individual and intra-organisational interest. We have found that there is a tendency both in legal regulations and in social science to consider cases of primarily personal interest and cases of public interest as converging towards a common procedural standard.[2] This trivialisation does neither imply that there have been no instances of reporting severely objectionable conditions, nor does it suggest that the concept and provisions inherently preclude whistleblowing on matters of significant public interest. However, we assume that the established concept and provisions have led to an extensive utilisation of whistleblowing channels for cases pertaining to minor conditions and/or

---

1 This is a new perspective on the institution of whistleblowing in Norway; our bibliography on whistleblowing research in Norway 1970–2023, prepared as part of the AFINO project, does not contain any publications of this type (Alm and Karlsen, 2024).
2 Even if we here distinguish between differences of importance, we later underscore that there is a continuum between whistleblowing cases of interest primarily to the individual and cases of public interest.

conditions which are of concern primarily or exclusively to the whistleblower in question. While we acknowledge that these cases may hold significance for the parties involved (which we will come back to), we contend that a trivialised whistleblowing concept and practice, as we interpret it, denotes a departure from prominent definitions of whistleblowing like Nader's.

The scope of cases covered by Norwegian whistleblowing provisions and organisations' guidelines is broad, ranging from minor personal cases to severe threats or damages to life, health, and economic affairs. Commonplace cases, such as those concerning unsatisfactory working environments, are likely to outnumber those that align with the whistleblowing concept as framed by Nader. We assume that a function of this tendency is that whistleblowing is at risk of being closely associated with personal cases relevant to the individual and, to varying degrees, their organisation, but less frequently of public interest. We understand the Supreme Court's expansive interpretation of the provisions of whistleblowing as the last significant example of this trivialising development with the risk of contributing to this inflation.

Against this background, we discuss if the prevailing broad understanding of whistleblowing impedes the strength of whistleblowing as a critical CSR instrument. We understand critical CSR (Midttun, 2013) as the company's integration of social and environmental considerations into its daily operations on a voluntary basis.

We further understand CSR as a business model focused on corporations' critical self-regulation in terms of social accountability, which can take various forms. Essentially, businesses engage in actions which benefit society, partly through self-criticism, through which they are also likely to enhance their reputation and, ultimately, their profits. Turning to whistleblowing procedures, corporations may encourage their employees to report in accordance with the organisation's whistleblowing routines if they, for instance, discover abuse/waste of public money transfers or social/psychological predicaments due to their organisation's activities. Such employer initiatives to facilitate whistleblowing constitute, however, an approach to stop or prevent wrongdoing, not to promote positive actions. Yet, if we broaden our understanding of CSR to encompass practices which make businesses more conscious of their societal impact, it might be reasonable to view a corporation's whistleblowing procedures as part of its CSR activities. One could argue that, ideally, CSR procedures should prevent the circumstances necessitating whistleblowing. A more pragmatic approach, however, acknowledges that, despite the best preventative measures, issues of concern in the organisation's undertakings may arise in the real world.

However, this broader interpretation of CSR raises a critical issue regarding what tangible or theoretical benefits are gained by associating whistleblowing with CSR, given such an expansive definition of the CSR concept. Several studies take this point of departure, though, focusing on whistleblowing as an instrument for CSR to promote disclosure of severe violations of norms, which is in the public interest to be informed about. Examples are financial fraud (Brink et. al., 2018), mistakes in products in the pharmaceutical industry (Stankiewicz-Mróz, 2015), positive attitudes to CSR as a driver

for employees' whistleblowing intentions as to fraud (Yang and Mahdy, 2022), and CSR as a driver for auditors to use whistleblowing channels (Sallaberry et.al., 2020). We have not found any studies discussing how whistleblowing regulations, despite perhaps the best intentions, may impede the potential of whistleblowing to detect when CSR fails.

Our analysis shows that whistleblowing *in the sense of speaking out about concerns of public interest* is insufficiently visible. Consequently, we problematise whether whistleblowing currently is as strong an instrument for restoring CSR as its potential. This work will suggest strategies to mitigate these issues.

2 Methodology

We conduct a discourse analysis of the whistleblowing conceptualisation in the legislation, selected representative types of social science research, and official reports. The approach to the analysis of these documents is inspired by Foucault's methodology for discourse analysis in his *The Archaeology of knowledge* (1972 [1969]), whose object is the identification of *statements* (*énoncés*). Through the collection and observations of concrete ways of speaking – *enunciations* (*énonciations*) – in a delimited set of documents, by the researcher, within the assembled *archive*, the researcher identifies *statements*. These are regularities or patterns the researcher observes across the collected enunciations, as well as their functions or effects in relation to other identified statements within the discourse (Karlsen, 2023). However, Foucault's methodology deals with the whole "mass of said things" (Foucault, 1969). This encompasses a broad spectrum, ranging, for instance, from the most formal documents to the most informal, everyday expressions. To analyse this broad spectrum is largely beyond the scope of this work. Yet, sections 3 and 4 conduct a discourse analysis of the mentioned documents and specify discursive statements across the whistleblowing articulations in this corpus.

Section 3 analyses the political legislative discourse. We have included every formulation of the legal provisions in the analysis, as well as government-initiated reports (NOUs) and government propositions which discuss whistleblowing.

Our selection of documents for analysis in section 4 is based on our bibliography on whistleblowing research in Norway from 1993 to 2023 (which will be published in 2024). The bibliography covers 250 entries of works on whistleblowing.[3] It discloses that a tradition of whistleblowing research in Norway did not start when we assumed,

---

3 Our bibliography discloses that a tradition of whistleblowing research in Norway did not start when we assumed, which is in the 1970s, after the students' uprising in Paris in 1968, but rather much later, around the year 2000.

which is in the 1970s, after the students' uprising in Paris in 1968, but rather much later, around the year 2000.

We have selected publications for analysis that, according to the bibliography, are the most dominating research on whistleblowing in the period: organisational sociology that has a quantitative design.

Our analysis of these strands of whistleblowing research is delimited in time to the period between 2005/2007 (when the whistleblowing legislation was drafted and then adopted by the parliament) and 2023. The whistleblowing legislation gave impetus to a significant number of research publications. In other words, 2005/2007 and the following years represent an intensification of Norwegian whistleblowing research, which makes it natural to start our analysis then.

We select three publication periods through these 15 years, the beginning, the middle, and the end, assuming this is sufficiently representative to grasp a possible development as to the topic of our hypothesis. Concerning the most influential type of whistleblowing research, organisational sociology, we select the researchers who, according to the bibliography, published most frequently from 2007 to 2023 and are most frequently cited. These include Sissel Trygstad, Fafo (27), and Marit Skivenes, University of Bergen (13).

3 Analysis of the history of the Norwegian whistleblowing legislation

In 1999, an official report on freedom of speech (NOU 1999:27) recommended that whistleblower provisions come into place in Norway. About 7 years later, the enactment of a whistleblowing legislation took place. Multiple articles and comments discussed and criticised the enacted provisions, and the parliament has enacted revisions of the provisions up until recently. This section briefly analyses this history with an emphasis on the whistleblowing concept it articulates and the implications of whistleblowing practice it entails. Legislation on whistleblowing was first put into place in 2007 and revised provisions were enacted in 2017 and 2020.

As we shall see in this section, whistleblowing legislation was put in place to achieve two main objectives. Strengthening employees' freedom of speech is one of them, partly – in a deontological ethical perspective – as an action with value in itself, partly – in a consequentialist perspective – as a means to ultimately achieve a society with less economic abuse, life-threatening accidents, health damage, and intolerable environmental destruction. The latter – the means – is the predominant issue the whistleblowing legislation was designed to deal with. To achieve both these objectives, however, an initially debated concern is to demarcate an employee's right to freedom of speech in light of the duty to protect the public interest and prevent harm, from the duty of loyalty to the organisation. The rationale for this debate is that this

boundary is unclear to employees and that Norwegian employees do not blow the whistle as much as justified.

Sub-sections 3.1 to 3.4 present enunciations of the whistleblowing concept in the selected documents and discuss their reception and context, whereas sub-section 3.1.5 concludes the section by specifying the discursive statements we have identified based on the patterns across these enunciations.

3.1 Legal whistleblowing discourse prior to the enactment of the whistleblowing provisions in 2007

In the preparatory works leading up to the enactment of the whistleblowing legislation in Norway in 2007, a recurrent topic of discussion was the delineation between an employee's loyalty obligation to their employer and their duty or right to report objectionable conditions. The notion of a right to report, albeit with no whistleblowing legislation yet, must be seen in accordance with the general right to freedom of expression in the Norwegian constitution. Additionally, there was debate regarding the problem of retaliations against employees who exercised their freedom of speech to expose objectionable conditions in their organisation. Legal experts also scrutinised the extent to which the early draft provisions would practically alter the existing status quo.

In 1999, an official report on freedom of expression declared that whistleblowing should be understood as an employee's right to freedom of expression. It recommends a clarification of this right by regulating it in law. The report depicts external whistleblowing as a form of disloyalty to the employer, which under certain conditions is legal and desirable. There is, however, an internal "negotiation" of the concept of loyalty in this report: public criticism from employees may also *benefit* workplaces and thus not be considered disloyalty:

> We would generally warn against that insinuations of disloyalty are played out as the first response to public criticism from employees. One should rather take the employee's initiative as a positive challenge and opportunity for further public exchange of words – to the benefit of both the general public and the workplace (NOU 1999:27, p.173).[4]

This report emphasises that loyalty to the public takes precedence over "apparent" loyalty to the organisation in certain circumstances:

> Under specific conditions, apparent disloyalty can be allowed and desirable. We refer here to situations where the employee alerts the public that their workplace is involved in corrupt, illegal,

---

4 All translations of Norwegian documents in section 3 were performed by us, with assistance from large language models (ChatGPT 4), unless otherwise specified.

immoral, or other harmful activities. Such reporting is referred to in English as 'whistleblowing' (NOU 1999:27, p. 175).

We see that the report articulates whistleblowing as usage of employees' freedom of speech to inform the public about their organisation being involved in, for instance, economic crime and other damage to public interest. Thus, this whistleblowing enunciation emphasises employees' freedom of speech as a right of value in itself while underscoring the use of it as an action in public interest. The latter concerns responsibility and indirectly CSR. Whistleblowing is articulated as a means to foster the cessation of harmful activities by organisations. When must loyalty – or perhaps more accurately, obedience and silence – yield to the right (or possibly the obligation) to speak out about reprehensible conditions? A space opens for organisations' critical self-regulation.

An official report on the Norwegian working life published in February 2004 included an analysis of whether there was a need for whistleblower legislation to clarify the boundary between employees' freedom of speech and the obligation of loyalty to the employer. This committee concluded that there was no need to legislate the delimitation of employees' right to notify (NOU 2004:5, p. 32). The same report recommended, however, to legislate prohibition of retaliations against loyal whistleblowing. It proposed the following formulation: "It is prohibited to use retaliation against employees who inform in a loyal manner about objectionable conditions in the organisation" (NOU 2004:5, p. 452). *Loyal whistleblowing* implies, among other elements, that internal procedures have been tried before going public, according to the report (NOU 2004:5, p. 445). In terms of balancing the interest of employees utilising their freedom of speech and the consideration for the employer (duty of loyalty), the committee behind the report emphasises the utility of whistleblowing:

> The consideration for the public generally argues in favour of extensive freedom of speech for employees. Serious objectionable conditions that pose a threat to life or health, legal violations, corruption, etc., are examples of matters that it would be justified to inform the public about (NOU 2004:5, p. 446).

> The committee places great emphasis on the societal benefits of reporting objectionable conditions within an organisation [. . .] It is also noteworthy that employees participate in public debate about their field of work. Therefore, it is important to facilitate the uncovering of objectionable conditions. At the same time, it is crucial to protect employers and the potential harmful effects that negative publicity can have on a business. In particular, employers should be protected against the release of unfounded claims and have the opportunity to address the issues internally first. (NOU 2004:5, p. 449)

In sum, whistleblowing enunciations in this report show a normative concern for facilitating whistleblowing, by suggesting prohibiting retaliations against whistleblowers. Simultaneously, they express the requirement of loyalty towards the employer, which, according to this report, already is well demarcated from employees' freedom of expression in existing provisions. By suggesting the concept of *loyal* whistleblowing, the committee calls out for guidelines in the organisations to assess the loyalty of notifica-

tions and prepares the ground for the later distinction between internal and external whistleblowing.

In line with the recommendations in this report, the government put forward a proposition for whistleblowing legislation, as a part of the revisions of the Working Environment Act (Ot.prp. nr. 49 [2004–2005]). Yet, the parliament adopted a different formulation of the provisions (see next paragraph). The revised Working Environment Act was enacted from the start of 2006. The provisions on whistleblowing were, however, postponed, due to the government's expressed need for specific legislative proposals that could further strengthen employees' real freedom of expression. The Ministry of Labour and Social Affairs was set to carry out this investigation and present it to the parliament during the spring of 2006.

The provisions from 2005, which were adopted by the parliament but whose enactment was postponed, read as follows:

> Prohibition of retaliation as a result of notification. Employees have the right to notify the public about objectionable conditions in their organisation as far as this is not contrary to other laws. Retaliation against an employee who gives notice pursuant to the first paragraph is prohibited (Ot.prp. nr. 49 [2004–2005]).

Lawyer Erik C. Aagaard (2005) criticised this draft for its failure to clarify the demarcation between freedom of speech and the obligation of loyalty. He further argued that weaknesses of the draft were its lack of a specification of "objectionable conditions" as well as its restriction to public reporting. Aagaard's criticism of the draft's failure to clarify the boundary between an employee's freedom of speech and their loyalty obligation builds upon the study of legal scholar Kyrre Eggen (2004). Eggen conducted a thorough analysis of this legal landscape prior to the 2004 draft provision. He claimed that employees already in reality had a more extensive right to exercise their freedom of speech, as per paragraph 100 on freedom of speech in the Norwegian constitution, than what was commonly practiced. According to Eggen, a probable explanation for this limited exercise of employees' freedom of speech right lay in an uncertainty regarding the boundary between the obligation of loyalty and freedom of speech. Consequently, he advocated for a whistleblower legislation specifically aimed at clarifying this distinction for employees. Eggen's work primarily consists of an analysis of the current legal situation and proposals for how it can be clarified. Yet, in the broader discourse it is part of, we may see this advocation as an articulation of whistleblowing as a solution to a deficiency of responsibility: employees need a clarification (in terms of whistleblowing legislation) to more frequently report objectionable conditions in their organisations, which entails being responsible for a common good. However, the critical point is whether the conceptualisation of whistleblowing in the draft provisions was too vague to be seen as a solution.

The discourse in the documents analysed in this sub-section establishes that employees should utilise their freedom of speech more, to the benefit of society. There are certain tensions and disagreements within this discourse related to how this can

best be achieved. The discourse includes disagreements regarding whether the existing provisions provide a sufficiently clear distinction between the notion of employees' freedom of speech and their duty of loyalty towards the employer. There is also an ambiguity as to what counts as notifiable conditions, to whom they should be reported, and if, or how, this should be specified in the law. Regarding the question of *to whom* objectionable conditions should be reported, we saw that the proposed provisions of 2005 emphasised *public* reporting. *Public* reporting would have entailed that the objectionable conditions would likely be of *public interest*. We also saw that public reporting and public interest were notions underscored in the freedom of speech report of 1999. Importantly, in this early whistleblowing discourse, also the frequent articulations of the employees' freedom of speech entail a discursive emphasis on public reporting and interest; the question of freedom of speech versus loyalty to the organisation is less relevant in relation to "smaller" cases/personal cases reported through internal whistleblowing channels.

The notions of public reporting and public interest will, as we shall see, be downplayed in the later legislation.

3.2 The whistleblower legislation of 2006/2007

The enacted whistleblower legislation in the Working Environment Act from the first of January 2007 reads as follows:

> Notification of objectionable conditions in the organisation. (1) Employees have the right to report objectionable conditions in their organisation. (2) The employee's procedure for giving notice must be justifiable. In any case, the employee has the right to notify in accordance with the obligation to notify or the organisation's whistleblowing routines. The same applies to notification to supervisory authorities or other public authorities. (3) The employer has the responsibility for proving that a notification eventually has occurred in violation of this provision (that it has not been justifiable [*forsvarlig*])"[5]

Additionally, the law specifies protection against retaliation, employers' obligation to facilitate whistleblowing, and *employers' responsibility* for proving that discrimination has *not* taken place if an employee has reported discrimination.

There is neither a definition in the law of "objectionable conditions" nor a specification of distinct degrees of severeness of objectionable conditions. As surveys had indicated that employees dreaded or were uncertain regarding exercising their right to speak out about objectionable conditions in their organisation, one intention behind these provisions was to lower the threshold for such reporting. Specifying the

---

5 The complete provisions, as originally enacted, can be consulted here (in Norwegian): https://lov data.no/dokument/LTI/lov/2006-12-01-64. The parliament adopted directly the provisions recommended in the proposition the Ministry of Labour and Social Affairs presented in June 2006.

concept of "objectionable conditions" could conflict with such an intention, according to the government's proposition to the parliament in 2006 (Ot.prp. nr. 84 [2005–2006]). However, the government's proposition includes a section which nevertheless attempts to define the scope of this concept. Building on the specification of "objectionable conditions" in the government's proposition to the parliament back in 2004, the proposition of 2006, which was enacted, holds that objectionable conditions should be understood as criminal and illegal activities and "probably violations of other ethical norms" (Ot.prp. nr. 84 [2005–2006], p. 23). In other words, the concept of objectionable conditions remains vague; the blurring of a fundamental distinction continues to play out.

3.3 The reception and evaluations of the whistleblowing legislation of 2006/2007

Subsequent to the enactment of the whistleblowing legislation, a great number of studies, including legal scholarly articles and trade journal articles, as well as master theses at Norwegian universities, analysed the notion "justifiable" (*forsvarlig*) in the law formulation (see section 4). This formulation emphasises the manner through which the notification should take place. According to the committee behind the law proposition, this formulation provided a tool for assessing whether a notification respected the notion of loyalty to the employer.

That the notification must be justifiable entails that it must have a factual basis (*saklig grunnlag*). Furthermore, internal whistleblowing should be attempted as the first step. Also, the reported objectionable conditions should be of public interest if the employee blows the whistle externally (Ot.prp. nr. 84 [2005–2006], 2006). The justifiability criterion received, however, criticism for being unclear (Flaatten, 2007; Støver, 2009), problematic (Eggen, 2008), and not suitable to take care of the considerations which are the basis for the whistleblowing rules (Olsen, 2007). More specifically, the notion of "factual basis" (*saklig grunnlag*) posed a problem. The whistleblower must have acted in good faith, convinced that the notification is based on factual conditions. According to Steen (2012), case law has shown that this notion leads to excessive focus on the whistleblower's intention at the expense of the objectionable conditions they have reported. Thus, this contributes to drawing the attention away from clarifying what "objectionable conditions" should be understood as. Also, the already mentioned study by Eggen (2004) warned – already before the whistleblowing provisions came into place – against significant emphasis on the whistleblower's intention and suggested instead increased emphasis on the public interest of the reported objectionable conditions. Building on Steen and Eggen's observations of the effects of the justifiability criterion, we argue that this whistleblowing enunciation – likely spurred by the attempt to clarify the distinction between loyalty to the employer and protection of the public interest – entails a slide in focus. This is a slide in focus from whistleblowing as acting in the public interest to-

wards internal procedures within the organisation. Routines in the organisation for assessing a whistleblower's credibility and time and resources for processing the notification's basis become more predominant in the discourse of whistleblowing, at the expense of its public interest dimension.

As we saw in the last section, the concept of objectionable conditions is vague in the provisions. What counts as ethical norms, which ethical norms are relevant in this context, agreed to by whom, and what counts as violations of these norms? Furthermore, these violations "probably" add up to objectionable conditions which one should have the right to report. However, it is an easy exercise to criticise this notion for not being concrete and clear; formulations of the provisions can't be too specific as this would lead to problems of exhaustiveness. Also, they are supposed to have relevance over a longer period of time. Yet, it seems to us that the concept could have been further specified without running into these potential predicaments. It is understandable though that the wish to motivate a greater number of employees to come forward when observing objectionable conditions results in provisions which do not delimit what constitutes objectionable conditions. This seems to be the main explanation for the vagueness of the concept in the provisions, as well as for the scarce description of what the obligation of loyalty to the employer entails. Nevertheless, although this reasoning is at first glance understandable, the result instead has been that whistleblowing has drifted from public to individual interests. One uses the channels for internal whistleblowing to a large extent to report conditions one perceives as objectionable in accordance with one's personal situation and values. These are conditions of which oneself – and often only oneself – is victim. In many cases the legal and ethical landscape of freedom of speech and obligation of loyalty is not relevant for these cases. Reporting a colleague for being difficult to collaborate with could be an example. Another example could be reporting one's superior for not showing understanding regarding one's life situation. We acknowledge that such conditions may be painful for the person reporting them. Such cases are, however, likely to be best handled by routines for individual working environment problems. In contrast, the legal and ethical landscape of freedom of speech and obligation of loyalty is relevant when it comes to systemic working environments problems and other violations of ethical norms of broad acceptance. To navigate in such a landscape, it is possible that employees would benefit more from clear guidance rather than vaguely defined concepts.

Commissioned by the Ministry of Labour and Social Affairs, the research foundation Fafo carried out an evaluation of the whistleblowing provisions (Fafo, 2014). They found that almost 50 percent of Norwegian employees do not report objectionable conditions. Also, whistleblowers experience retaliations (the percentage varies among studies, with about 12 as the average percentage), and there are varieties regarding employees' awareness of the whistleblowing provisions. Furthermore, whistleblowing routines facilitate whistleblowing and reduce the risk of retaliations. Due to the vague concept of "objectionable conditions", a problem, however, is that these numbers do

not necessarily inform us as to what extent they apply to personal cases, cases of obvious public interest, or others. The report shows multiple operationalisations of the concept of "objectionable conditions" across studies. This is a significant problem, given that the predominant interest in more whistleblowing is to prevent or stop damage to the economy and population. Yet, this report recommended, based on international research, that the concept of "wrongdoing" or "objectionable conditions" in the provisions remained wide (Fafo, 2014, pp.18-20; 199; 201).

Based on their evaluations, the Ministry of Labour and Social Affairs submitted proposals for revisions of the whistleblowing provisions (Prop. 72 L, [2016–2017]). Among the proposals was the duty for organisations with five or more employees to have whistleblowing routines. The proposal did not suggest specifications of the scope of objectionable conditions but recommended the organisations to operationalise the concept in their whistleblowing routines and possibly provide relevant examples of objectionable conditions. The parliament adopted the proposals in line with the committee's recommendation (Inst. 303 L, [2016–2017]), with effect from July 1, 2017.

3.4 Recent political discourse and revised provisions

3.4.1 Official report on whistleblowing (NOU 2018:6)

In 2018, an official report on whistleblowing was published (NOU 2018:6), commissioned by the Ministry of Labour and Social Affairs. The committee takes a clear stance on whistleblowing; the report opens with their conclusion that whistleblowing is a value, not a problem; it is a socially profitable value to be promoted until we reach the "zero vision" of no objectionable conditions. They continue by highlighting that employees can more easily become aware of objectionable conditions than external parties. Furthermore, society, as well as the organisation in question, benefits from objectionable conditions not taking place, or from them ceasing. According to the report (NOU 2018:6, p.13), whistleblowing is socially profitable, a conclusion supported by the Ministry's commissioned research conducted by the economic consulting firm Oslo Economics.

The report addresses the concept of "objectionable conditions", showing that many find it unclear and challenging to interpret: how does the concept relate to distinctions between personal conflicts and work environment conflicts affecting multiple individuals, or professional disagreements and improprieties? This ambiguity makes it hard for both employers and employees to navigate. The committee acknowledges that it is aware "that there are many who perceive the use of the whistleblowing rules in connection with individual labour conflicts as one of the biggest challenges with the regulations, and which could contribute to casting whistleblowing in a negative light" (NOU 2018:6, pp. 155–156). As an attempt to clarify this landscape, the report suggests an inclusion in the provisions of a *non-exhaustive* list of examples of relevant objectionable con-

ditions, not an exhaustive list as is the case in the whistleblowing legislation of, for instance, in England, Ireland, the Netherlands, and New Zealand (NOU 2018:6, p.157). Their argument against an exhaustive list is that it could hinder a dynamic understanding of what is considered objectionable conditions, which may change over time. They also highlight that the preparatory work for the current rules does not provide such an exhaustive list (NOU 2018:6, p.157).

Overall, this report repeats and reinforces the discursive position that whistleblowing is in the public interest. Not least, the Ministry's commissioning of research on the societal benefits of whistleblowing shows that. Correspondingly, the report also repeats and reinforces the discursive position that the concept of objectionable conditions must be wide to facilitate public interest whistleblowing. Although acknowledging that many contemporary whistleblowing cases do not concern objectionable conditions within the public interest intent behind the legislation, the report is thus prudent in its suggestions to address this challenge.

Based on this report, the Ministry of Labour and Social Affairs submitted a proposal to the parliament in 2019 for revised whistleblowing provisions (Prop 74 L. [2018–2019]). The proposal aimed at specifying the scope of objectionable conditions, as well as the concepts of "justifiable procedure" ("forsvarlig fremgangsmåte") and "retaliations" ("gjengjeldelser") in line with recommendations in the report. One intention behind this proposition was to avoid purely personal matters from being covered by the whistleblowing provisions. Moreover, the report proposed regulations regarding the employer's obligations when receiving a notification. The revised provisions became effective from 1 January 2020 (Lovvedtak 70 [2018–2019]; Working Environment Act, 2005).[6]

Regarding the specification of objectionable conditions, referred to as "issues of concern" in the English translation of the legislation, the formulation reads as follows: "issues of concern include breaches of legislation, written ethical guidelines in the undertaking or ethical norms on which there is broad agreement in society", followed by a non-exhaustive list of examples. Two of these are: "the abuse of authority" and "an unsatisfactory working environment". Regarding the limitation of the legislation's application, the formulation reads as follows: "Questions raised that only relate to the employee's work situation shall not be considered whistleblowing [. . .] unless the matter also involves issues of concern as described [above]" (Working Environment Act, 2005). A pressing question is whether the provisions in practice work in accordance with the intentions, or whether employees still are inclined to use whistleblowing channels for personal matters, referring to "abuse of authority" and/or "unsatisfactory working environment".

Jon Hustad (2018), journalist, historian, and writer, reviewed the official report on whistleblowing, on which these last revisions are based. Most important in his view is the lack of nuances in the report's take on whistleblowing. "The committee transmits a

---

6 The Working Environment Act is available in English here, with the latest provisions from 2020.

viewpoint that whistleblowing always is a positive practise: it is defined as a value, not a problem" (Hustad, 2018, p. 308). However, many cases concern minor matters – such as errors or deficiencies, which can be reported directly to a manager or other authority in the high-trust Norwegian society – and many are cases where whistleblowers are wrong, and the employer is right, according to the reviewer. The latter could, for instance, be whistleblowers who themselves are the ones acting destructively to the working environment, according to the reviewer (Hustad, 2018). Hustad wrote a book in 2007 on a well-known Norwegian paediatrician and whistleblower. Subsequent to this publication, many people contacted him for help to bring public attention to their cases. He found that in most of these cases the employer was right. The alleged whistleblowers were themselves the problem without realising it (Hustad, 2018). Furthermore, Hustad highlights the varieties in the whistleblowing definitions and large estimates in the numbers in the research from Oslo Economics. Furthermore, Hustad highlights the variations in whistleblowing definitions and the wide range of estimates on the economic impact of whistleblowing provided by Oslo Economics, whose research forms the basis of the report. Oslo Economics aimed to determine the extent to which whistleblowing benefits society, concluding that the Norwegian society gains between half a billion and 12 billion NOK. He also refers to Oslo Economics' application of the concept of whistleblowing, which "means that whistleblowing encompasses everything from verbal alerts that can be resolved simply and non-bureaucratically by the immediate supervisor, to extensive whistleblowing cases that ultimately can end in media coverage and significant public attention" (NOU 2018: 6, p. 32, referred in Hustad, 2018, p. 309). He criticises the committee for their uncritical application of this research on the social benefits of whistleblowing. They thereby downplay the problem of the many commonplace cases which today are directed through the organisations' whistleblowing channels (Hustad, 2018).

3.4.2 Evaluations of the current whistleblowing institute in Norway

In the newspaper article "Et misforstått varslingsinstitutt?" (A misunderstood whistleblowing institute?), the lawyer Eldrid Huseby Gammelsrud (2022) refers to research showing that Norwegian employees most often blow the whistle about psychosocial conditions in the workplace, primarily related to bullying/harassment and destructive leadership. Such cases are often personal conflicts, according to Gammelsrud. She argues that the whistleblowing provisions are likely to complicate these processes and contribute to increased levels of conflict due to the thorough investigation mandated by the whistleblowing regulations. Often, organisations' insufficient legal expertise in managing whistleblowing cases necessitates the assistance of lawyers. The whistleblowing provisions are apt for handling cases of public interest, such as, for instance, economic and environmental crime, which require extensive investigation (Gammelsrud, 2022). The author mentions whistleblowing regulations in other nations, empha-

sising the commonly established precondition that whistleblowing should address matters of public interest, like the EU Whistleblowing Directive. She suggests that the aspect of public interest should be more emphasised in the Norwegian whistleblowing provision to enhance their functionality and scope.

A lawyer, with 20 years of experience in a Norwegian trade union, stated in an interview we conducted 29 February 2024[6] that they receive a large number of whistleblowing cases today. However, the majority of them are not cases in the sense intended by the Working Environment Act. Whistleblowing, as stated in the Working Environment Act, is intended to regulate objectionable conditions for which the organisation is responsible, such as the employer's violation of a norm within the organisation. It is this type of cases – which it is in the public interest to reveal and correct – of which the whistleblowing institution should ensure investigation. On the other hand, personal matters are the responsibility of the employer to address. These may be cases where an employee blows the whistle on another, for example, because they perceive their colleague's behaviour or actions as inappropriate. These cases mainly revolve around determining whether an individual has done something wrong, not whether the organisation is behaving improperly. Unless an individual with power in the company, such as a manager, engages in such objectionable behaviour to the degree that constitutes a systemic failure, the whistleblower provisions of the Working Environment Act are not suitable for handling them. The overwhelming majority of whistleblowing cases that this lawyer and their colleagues deal with should have been treated as personal matters but are erroneously funnelled into the whistleblowing system.

Regular personal matters are probably deliberately funnelled into the whistleblowing system because employers lack competence in handling whistleblowing cases, the lawyer explains. A crucial nuance is absent: the ability to differentiate between objectionable behaviour carried out by an individual and objectionable conditions within the organisation. The latter requires a different, more comprehensive investigation. Whistleblowing ("varsling") has become such a commonplace term. It is taken out of its legal context and has become part of ordinary, everyday vocabulary. That is where employers often face a comprehension barrier. Some sort of misguided understanding on the employer's part or fear of not taking matters seriously enough may be the reason why cases that should have been treated as personal matters are instead treated as whistleblowing cases. They then fall under the procedures for handling whistleblowing cases. Consequently, since the employer is obligated to take all reports seriously, the reports are not filtered.

The lawyer describes this situation as a catastrophe because the employer uses the whistleblowing institution against the individual. By this, they mean that extensive investigations are launched, as required by the whistleblower provisions in the Working Environment Act, and the employee faces problems as a result. A case that could have been resolved through consultation with their immediate supervisor unnecessarily grows larger, involving, for example, external lawyers. The employer can then point to

having taken action. In reality, however, the matter is entrusted to others, and it often ends with nothing actually happening. Instead, a report is usually received stating that nothing objectionable in the legal sense has taken place. However, the employer's handling of the case is not scrutinised. Yet, if it were to be revealed that the whistleblower committed an unlawful act, it could lead to termination. However, the whistleblowing regulations are not intended to be used for firing people, the lawyer points out.

The lawyer relates this trend to a more general observation, namely that the working environment is becoming tougher. It is getting tougher in the sense that there is less space for individuals who stand out. Much has become more streamlined. More and more things need to be documented and implemented in procedures. Not everyone is able to keep up with this development. As a result, we have a poorer working life and a poorer society, where we weed out those who are a little different. There may be a whistleblowing case in a situation where an employee, for example, does not document in their calendar what they are working on at any given time. Such cases are coming in as a package; that is the tendency, according to the lawyer. They find that it has become more difficult for labour lawyers on the employee side to achieve good solutions now than it was before. Out of fear for their own reputation, employers let this be handled according to the whistleblowing procedures. This means it can take weeks from the moment the employee finds out that someone has blown the whistle on them, until the content of the case is disclosed. This can be a significant burden for the employee. Furthermore, the case is likely to be outsourced to investigation firms, which are not necessarily bound by guidelines on how to proceed.

Another example of cases that are mistakenly treated as whistleblowing cases can be a situation where an employee has made a statement that has been perceived as having sexual undertones. Certainly, in some cases of this type, it may be that the person who is being reported has behaved inappropriately. Yet, in any case, these should not be treated as whistleblowing cases. Instead, these are cases that the employer should address, based on their management authority and personal responsibility, concludes the lawyer in the interview by 29 February 2024.

3.4.3 Norwegian Supreme Court's verdict on whistleblowing

The Norwegian Supreme Court's verdict on 21 December 2023 on whistleblowing establishes a precedent that solidifies a low threshold for what should be considered whistleblowing cases (The Supreme Court of Norway: HR-2023-2430-A, 2023). The majority of the Supreme Court of four justices concluded that an email sent from an employee representative to an HR manager met the requirements for whistleblowing. A minority of one justice took dissent, concluding that the email should be considered general criticism in working life, not whistleblowing[7].

The employee representative had assisted a colleague in a meeting where an HR manager had given the colleague an oral warning. Subsequently, the employee representative sent the mentioned email to a manager in the organisation. In harsh language, he criticised the HR manager's behaviour in the meeting and for having harassed this employee several times. This email resulted in a written warning and the employee representative was reassigned. Followingly, in his claim for redress and compensation, the critical concern was whether the email met the requirements for whistleblowing, to which the Court of Appeal's judgement had concluded negatively. This judgement was set aside, as the Supreme Court's majority found that the email met the requirements for whistleblowing.[7]

As the basis for their interpretation, the majority states:

> The law has no further definition of the term whistleblowing than speaking out about objectionable conditions in the workplace [. . .] The law does not stipulate any requirements regarding the form of the notification. Therefore, whistleblowing can be made both in writing and orally, and in any context [. . .] It is also not required that the employee's procedure when notifying is responsible ('forsvarlig'). (The Supreme Court of Norway: HR-2023-2430-A, 2023).

Moreover, in their interpretation, there is no criterion in the law that the conditions reported must be in the public interest to qualify as whistleblowing. They highlight that the law defines "objectionable conditions" in a wide sense. However, they point out that it is required that the objectionable condition violates a specified norm. In this concrete case, the relevant norm, according to the majority, is the company's work regulation stating that "everyone must act considerately and correctly towards managers and colleagues". The majority considered the email – specifically the alleged harassment of the colleague – to express more than the employee representative's disagreement with the organisation's warning to the colleague: "It described behaviour contrary to a rule in the undertaking's work regulations on considerate and correct conduct, and thus an issue of concern in the undertaking", according to the majority (see the English summary of the judgement, referred to in footnote 6). This last point is decisive for the majority's conclusion.

The minority of the Supreme Court, Cecilie Østensen Berglund, states: "as the first voter interprets the statutory provision, I believe that the distinction between general criticism and the whistleblowing rules are erased. I can't see that this has been the legislator's opinion [. . .] whistleblowing is a specific and specially protected form of expression, justified by society's need for certain types of information to be disclosed". Here, she refers to the latest official report on freedom of expression (NOU 2022:9), which warns against a chilling effect if criticism mistakenly is treated as whis-

---

7 The entire verdict, including the email in question, is available in Norwegian here: https://lovdata.no/dokument/HRSIV/avgjorelse/hr-2023-2430-a See the official English summary here: https://lovdata.no/dokument/HRENG/avgjorelse/hr-2023-2430-a-eng

tleblowing (see also section 3.5). Furthermore, Østensen Berglund refers to law documents establishing the importance of collaboration between employers and employee representatives for ensuring a well-functioning work environment. Frequent doubts about what constitutes a notification could weaken this cooperation. "In assessing whether a notification from an employee representative is formulated in such a way that the employer has reasonable grounds to perceive that the statement is a notification, the content, role expectation, and context will be significant". She describes sections of the email as "a pure scolding," reflecting the employee representative's reputation for harsh language, suggesting the email is an example general criticism, in line with the individual's tone. Furthermore, she highlights that it is reasonable to expect an employee representative to use the organisation's whistleblowing channels or label the criticism as whistleblowing, specifically to distinguish such notifications from ordinary interactions by virtue of being an employee representative.

Labour law attorneys have strong opinions about the judgment (see, for instance, the legal advice website Juridisk ABC, 2024). A consequence of the judgment is that the concept of whistleblowing is broadened. It takes little for an expression to be considered whistleblowing (Hagen, 2024; Aastveit, 2024). This blurs the distinction between general criticism and whistleblowing (Codex Advokat, 2024: Juridisk ABC, 2024). More internal communications within organisations will be considered whistleblowing. This means more time and resources for organisations and more work for labour law attorneys to assist the employer in assessing whether objectionable conditions in accordance with the whistleblowing provisions have taken place (Juridisk ABC, 2024). "The Supreme Court, in accordance with the legislator's intent, has defined a concept of whistleblowing that can easily be interpreted such that many more personal conflicts lead to whistleblowing" (Juridisk ABC, 2024). A too-broad concept of whistleblowing could eventually lead to whistleblowers not being taken seriously. It may imply that one is necessarily considered a whistleblower if criticising one's workplace. Employees may thus be less likely to engage in commonplace use of freedom of speech (Codex Advokat, 2024). Seen from the employer's point of view, if all types of criticism are framed as whistleblowing, they might rather prefer to protect the firm than to find solution (NOU 2022: 9, p. 315).

3.5 Analysis and results so far

The discourse in the examined political and legal documents on whistleblowing in Norway includes the following discursive statements:

1. *Increased use of employees' right to freedom of speech is good for the organisation and society.* Furthermore, to facilitate this:
2. *The criteria for reportable issues should be broad enough to ensure that employees do not refrain from reporting due to uncertainty about whether the specific matter qualifies.*

3. *To ensure that employers handle notifications seriously, transparent whistleblowing routines should be in place.*
4. *A certain loyalty towards the employer is necessary; therefore the whistleblower's procedure should be in accordance with the organisation's whistleblowing routines to be protected against retaliations.*

In our ongoing research on the early history of the whistleblowing concept, we have found that its rise should be understood in light of population politics (Foucault, 2001; 2004). Starting already in the 18th century, the term population takes on a new meaning, as Foucault has observed. "[P]opulation comes to appear above all else as the ultimate end of government [which] has as its purpose [. . .] the welfare of the population, the improvement of its condition, the increase of its wealth, longevity, health, and so on" (Foucault, 2001, p. 216). Initiatives which influence public health, fertility, and the raising of children, as well as measures which affect the economy, increase production, facilitate the export of goods, and create opportunities for employment, become important.

The public interest aspect across the discourse we have analysed articulates population politics. Whistleblowers are necessary to protect the economy and the population's health and productivity. Organisations need whistleblowers to detect irregularities, which in the long run risk negatively affecting the organisation's productivity and reputation. These irregularities may, for instance, be fraud, waste of public money, an unhealthy work environment, or security issues. This may lead to economic costs, less efficient welfare institutions, accidents, sick leave, and loss of competence, which in turn affect variables on which the population depends, such as health, employment rate, and productivity. We see that the public interest aspect of the discourse frames whistleblowing as a CSR instrument. In a complex knowledge society, insiders' knowledge is indispensable to uncover risk of harm to the population and the economy and thus contribute to the common good. However, the public interest concept in this context is not limited only to population political issues; it also designates, for instance, environmental concerns and human and animal rights.

A premise for the discourse we have identified has been (mistakenly) that the wider the concept of objectionable conditions (albeit in combination with whistleblowing routines and protection against retaliations), the more cases of objectionable conditions related primarily to productivity and the quality of the population will come to the surface. However, a function of this discourse (encompassing the whistleblowing provisions) appears to be the predominance of personal-related cases within whistleblowing channels. Another aspect of this discourse is a change in the concept of whistleblowing: primarily signifying it as a process internal to organisations. Consequently, this emphasis shifts the focus away from public interest – a fundamental element for garnering public attention in cases of external whistleblowing – relegating it to a more remote consideration. Furthermore, whistleblowing is likely to become a "managerial tool" to regulate information in cases of public interest (see Du

Plessis, 2022). For cases of public interest, employers' obligations related to the processing of the notifications could even be sharpened, as Gammelsrud argues (2022).

Organisational psychologist Inge W. Brorson argues that although many whistleblowers report a professional reception of their notification, several experience stigmatisations (Brorson, 2023). He refers to well-known examples from contemporary Norway, such as the army and health services. According to Brorson, the strategies for rejection of whistleblowers can be subtle, including ignoring or trivialising the criticism.

If personal cases and conflicts regarding the working environment, for instance, reach a certain volume they may indicate a systemic problem of malaise in the workplace. Whistleblowing can afford an opportunity for this to come to light and produce outcomes. An organisation which systematically harasses its employees or expects unreasonably high productivity based on insufficient personnel or professional resources *is* of public interest. Institutionalised ill-treatment of employees or the burdens of too demanding productivity expectations in an organisation may lead to health problems and sick leaves, which have consequences for the individuals, the organisation's productivity, and the welfare state. Yet, a whistleblowing case of *public interest* in such a situation would be one which points out the *systematic* character of the ill-treatment, based on knowledge of the many individual cases. For this to be possible, personal cases must be registered. By problematising that multiple commonplace personal cases pass through whistleblowing channels, we do not intend to downplay the potential severity of such cases. However, as the lawyer in the interview on 29 February 2024 points out, they should be addressed based on management authority and personal responsibility.

There is no tension in the discourse we have identified as to the importance of ensuring a low threshold for employees' right to safely and publicly – when appropriate – criticise their organisation. Neither are there disagreements as to whether personal matters and conflicts should be solved as efficiently and painlessly as possible. Also, there is consensus as to facilitate the reporting of objectionable conditions in an organisation's undertaking in accordance with the Working Environment whistleblowing provisions. However, we have seen that the current whistleblowing institute and practice of it tend to blur the distinction between general criticism, personal matters, and objectionable conditions in the organisation's undertaking. We have identified a significant emphasis on *whistleblowing in the public interest* in the discourse. However, given the current formulations of the provisions, the majority of the Supreme Court's conclusion makes sense. It makes sense from a strict jurisprudential interpretation of the letter of the law. However, as we hope to have made clear by now, this verdict is unfortunate for the institution of whistleblowing, and we endorse the dissenting justice's argument. Yet, in our opinion, there are weaknesses in the legislation that pave the way for a conclusion like the one reached by the majority of the Supreme Court. This conclusion further dilutes the above-mentioned distinction. A trivialised whistleblowing concept means that commonplace cases – such as personal

matters of which there are many daily in Norway – and general criticism – of which the threshold should be low – become fuelled into the same system and, in principle, require analogous rigorous investigation, as systemic failure in the organisation's undertaking. The latter would often need thorough investigation by external parties, and its revelation and correction are likely to be in the public interest. The erasure of this distinction makes it more difficult to generate knowledge regarding the efficacy of whistleblowing as a CSR instrument and inform the public about severe problems. Furthermore, personal cases are often resolved with less conflict and stress if the organisation takes personal responsibility, without involving external parties. Thus, the organisation would save resources and costs. We have seen that employers may be saturated by the large number of resource-demanding whistleblowing cases and pursue subtle ways to protect the organisation and silence criticism. Relatedly, there is a risk of a chilling effect: employees may be more reluctant to engage in critical dialogue in their organisation if the distinction between ordinary criticism and whistleblowing is unclear. This could also result in less exposure of public interest cases.

4 Analysis of the history of Norwegian whistleblowing research

In this section we analyse statements in the most influential type of whistleblowing research, organisational sociology, published from 2005 to 2018. In the methodology section, we have elaborated on the reasons why we have chosen these examples. The limited number of publications we selected implies that our ambition is to uncover a trend in the history of Norwegian whistleblowing research.

4.1 A historical development in organisational sociology

We have analysed methodological and definitional statements in four influential surveys illustrating a continuity in diluting the understanding of whistleblowing in organisational sociology.

This diluting understanding in the surveys functions as the basis for the collection of data among many employees in Norwegian working life, concerning their understanding of whistleblowing. Thus, the answers from these respondents are, to a large degree, determined by the wide scope of these diluting methodological and definitional statements.

The surveys are "Whistleblowing in Norwegian working life: What does it mean and what do we know?" (Skivenes and Trygstad, 2005), "Whistleblowers: About employees who speak up!" (Trygstad and Skivenes, 2006), "Explaining whistleblowing processes in the Norwegian labour market: between power resources and institu-

tional arrangements", (Trygstad and Skivenes, 2015), and "Whistle blowing in Norwegian working life. The person as object of whistleblowing, receivers and reactions" (Trygstad and Ødegaard, 2018).

These statements reflect a development characterised by a three-stage pattern of dilution. The statements are from the one survey to the next linked to a definition and an expansive, differentiated, and comprehensive methodology, significantly broadening the concept of objectionable conditions to include a large plurality of problems. The objectionable conditions span from the most serious problems in the public interest (such as corruption; unwillingness to correct deficiencies in services and products that could cause damage; breaches of rules on safety, health, and environmental conditions; use of illegal chemicals; bullying) through less severe problems more in the interest of an organisation (such as lack of willingness to discuss deficiencies in the service offered; lack of participation in decisions of importance to the workplace) to problems of more personal interest (such as questions related to working hours, lack of compensation for overtime, lack of respect from managers, incompetent managers). This dilutional understanding of the scope of problems has consequences for what the respondents necessarily are able to understand as whistleblowing. Whistleblowing is correspondingly understood very broadly, as reporting commonplace cases of interest primarily to the individual, general criticism, and reporting of cases in the interest primarily of the public.

This interpretation of a pattern of dilution occurring through these three categories of objectionable conditions and whistleblowing is indeed a reduction of the complexity of these survey's more detailed descriptions of both objectionable conditions and whistleblowing. We distinguish between these three categories of problems and reporting, pretending that they are sharply divided. However, in the reality of the surveys, they often overlap and thus represent more categories. There are grey zones. However, both showing grey zones and the more detailed variation of concepts would rather contribute to confusion because of the tendency of increased complexity than to simplify and give an overview of the trend of dilution, which is mainly caused by these three categories we have described.

4.2 Analysis of problems

4.2.1 The history of the broad methodology of organisational sociology

The primary driver of the risk of trivialisation of the concept and practice of whistleblowing is the sociological continuous use of a quantitative methodology in the form of a differentiated questionnaire, broadening the concept extensively.

The differentiated and very broad scope of the quantitative methodology means that the research promotes an unclear (Winnæss, 2007), i.e., fluid, concept (Wittrock, 2024) of what whistleblowing is in the eyes/face of its many respondents, facilitating

that the totality of their answers represents the same broad understanding of whistle-blowing structured roughly by the three categories mentioned. On the one hand, one can thus say that the fluid concept of whistleblowing legitimises and enables many different types of criticism of conditions in working life, the most serious types, the moderately serious, the less serious, and the least serious, making whistleblowing into something trivial, commonplace. This can have important advantages. Employees might be encouraged to criticise problems deserving of it. Problems might not go under the radar as serving for possible solutions. On the other hand, the fluid use of the term creates problems. Let us mention three.

4.2.2 Covering up subjective assumptions

The use of differentiated questionnaires as the ones we have analysed means that one equates a respondent with low and one with high tolerance for relatively trivial prob-lems, rudeness towards users, use of intoxicants at work, reluctance to correct mis-takes, neglect in relation to work tasks, and unfair advantages to certain employees/customers. The person with low tolerance will think that such relatively banal condi-tions should be notified; the person with high tolerance will think that they should not be notified. Such questionnaires level out these individually different assumptions behind the answers and thus shade individual assumptions that will serve as the basis for the trivialisation of whistleblowing that we have pointed out. Thus, neither do we get data about how large is the percentage of respondents representing the low tolerance and low threshold supporting a notification, nor do we get data about the respondents with the high tolerance and high threshold not supporting a notification. Consequently, the lack of such subjective data in these questionnaires means that even if we can identify a fluid understanding of whistleblowing in the way of present-ing comprehensive options in these questionnaires, we do not know the degree to which the subjectivity of the respondents contributes to the fluidityness and thus the trivialisation of the concept of the whistleblowing. One type of subjectivity might be of specific importance. The respondent with low tolerance for commonplace and triv-ial problems deciding to notify them might be those among the respondents who pri-marily contribute to the fluidityness and trivialisation of the concept. This could be understood against the historical background of a modern society of strong individu-alism and de-collectivising characterised by personalities critical first to common-place problems which could be disadvantageous for their successful achievements and carriers as individuals.

4.2.3 The fluid concept and the lack of distinction

The quantitative research's use of differentiated questionnaires inadvertently promotes a fluid and diluted concept of whistleblowing, which makes it almost impossible to determine what separates whistleblowing as employee criticism of conditions in working life from other types of criticism of such conditions by employees, ordinary criticism of commonplace problems in organisations, and criticism based on personal interest. Since determining the uniqueness of this term becomes almost impossible, it no longer appears important to use it. You can just as easily talk about employees' criticism of conditions in working life. Through dilution, the term seems to make its own meaning vague and meaningless, as the boundaries separating it from other criticisms are blurred.

4.2.4 Epistemology and the social responsibility of science

We continuously stress the importance of distinctions of this type for another reason as well. Science's ability to make precise distinctions and use concepts rigorously affects the quality of knowledge and the social usefulness of science. When science can discern, it creates valuable knowledge for society. But when the distinctions mentioned are not introduced, scientific investigations do not give society knowledge fine grained enough to be a guide to effective understanding and action. Are individuals who report serious issues such as corruption, of which society has an interest in being aware, facing severe repercussions like retaliation, harassment, or dismissal? Might there as well be severe consequences for society of such a negative development? And, on the other hand, are much less negative consequences linked to persons notifying about minor problems that are essentially of individual interest? On the other hand, are those reporting minor and personal problems suffering less dire consequences and milder retaliations? As we have shown in section 3, the discourse we have identified articulates a need for more whistleblowing in cases of public interest. These are cases primarily related to the economy and population. At this point, we only know there is a trend towards increasing retaliation against whistleblowers. Whether this trend specifically targets whistleblowers of public interest cases would be crucial to understand: has whistleblowing, which motivated the whistleblowing institute, become less effective? It would be likely to assume so if these whistleblowers are more likely to face retaliation. The broad concept of whistleblowing makes it difficult to gain insight into this.

   This distinction seems to be a reasonable working hypothesis for future research that aims at exploring the correlation discerning between different grades of severity of blowing the whistle and followed logically by exploring if this difference is linked to different grades of negative consequences for the persons blowing the whistle. The societal implication of such discerning is important. Based on our analysis of the Nor-

wegian whistleblowing research (in our mentioned upcoming comprehensive bibliography), If it turns out that not only the few examples of research we have analysed but also the whole body of whistleblowing research is unable to produce knowledge about whether revenge, harassment, and dismissal are hitting first whistleblowers who report serious problems such as corruption, illegal surveillance, dangerous technology, and products, this is a problem for both science and society. Society is then not equipped by science to protect these types of critics in a focused and sufficient way as a first step to solving these problems efficiently; thus the problems might increase. And science does not take its role of honouring its social responsibility seriously. Science rather neglects its social responsibility by ignoring society's need for accurate knowledge. Correspondingly, if the whole body of whistleblowing research is incapable of producing knowledge about what kind of consequences are suffered if it is small negative consequences such as collegial relocation at the workplace and lack of career progress that affect whistleblowers who report moderate problems such as collegial conflicts and dissatisfaction with leadership style, again both science and society have a problem. Society is then not equipped by science to consider these types of less significant critics and consequences as less in need of protection. Society might rather be confused and use the provisions of law to protect in an equally serious way cases of different importance. Science is not fulfilling its role responsibly; it fails to take social responsibility seriously and often ignores society's needs for acting on the right type of knowledge.

5 Conclusion

In Norway, the threshold for reports to be considered whistleblowing is notably low, a standard recently reaffirmed by the very first judgment of the Norwegian Supreme Court concerning whistleblowing. The Norwegian whistleblowing concept of 2007–2023 encompasses a broad range of objectionable conditions. We have shown that the Norwegian whistleblowing concept and legislation emerged to expose and rectify issues of public interest within organisations. As such, whistleblowing has the potential to detect when CSR fails and serve to restore it. However, we contend that the introduction of whistleblowing legislation and its development towards the current provisions undermines our capacity to gauge its impact as a CSR tool. Despite this uncertainty, we have identified a *risk* that the low threshold for classifying notifications as whistleblowing limits our ability to evaluate whether whistleblowing is effective when CSR fails.

Our analysis has spanned legal provisions, scholarly research within sociology and national surveys and reports on whistleblowing in Norway. Through our Foucault-inspired discourse analysis of these documents, we have found that the establishment and institutionalisation of whistleblowing practices are tied to a power interest in encouraging a greater use of employees' freedom of speech, predominantly to prevent or

halt severe damage – of public interest – to the population and the economy. Yet, a premise for this discourse has been that by broadening the concept of objectionable conditions and having loose formal criteria for what constitutes whistleblowing by law, more cases in the public interest will emerge. A consequence of this fluid discourse, however, is that the distinction between personal cases, general criticism, and whistleblowing is blurred. The possible predominance of personal-related cases within whistleblowing channels is a result. Consequently, it remains difficult to understand and gain insight into to what extent a function of whistleblowing makes organisations become more responsible for the common good. We have also seen that the increasing number of whistleblowing cases which labour attorneys face may be understood in the context of a working life that has become more streamlined and has less space for people who are standing out. Intuitively, one would perhaps think that a low threshold for reporting means a more humane working life. However, when whistleblowing shifts focus from addressing systemic issues within the organisation to targeting individuals who allegedly do not align with a streamlined working culture, there is less space for individuals who stand out. Thus, whistleblowing could become a means of enforcing conformity rather than promoting a more inclusive and humane workplace. Another risk after the decision in the Supreme Court is a further commercialisation of the whistleblowing institution. Labour attorneys are of the opinion that increased worries among employees on how to interpret the low threshold of the legal provisions according to the liberal decision in the Supreme Court might easily lead to extended use of labour attorneys.

The whistleblowing concept in the legislation is likely to be reflected in the approach of organisations to whistleblowing. We have seen that the law mandates that organisations (with five or more employees) must implement whistleblowing routines. This legal requirement suggests that organisations are likely to develop their whistleblowing procedures in alignment with the legislative concept of whistleblowing. Furthermore, in research, such as surveys to explore the prevalence and experiences with whistleblowing in Norway, one would assume that scholars are likely to base their understanding of whistleblowing on the legislative conceptualisation. This assumption is logical because employees' perceptions of whistleblowing are presumably shaped by how the concept is framed within their organisations' routines (but see chapter 10, where it is shown that translation of law into organisational practice affords significant leeway). Thus, it is reasonable to conclude that the depiction of whistleblowing in legislation, organisational practices, and survey research are interconnected. However, we have seen that the conceptualisation of whistleblowing in selected research publications is more complex.

Whistleblowing research at Fafo done by Trygstad and Skivenes has shown that since the turn of the millennium, the institution of whistleblowing has suffered significant drawbacks. On average, Norwegian employees, during a period of 15 years, decrease their whistleblowing activity significantly, experience more retaliation, are less interested in blowing the whistle again, and feel to a larger degree that their mes-

sage is not taken seriously by their recipients. However, as the broad whistleblowing concept blurs the line between whistleblowing in the public interest and individual grievance reporting, we have no systematic knowledge of whether the increase in retaliations applies to public interest reporting.

6 Suggestions for improving the Norwegian whistleblowing institute

We proceed to suggest some strategies to tailor the understanding of whistleblowing to facilitate notifications of objectionable conditions *of public interest.* These are also strategies to gain data on the prevalence of this type of notification in Norwegian society, as well as their impact and the safety for these whistleblowers.

Specification of concepts in the whistleblowing legislation: Investigative work should be conducted with the intention of renewing the whistleblowing legislation. It should be explored how the concept of whistleblowing can be more specifically refined to notifications of wrongdoing in the undertaking of the organisation, often in the public interest, to be revealed and corrected. In this context, many organisations will probably need clearer distinctions between general criticism, personal cases, and whistleblowing. Provisions which establish that whistleblowing procedures apply to objectionable conditions of *public interest* may result in a less ambiguous and more transparent whistleblowing institute.

Specification and differentiation of whistleblowing concept in larger surveys: Research reports using surveys to gather data on employees' experiences with whistleblowing – including threshold for reporting and consequences for the whistleblower – should provide specific data on the extent of whistleblowing of public interest. Thus, we suggest that future research should explore whether employees are more likely not to blow the whistle again about cases of public interest because of experiences with the burdens of retaliation. We also suggest more research on employees' experiences of the effectiveness of blowing the whistle about cases of public interest. If this tendency is negative, we as a democratic society have a problem.

Media research on whistleblowing and public debate: In light of the recent legislative development among the 27 member states of the European Union we suggest that future media research investigate (a) if media to a larger degree than before use whistleblowing cases primarily of personal interest, as to draw public attention to dramatic or sensational experiences of these individuals, and (b) if media to a more limited degree than before are able to identify and use hardly accessible information from whistleblowers about cases of public interest to raise public debate. We suggest that media research gather data if this partly is due to a limitation of the number of investigative journalists,

the most expensive type of journalism, which has suffered a decline in recent decades when media lost large income from advertisements to social media platforms.

References

Alm, K. and Karlsen, H. (2024) Employees freedom of speech as contribution to the modernization of the Norwegian welfare state. An annotated, digital bibliography on Norwegian research on whistleblowing with an introduction. 1970–2023.

Aagaard, E. C. (2005). Lovreguleringen om adgangen til varsling – noen spørsmål. *Arbeidsrett, 2*(3), 189–192. https://doi.org/10.18261/ISSN1504-3088-2005-03-06

Aastveit, I. H. F. (2024, February 1). *Høyesterettsdom / Arbeidsrett – Lav terskel for varsling*. Harris. https://harris.no/aktuelt/hoyesterettsdom-/-arbeidsrett-lav-terskel-for-varsling/

Arbeidsmiljøloven. (2005). *Lov om endringer i arbeidsmiljøloven (varsling)*. Lovdata. https://lovdata.no/doku ment/LTI/lov/2006-12-01-64

Bjørkelo, B., Ryberg, W., Matthiesen, S. B., & Einarsen, S. (2008a). When you talk and talk and nobody listens": A mixed method case study of whistleblowing and its consequences. International Journal of Organisational Behaviour, 13(2), 18–40.

Bjørkelo, B. & Matthiesen, S. B., (2008b). Sladrehank skal selv ha bank: Om" whistleblowing" i norsk arbeidsliv. Tidsskrift for norsk psykologforening.

Bjørkelo, B., Einarsen, S., Nielsen, M. B., & Matthiesen, S. B. (2011). Silence is golden? Characteristics and experiences of self-reported whistleblowers.

Bjørkelo, B., Thorsen, C., D'Cruz, P., & Mikkelsen, E. G. (2021). Whistleblowing and bullying at work: the role of leaders. Special topics and particular occupations, professions and sectors, 75–108.

Brink, A. G., Eller, C. K., & Green, K. Y. (2018). The effects of corporate social responsibility and wrongdoer rank on whistleblowing. Accounting and the Public Interest, 18(1), 104–128.

Brorson, I. W. (2023, May 21). Ledere bør lytte til varslere, ikke gi dem rødt kort. *Dagens næringsliv*. https://www.dn.no/innlegg/varsling/ledelse/laring/ledere-bor-lytte-til-varslere-ikke-gi-dem-rodt-kort /2-1-1450349

Christ, K. S. (2020). Varslingskultur i en petroleumsvirksomhet (Master's thesis, University of Stavanger, Norway).

Codex Advokat. (2024). *Ny Høyesterettsdom om varslingsbegrepet*. Codex Advokat. https://codex.no/aktuelt/ dom-varslingsbegrepet

de Ruiter, F. (2022). " 1001 gjengjeldelser?"-En kvalitativ studie av varsling i kommunal sektor (Master's thesis, University of Agder).

du Plessis, E. M. (2022). Speaking truth through power: Conceptualizing internal whistleblowing hotlines with Foucault's dispositive. *Organization, 29*(4), 544–576. https://doi.org/10.1177/1350508420984019

Eggen, K. (2004). Ansattes ytringsfrihet. Rettslige bånd eller gyldne lenker?1. *Arbeidsrett, 1*(1), 2–23. https:// doi.org/10.18261/ISSN1504-3088-2004-01-02

Eggen, M. (2008). *Arbeidsgivers tilrettelegging for varsling* [Master thesis, University of Oslo]. https://www. duo.uio.no/handle/10852/21946

Engebretsen, G. A. (2012). Varsling av økonomisk kriminalitet-sett fra Corporate Governance-perspektiv (Master's thesis).

Flaatten, I. (2007). *Arbeidstakers adgang til å varsle etter ikrattredelse av de nye varslingsreglene i arbeidsmiljøloven* [Master thesis, University of Oslo]. https://www.duo.uio.no/handle/10852/21090

Foucault, M. (1969). *L'archéologie du savoir*. Gallimard.

Foucault, M. (1972). *The Archeology of Knowledge*. Pantheon Books.

Foucault, M. (2001). *Dits et écrits, 1954-1988: 2 : 1976–1988* (Vol. 2). Gallimard.

Foucault, M. (2004a). *Naissance de la biopolitique: Cours au Collège de France, 1978–1979*. Seuil Gallimard.

Foucault, M. (2004b). *Sécurité, territoire, population: Cours au Collège de France, 1977–1978*. Seuil Gallimard.

Gammelsrud, E. H. (2022, October 23). Et misforstått varslingsinstitutt? *Dagens næringsliv*. https://www.dn.no/innlegg/arbeidsliv/arbeidsmiljoloven/varsling/et-misforstatt-varslingsinstitutt/2-1-1339315

Gravås, A.K. (2007) Forsvarlig varsling. (Master thesis)

Hagen, I. (2024, January 15). *Ny dom fra Høyesterett om varsling av kritikkverdige forhold – Hva er et «varsel»?* Stiegler Advokatfirma. https://stiegler.no/aktuelt/ny-dom-fra-hoyesterett-om-varsling-av-kritikkverdige-forhold-hva-er-et-varsel/

Hustad, J. (2018). Kven skal varsle om varslarane? *Nytt Norsk Tidsskrift*, *35*(3–4), 306–311. https://doi.org/10.18261/issn.1504-3053-2018-03-04-11

Innst. 303 L (2016–2017). (2017). *Innstilling frå arbeids- og sosialkomiteen om Endringer i arbeidsmiljøloven (varsling og arbeidstid)* [Inns]. Arbeids- og sosialkomiteen. https://www.stortinget.no/no/Saker-og-publikasjoner/Publikasjoner/Innstillinger/Stortinget/2016-2017/inns-201617-303l/?all=true

Juridisk ABC. (2024, January 16). *Fikk arbeidsrettsadvokater en førjulsgave fra Høyesterett?* Juridisk ABC. https://juridiskabc.no/arbeidsrett/fikk-arbeidsrettsadvokater-en-forjulsgave-fra-hoyesterett/

Karlsen, H. (2023). Foucault's archeological discourse analysis with digital methodology – Discourse on women prior to the first wave women's movement. *Digital Scholarship in the Humanities*, *38*(1), 195–208. https://doi.org/10.1093/llc/fqac022

Lovvedtak 70 (2018–2019). (2019). *Vedtak til lov om endringer i arbeidsmiljøloven (varsling)* [Lovvedtak]. Stortinget. https://www.stortinget.no/no/Saker-og-publikasjoner/Vedtak/Beslutninger/Lovvedtak/2018-2019/vedtak-201819-070/

Midttun, A. (2013). CSR and beyond: A Nordic perspective.

Norges Høyesterett. (2023). *Dom: HR-2023-2430-A*. Lovdata. https://lovdata.no/dokument/HRSIV/avgjorelse/hr-2023-2430-a

NOU 1999: 27. (1999). *«Ytringsfrihed bør finde Sted» – Forslag til ny Grunnlov § 100*. Justis- og politidepartementet. https://www.regjeringen.no/no/dokumenter/nou-1999-27/id142119/

NOU 2004: 5. (2004). *Arbeidslivslovutvalget – Et arbeidsliv for trygghet, inkludering og vekst*. Arbeids- og administrasjonsdepartementet. https://www.regjeringen.no/no/dokumenter/nou-2004-5/id384819/

NOU 2018: 6, A. (2018). *Varsling – verdier og vern – Varslingsutvalgets utredning om varsling i arbeidslivet*. Arbeids-og sosialdepartementet. https://www.regjeringen.no/no/dokumenter/nou-2018-6/id2593665/

NOU 2022: 9. (2022). *En åpen og opplyst offentlig samtale – Ytringsfrihetskommisjonens utredning*. Kultur- og likestillingsdepartementet. https://www.regjeringen.no/no/dokumenter/nou-2022-9/id2924020/

Olsen, I. (2007). Varsling i arbeidsforhold: En fremstilling og vurdering av forsvarlighetskravet i aml. § 2–4 (2) første punktum ved varsling om kritikkverdige forhold i virksomheten (Master's thesis).

Ot.prp. nr. 49 (2004–2005). (2005). *Om lov om arbeidsmiljø, arbeidstid og stillingsvern mv. (Arbeidsmiljøloven)* [Proposisjon]. Arbeids- og sosialdepartementet. https://www.regjeringen.no/no/dokumenter/otprp-nr-49-2004-2005-/id396602/

Ot.prp. nr. 84 (2005-2006). (2006). *Om lov om endringer i arbeidsmiljøloven (varsling)*. Arbeids- og inkluderingsdepartementet. https://www.regjeringen.no/no/dokumenter/otprp-nr-84-2005-2006-/id189449/

Prop. 72 L (2016–2017). (2017). *Endringer i arbeidsmiljøloven (varsling og arbeidstid)* [Proposisjon]. Arbeids- og inkluderingsdepartementet. https://www.regjeringen.no/no/dokumenter/prop.-72-l-20162017/id2545705/

Prop. 74 L (2018–2019). (2019). *Endringer i arbeidsmiljøloven (varsling)* [Proposisjon]. Arbeids- og inkluderingsdepartementet. https://www.regjeringen.no/no/dokumenter/prop.-74-l-20182019/id2639812/

Sallaberry, J. D., Conesa, I. M., & Flach, L. (2020). Relationships between the corporate social responsibility orientation and whistleblowing: A study with Hispano-American auditors. A Contabilidade e as Novas Tecnologias, 1-17.

Skivenes, M., & Trygstad, S. C. (2005). Varsling i norsk arbeidsliv: Hva betyr det og hva vet vi ?

Skivenes, M., & Trygstad, S. C. (2006). Varslere: Om arbeidstakere som sier ifra!

Skivenes, M., & Trygstad, S. C. (2010). When whistle-blowing works: The Norwegian case. Human Relations, 63(7), 1071–1097.

Skivenes, M., & Trygstad, S. (2015). Explaining whistle blowing processes in the Norwegian labour market: Between individual power resources and institutional arrangements. Economic and Industrial Democracy, 36(1), 119–143.

Stankiewicz-Mróz, A. (2015). Ethical code and whistleblowing as CSR tools in pharmaceutical companies. PRACE NAUKOWE, 158.

Steen, A. H. (2012). *Arbeidsmiljølovens bestemmelser om varsling* [Master thesis, University of Oslo]. https://www.duo.uio.no/handle/10852/35499

Støver, L. (2009). *Varsling – Et spørsmål om lojalitet?* [Master thesis, Norwegian University of Science and Technology]. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/268146

Trygstad, S. C., Skivenes, M., Steen, J. R., & Ødegård, A. M. (2014). *Evaluering av varslerbestemmelsene | Fafo-rapport 2014:05*. Fafo. https://www.fafo.no/zoo-publikasjoner/evaluering-av-varslerbestemmelsene

Trygstad, S. C., & Ødegård, A. M. (2018). Varsling i norsk arbeidsliv. Den det varsles om, mottak og håndtering.

Winnæss, K. (2007). Arbeidstakers varsling på grunnlag av påståtte kritikkverdige forhold i virksomheten.: En fremstilling og vurdering av arbeidstakers ytringsfrihet sett i forhold til arbeidstakers ulovfestede lojalitetsplikt overfor arbeidsgiver (Master thesis).

Wittrock, C. (2024). Criticism of an early draft. Presented at the AFINO workshop at OsloMet, January 2024.

Working Environment Act. (2005). *Act relating to the working environment, working hours and employment protection, etc.* Lovdata. https://lovdata.no/dokument/NLE/lov/2005-06-17-62/KAPITTEL_3#KAPITTEL_3

Yang, L., & El Mahdy, D. (2022). Social Networking (Guanxi) and Whistleblowing Intentions: Does CSR Ring the Bell? Journal of Forensic and Investigative Accounting, 14(2).

Caroline D. Ditlev-Simonsen

Chapter 10
Acting on the Norwegian Transparency act: interpretation and implementation

The chapter delves into the ramifications of Norway's Transparency Act (Åpenhetsloven, 2021), which was enacted on 1st of July, 2022, compelling businesses to foster human rights and fair working conditions in their supply chains through enhanced transparency. It scrutinizes the interpretation and operationalization of The Act within two distinct companies, employing the Knowledge Transfer as Translation (KTT) theory – traditionally applied to knowledge transfer within corporate culture – to navigate The Act's conversion into corporate actions. This exploration uncovers the obstacles and divergent compliance strategies among the firms, showing that The Act's indeterminate language and the specific resources and individuals within each company lead to varied corporate reactions. Despite The Act's objective to improve supply chain transparency, the lack of clear norms or a unified understanding of the legislation at this early stage results in inconsistent applications. The study also posits that KTT offers a valuable framework for examining the enactment of not only abstract cultural issues but also tangible legal mandates, suggesting its broader applicability in legal interpretation and corporate action alignment.

1 Introduction

Industrialization and mass production started off with poor and uncontrolled working conditions: child labour, slavery, no worker organization opportunities, non-existing workers' rights, low pay, long working days, hazardous working conditions, etc., were common characteristics of factories in the 19th century. Gradually, as the economy grew and democracy and strong governmental organizations developed in the western world, so did rights and demand from workers, equality, workers' rights, pay, etc., increase. The creation of the International Labour Organization (ILO) in 1919 is an example that illustrates the move towards improving working conditions. However, as working conditions improved in industrialized countries in Europe and the United States, so did the cost of production. As a result, many companies selling products requiring manual work but no education moved to Asia, where the working conditions in many cases mirrored those in the West before the rights and organization of workers were established.

Production relocation in low-cost countries started in the sixties and has grown tremendously. Companies take advantage of this opportunity for low-cost production

based on violation of human rights, as it reduces costs and increases profit. As many of these low-cost countries have poor regulations in the field of working condition standards and an even less functioning system to follow up these regulations, the problem is still there. We continuously witness tragedies and scandals associated with human rights violation issues.

Global companies with well-known brands have worked actively through voluntary initiatives like the UN Guiding Principles on Business and Human Rights (UNGP) and the OECD Guidelines for Multinational Enterprises on Responsible Business Conduct, to ensure acceptable working conditions in their supply chain. Still these companies are frequently involved in tragedies and scandals associated with human rights issues. Examples of such events are child labour in mines in Kongo (Apple, Google, Microsoft, Tesla, etc.), unpaid overtime and poor working conditions, modern slavery in textile factories in Bangladesh (H&M, Zara, etc.), and race-based disparities in pay and promotions (Coca-Cola). Similar cases involving Norwegian companies are, for example, Telenor suppliers involved in child labour and unacceptable working conditions in Bangladesh, illegal African immigrants picking tomatoes without pay in Southern Italy, and unpaid overtime among producers of wines sold at Vinmonopolet (the Norwegian monopoly for wine and spirits). And it is not only in low-cost countries that such breaches are experienced. Norway is continuously experiencing breaches of human rights. For example, Polish workers through municipality contracts are paid less than 69 NOK per hour (less than US$7) (Aarseth, 2023), and employees in the Espresso House cafes did not receive wages for several months (Wig, 2023). In addition, human trafficking of sex slaves is constantly being uncovered in Norway. In chapter 6 Midttun discusses how H&M addresses breaches of human rights among suppliers in Xinjiang, where 85% of the company's cotton is produced.

Numerous tragedies and scandals resulting from corporate mismanagement spurred various initiatives that reconsider the extent of a company's responsibilities. The concept of Corporate Social Responsibility (CSR) is among the most recognized and is increasingly embraced by various stakeholders. Different definitions of CSR have been proposed, but the EU definition is maybe the most relevant to mention here: "the responsibility of enterprises for their impact on society" (European Commission). This is a vague definition and it is understood and interpreted differently by different stakeholders. Around the turn of the century, the use of the term CSR expanded exponentially; still, CSR was not directly linked to regulations and therefore was often perceived as a buzzword (see chapter 1 for a discussion of CSR as window dressing).

In Norway, the term CSR has been used as an abbreviation, as well as translated into the term "bedrifters samfunnsansvar" – which can be directly translated into "corporate social responsibility". In Norway "societal responsibility" includes environmental and social issues, working environment, equality, non-discrimination, compliance with human rights, and repudiating corruption. According to Norwegian Accounting Law, in 2013 large companies, around 1000 of Norway's 500,000 companies, were required to report on "societal responsibility" (Endringslov til regnskapsloven, 2013). Still

the format and scope were unclear and interpreted differently by corporations. For more detailed information about Norwegian law associated with environmental and social issues, see Rapportering om Samfunnsansvar i små foretak (CSR reporting in small corporations) (Ditlev-Simonsen, 2016).

Globally, large corporations' focus on CSR reporting increased dramatically. Companies wanted to communicate their CSR engagement – and avoid criticisms. Internationally there was a proliferation of CSR or corporate reporting initiatives, like the OECD Guidelines for multinational enterprises, Global Reporting Initiative (GRI), UN Guiding Principles, UN Global Compact 10 principles, and Fairtrade, and the same happened at the national level in Norway, with initiatives like certifying the environmental credential of products. However, all these reporting and certifications were *voluntary.* Such voluntary initiatives were obviously not enough to avoid unacceptable social behaviour among corporations.

Based on the continuous violations of human rights, unacceptable working conditions, and the inadequate achievements of voluntary initiatives, the UK and France introduced *regulation* for large companies to ensure good or adequate measures in the supply chain. The shift from soft to hard laws, including fines for breaking these laws, led to a change in attitudes and operations for companies with production in low-cost countries.

In Norway on 1 July 2022, the Transparency Act (The Act) (åpenhetsloven) was passed. The purpose of the law is to make companies responsible for respecting basic human rights and decent working conditions by demanding increased transparency. This study investigates the impact of The Act on two companies, one large and well known, and one medium sized, but still covered by the law. As The Act is vaguely formulated, the corporate interpretation and activities anchored in the law might differ. At first, it only required companies to respond to questions about their supply chain, but 1 year later, on 30 June 2023 the same companies were required to carry out a due diligence process and report on this on their website. The interviews were conducted in November and December 2023, which is very timely for looking closer at the impact of the law. The fact that the norm and interpretation of the law are not yet established makes the study particularly interesting.

In each of the two companies I have interviewed, the person who has had the key role in managing or organizing the compliance with The Act. Then I investigated the process of implementing The Act by applying the Knowledge and Translation Theory (KTT).

I begin with a review of the background and the content of The Act. Thereafter, the theory and method of KTT are introduced and discussed. The study goes beyond the already explored application of KTT on the translation of corporate and country cultures to applying the theory on interpretation of concrete laws: The Act. The two cases as well as the subject of the interviews and the result of the interview will be presented. The findings will be analysed and discussed before the conclusion.

2 The background for the Norwegian Transparency Act

In response to the challenges above, the first law on transparency in the supply chain, the California Transparency in Supply Chains Act (TiSC), was enacted in 2010 (Harris, 2015). The law requires Californian companies to report on their efforts to combat slavery and human trafficking. However, it was after the Rana Plaza disaster in Dhaka, Bangladesh, in 2013 that awareness of unacceptable working conditions in low-cost countries came to the forefront. Over 3,000 textile workers, mainly women, working in this factory when it collapsed were injured, and over 1000 died. The image of the collapsed building and the tragedies associated with it made headlines around the world. Since well-known brands like Zara and Mango contracted with companies in this factory, clothing that many readers might have in their closet, the accident increased awareness and engagement among people in general. This made people and regulators see that soft laws, voluntary initiatives by corporations, were not enough. Hard laws were a necessity.

The first law to address unethical labour conditions in the supply chain in response to the Rana Plaza tragedy was the UK Modern Slavery Act of 2015 (Legislation. gov.uk, 2015). It was introduced to prevent slavery, forced labour, and human trafficking in England. The law builds on TiSC and requires businesses to identify and report on the risks and efforts to eradicate modern slavery within their organization and supply chain. Company boards must approve documents in which the company explains its work on human rights, and this information must be available on the company's website. The fact that the document must be approved by the board requires knowledge and awareness in leadership, which in turn contributes to the work on human rights across the company. The law applies to large companies with turnovers of over 36 million GBP (approximately 480 million NOK). In total, about 12,000 companies were affected.

The French Corporate Duty of Vigilance Law was enacted in 2017. The initiative for this law also came in the wake of the Rana Plaza disaster in Bangladesh in 2013. A central element of the law is mandatory company review. This law establishes a legal obligation for large companies to identify and prevent negative human rights and environmental impacts resulting from their own operations, activities in companies they control, and activities involving their subcontractors and suppliers with whom they have an established commercial relationship (European Coalition for Corporate, 2017). Responsibility lies with the company when they fail to meet their obligations, such as not having a plan for the above-mentioned challenges or not following up on the plan.

The fact that the French law applies to companies with at least 5,000 employees in France and at least 10,000 globally means that it "only" applies to 150–200 companies. At the same time, these companies, through their supply chain, account for

about half of French exports. This is how laws that initially apply to very few companies have a significant impact both globally and nationally. To conduct due diligence on their own operations, companies must gather information about conditions at subcontractors that may not necessarily be covered by the law. This indirectly affects smaller companies.

The majority of laws in the wake of the Rana Plaza tragedy are oriented towards transparency associated with human rights in the supply chain. The French Vigilance Law, however, includes not only human rights issues but also *environmental* issues. Companies are obligated to identify and act on environmental challenges in their supply chain. Examples of what is not perceived as acceptable in the supply chain related to environmental issues might be better conveyed through cases where French companies have been reported as breaking the law. Danone is reported for breaking the law based on the negative impact of its use of plastic. Another example is BNP Paribas, which contributes to pollution by financing oil and gas extraction. So far, it has mainly been the press and NGOs that have been following up on the law from the outside and have reported companies for not complying with the law (Business and Human Rights Information Center, 2024)

Following the aforementioned laws, there have been initiatives for more or less similar laws in countries such as Germany, Canada, the Netherlands, Switzerland, and more. Furthermore, the EU is in the process of developing its own directive on sustainable due diligence, the Corporate Sustainable Due Diligence Directive (CSDDD). The CSDDD will also address environmental matters and was approved on 14 December 2023.

So, in response to transparency regulations related to human rights and the supply chain in other countries, the Norwegian Ministry of Children and Families established the Ethics Information Committee in June 2018. The purpose was to assess whether businesses should have a duty to provide information related to social responsibility and supply chain. The Committee submitted its report to the Ministry of Children and Families in November 2019. This resulted in Proposition 150 L (2020–2021) and The Act, which was passed in June 2021 (Åpenhetsloven, 2021). The law came into effect on 1 July 2022, when companies were obligated to answer requests from stakeholders regarding their supply chain and, by 30 June 2023, required to make the report on their due diligence available online. The law is based on international commitments, especially the UN Guiding Principles on Business and Human Rights, OECD's guidelines for multinational enterprises, and UN Sustainable Development Goals.

3 The content of and the approach to the Transparency Act

The purpose of the Transparency Act is to promote respect for human rights issues in the supply chain from a company point of view. More specifically, as § 1 states, "The Act shall promote enterprises' respect for fundamental human rights and decent working conditions in connection with the production of goods and the provision of services and ensure the general public access to information regarding how enterprises address adverse impacts on fundamental human rights and decent working conditions."(Åpenhetsloven, 2021).

The law applies to companies that meet at least two of the following three criteria: over 70 million NOK in sales revenue, over 35 million NOK in balance sheet total, and an average of over 50 employees in the fiscal year. Initially, this applies to approximately 8,800 companies in Norway.

The law includes three central obligations: the obligation to conduct due diligence, the obligation to account for due diligence, and the right to information (§§ 4–6). According to the Transparency Act, due diligence means (§ 4):

Section 4. Duty to carry out due diligence

"The enterprises shall carry out due diligence in accordance with the OECD Guidelines for Multinational Enterprises. For the purposes of this Act, due diligence means to

a) embed responsible business conduct into the enterprise's policies

b) identify and assess actual and potential adverse impacts on fundamental human rights and decent working conditions that the enterprise has either caused or contributed toward, or that are directly linked with the enterprise's operations, products or services via the supply chain or business partners

c) implement suitable measures to cease, prevent or mitigate adverse impacts based on the enterprise's prioritisations and assessments pursuant to (b)

d) track the implementation and results of measures pursuant to (c)

e) communicate with affected stakeholders and rights-holders regarding how adverse impacts are addressed pursuant to (c) and (d)

f) provide for or co-operate in remediation and compensation where this is required.

Due diligence shall be carried out regularly and in proportion to the size of the enterprise, the nature of the enterprise, the context of its operations, and the severity and probability of adverse impacts on fundamental human rights and decent working conditions. The Ministry may issue regulations regarding the duty to carry out due diligence."

Section 5. Duty to account for due diligence

"The enterprises shall publish an account of due diligence pursuant to Section 4. The account shall at least include

a) a general description of the enterprise's structure, area of operations, guidelines and procedures for handling actual and potential adverse impacts on fundamental human rights and decent working conditions

b) information regarding actual adverse impacts and significant risks of adverse impacts that the enterprise has identified through its due diligence

c) information regarding measures the enterprise has implemented or plans to implement to cease actual adverse impacts or mitigate significant risks of adverse impacts, and the results or expected results of these measures.

Section 6, second paragraph (c) and (d), third and fourth paragraph correspondingly apply to the duties pursuant to the first paragraph.

The account shall be made easily accessible on the enterprise's website and may form part of the account on social responsibility pursuant to Section 3–3 (c) of the Accounting Act. The enterprises shall in annual reports inform of where the account can be accessed.

The account shall be updated and published no later than 30 June of each year and otherwise in case of significant changes to the enterprise's risk assessments. It shall be signed in accordance with the rules in Section 3–5 of the Accounting Act." (Åpenhetsloven, 2021).

According to the Transparency Act § 6, anyone has the "right to information from a business in writing about how the business handles actual and potential negative consequences under § 4. This includes both general information and information related to a specific product or service that the business offers." (Åpenhetsloven, 2021). The law has detailed rules on when a request for information can be rejected.

When the law came into effect on 1 July 2022, many believed that businesses would be overwhelmed with requests for information. However, it turned out that even well-known companies received a limited number of inquiries in the first few months after the law came into effect. Many requests from external stakeholders may involve costs to follow up. On the positive side, the requirement increases awareness of working conditions in supply chains, leading to improvements.

Since 30 June 2023, businesses subject to the Transparency Act have been required to publicly disclose a report on the legal due diligence they have conducted. It's observed that companies' reporting on their website on their due diligence varies from no report at all to over 30 pages. The level of reporting is often related to the company's size and level of recognition.

If anyone believes that a business subject to the Transparency Act (The Act) is not fulfilling its obligations under the law, they can file a complaint with the Consumer Authority. The Authority has already completed two complaint cases. In July 2022, shortly after the Transparency Act (The Act) came into effect, the organization Future in Our Hands requested information from IKEA regarding the follow-up on human rights at factories in Bangladesh and Pakistan. The Authority determined that IKEA did not need to conduct further investigations beyond what they had already done. More information about these cases is available on the Consumer Authority's website (https://www.forbrukertilsynet.no/english). In the future, the Consumer Authority will conduct inspections to ensure compliance with the law, not just address incoming complaints.

In the fall of 2022, Klassekampen (newspaper) filed a complaint against Posten (the Norwegian postal service) because they refused to disclose a list of their transport providers with vehicles weighing under 3.5 tons. The newspaper wanted to verify if these providers were being monitored as reported by Posten. The complaint was dis-

missed because the Consumer Authority believed it was unnecessary to disclose the names of these providers.

4 Theory of knowledge transfer as translation (KTT)

The Act is, as presented in the previous section, rather unclear. Many companies have several hundred suppliers: are all these supposed to be reviewed? How far down the supply chain are companies expected to report on? If breaches of human rights are detected in the supply chain, how much are companies required to spend on remediation? How companies understand the law varies. Some companies write almost 30-page due diligence reports, whereas others do not publish anything, i.e., not fulfilling the law. How can the understanding be so varied – and what is the story behind what is made available on the companies' website. I thus wanted to understand how The Act has been interpreted and implemented, i.e., how the law is interpreted and applied in companies.

The Knowledge Transfer as Translation (KTT) theory, introduced by Røvik in the seminal work "Trends and Translations – Ideas that Shape the 21st Century's Organizations," presents a model that navigates the intermediary space between modernistic perspectives and social constructivist paradigms. Rooted in pragmatic institutionalism, the KTT model emphasizes the pivotal role individuals play in moulding and re-interpreting ideas. It shares synergies with other pragmatic research methodologies, notably Weick's sensemaking framework (Weick, 1995) and Czarniawska-Joerges' concept of "storytelling" (Czarniawska-Joerges, 1997). The model suggests that individual translators are influenced by factors such as identity, retrospection, enactment, social interactions, current events, cues, and plausibility, which are integral to Weick's sensemaking process.

Translation theory is primarily applied to study how ideas and cultures are communicated and understood (Demir & Fjellström, 2012; Gold & Pedler, 2022; Røvik, 2007, 2016), for example, how business practices are communicated and understood between companies and subsidiaries in different countries of different cultures (Gold & Pedler, 2022). Another application of translation is on cultural variations emerging when ideas or practices are transferred from one country to another, like how Western economic ideas were introduced and applied in China (Chao, 2022). An even more encompassing challenge with translation is where the understanding of terms like "Corporate Social Responsibility" varies among people, organizations, and corporations (Dilling, 2011). In this study I applied the KTT through a new approach, namely I studied how companies, via individuals, interpret a concrete law and what concrete actions within the company follow.

With the introduction of The Act, companies find themselves in uncharted territory, lacking precedents to guide their compliance efforts. The interpretation and inte-

gration of The Act within corporate practices are significantly shaped by the individuals' or entities' comprehension of the law, essentially boiling down to their ability to make sense of the legal text. The KTT model not only illuminates the sensemaking activities surrounding the adherence to The Act within organizations but also progresses to assess the law's impact post-implementation.

Moreover, the KTT framework proposes practical and empirically validated methods that offer a structured approach for case comparison. This model stands out for its contribution to understanding the nuanced processes through which laws like The Act are internalized and operationalized within corporate environments, providing valuable insights into the dynamic interplay between legislation and organizational behaviour.

The process of translating organizational ideas into practice is intricately shaped at the juncture of two critical areas: 1) the supply of organizational concepts and innovations and 2) their adaptation and reception within a company. Research employing the Knowledge Transfer as Translation (KTT) theory has shed light on the evolution of other social and environmental topics, like Corporate Social Responsibility (CSR). This study's approach is inspired by the insights from the article *From corporate social responsibility awareness to action?* (Ditlev-Simonsen, 2010). Both studies are about the interpretation of relatively vague concepts and translation to integration in corporate operations. Whereas the CSR study looks into how corporate social responsibility is understood and what it is translated into in a company, this study, as explained, goes a step further by investigating how a concrete law is translated into corporate activities.

5 Method and data collection

The nascent understanding and application of The Act necessitate an exploratory research approach. To deepen our comprehension and enhance the conceptualization of the translation process, it is imperative to meticulously examine the journeys companies have embarked upon to integrate The Act into their operations. The comparative case study method, with its inductive reasoning to identify both similarities and differences across selected instances, stands out as particularly apt to tackle such research inquiries (Andersen, 2003; Eisenhardt, 1989a; Yin, 2000). It has been posited that while multi-case research tends to align more with the expansive narrative scope of books, single-case studies are often deemed more fitting for the concise format of academic articles, due to the detailed empirical depth involved in multi-case examinations (Eisenhardt & Graebner, 2007a).

I have opted for the multi-case study approach and addressed the challenge of limited space by succinctly presenting the cases. I posit that examining a range of distinctly contrasting cases will yield a more comprehensive understanding of the context at hand. Additionally, employing the Knowledge Transfer as Translation (KTT)

framework for case presentation not only mitigates space constraints but also enhances the clarity and accessibility of the findings.

The two cases selected for this article are based on convenience sampling. As the cases are very different with respect to size, market, recognition and familiarity among people in general, sector, etc., this polar type of cases contributes to illuminating the phenomenon (Eisenhardt, 1989b; Eisenhardt & Graebner, 2007b). Even though the companies are very different, they have in common that they are required to implement The Act. Companies that have implemented The Act are generally more willing to come forward, sharing their experiences and reports. Conversely, it proves more difficult to engage companies that have not achieved similar success or feel uncertain about how to proceed in sharing their experiences. This dynamic can intimidate companies struggling to get started, as it often appears that only the successful ones, typically those with ample resources, are able to share their stories. Consequently, including a company with limited experience and resources presented a challenge, yet it added a unique dimension to the study.

The translator's motivation, their distinct characteristics, and the principles and patterns guiding the translation, as well as the impact of The Act translation, are thus pivotal in grasping how The Act is decontextualized and subsequently recontextualized within organizations. This understanding is articulated through four essential questions in the Knowledge Translation Theory (KTT) framework (Røvik, 2007), specifically tailored to the introduction of The Act:

1. *The Translator's Incentives/Motivation:*
   Discussion revolves around questions such as: "What motivated both the company and you personally to adopt The Act, and how does this motivation relate to your role within the organization?"
2. *The Characteristics of the Translator:*
   Explore questions like: "What role did you play in the process of addressing the concept of The Act in the company, and what is your expertise and interest in social issues within the supply chain?"
3. *Translation Rules:*
   Delve into questions such as: "How was The Act implemented across different organizational contexts, and what impact has it had in various parts of the company?"
4. *The Effect of the Translation:*
   Examine questions like: "What has been the internal and external effect of The Act on the company?"

The interview was semi-structured, based on the four key questions presented above, and lasted for about 45 minutes to one hour. Afterwards a summary of the interview was written and sent to the interview subject for review. The interviewee had the opportunity to make changes.

6 Two cases describing how the Transparency Act (TA) was introduced and translated into corporate activities

6.1 Case 1: Posten Bring AS

Posten Bring AS is a "Nordic postal and logistics group that develops and delivers comprehensive solutions within post and logistics. We have over 12,500 skilled employees, and meet the market with the two brands Posten and Bring" (Posten, 2024). Posten has over 14,000 suppliers globally, of which over 2,000 are external transporters and 15 subsidiaries. The company is a "limited liability company wholly owned by the Norwegian government and is the parent company of the Group" (Posten Bring, 2022) and has a leading role in following up on social responsibility in the supply chain as well as being a responsible employer. "Posten Bring actively works to ensure that we respect fundamental human rights and decent working conditions in our own operations and in our supply chain." (Posten, 2023)

6.1.1 Interview

The interviewee, hereafter referred to as "the translator", was recommended by the company's Senior Vice President on Sustainability. This referral was part of the effort to find the right individual to discuss the implementation of the Transparency Act (The Act). The organization of the company's sustainability work has been reorganized during the last year. The team for responsible sourcing is now an integrated part of Group Procurement. As a consequence, the approach to sustainability issues has been slightly revised dependent on where in the company structure the sustainability group was placed. These changes in organization have happened independently of the Act. Still, independent of the location of the responsible procurement team, collaboration across the company has been a key issue and the team receives support from the law department, HR, HSE (Health, Safety, and Environment), and Purchasing. The translator is managing a team of three persons and has the overall responsibility for the integration of The Act issues.

6.1.2 The translator's incentives/motivation

As The Act came into force and the law was associated with ethics in the supply chain, the translator became the person in charge of managing the implementation of the law. Posten has for many years worked systematically on ensuring decent working conditions within the company as well as in the supply chain. Initially the law did

not explicitly go beyond activities the company was already voluntarily involved in. Overall, the company had the impression that the new law covered what was already addressed as soft law in operations.

6.1.3 The characteristics of the translator

Implementing The Act, a law with potential negative economic consequences if not adhered to, differed significantly from implementing voluntary principles, frameworks, policies, or self-developed standards. In a legal context, the translator's characteristics were less pronounced compared to voluntary initiatives. However, identifying priorities and approaches was still based on the translator's competencies and interpretation of the law.

6.1.4 Translation rules

Given the company's extensive experience in monitoring the supply chain, The Act mainly contributed to systematizing existing standards and activities related to decent working conditions. A shift resulting from The Act was a more formal procedure where the due diligence report must be processed and thoroughly reviewed by the group management, board, and its audit committee. When considering the law's impact in terms of anchoring, this requirement stands out as most crucial.

The requirement for Posten to respond to customer inquiries about its supply chain enhanced internal competency regarding supply chain working conditions. Posten has received a large number of requests for information, mostly in survey forms, to meet their reporting obligations. Although most required information is available in Posten's 31-page due diligence report, most customers prefer custom responses. The due diligence report itself is evolving during this process of responding to information requests. Posten's established educational system on ethical supply chain issues has been updated to include The Act–related content. Still, within the company, other factors over the past 2 years have been as influential as the law itself in garnering attention for the subject area.

6.1.5 The effect of the translation

The translation process has led to a more formal and aligned approach to managing decent working conditions in the supply chain, in particular in the area of risk assessment and reporting and enhanced organized transparency. Implementing The Act also exposed some unacceptable conditions in the supply chain, requiring follow-up. This increased focus has made employees more aware of and engaged in social issues

within the supply chain. However, the law's specificity regarding openness require-
ments remains a topic for ongoing discussion. Also, the fact that the law contains
room for interpretation has been met by additional requirements imposed by The
Consumer Authority (Forbrukertilsynet). As the role of the authority is to supervise
the market, not to extend existing laws, this has been a challenge.

Despite The Act's goal to improve awareness and conditions in the supply chain,
it is a challenge that the reporting element gets too much attention, relative to im-
proving conditions in the supply chain.

Ensuring that The Act doesn't become a bureaucratic burden remains a challenge.
Theoretically, if 9,000 companies, which is the number of companies The Act actually
applies directly to, request information from each other multiple times, it could lead
to significant additional work. Integrating this reporting into an existing system, like
the Brønnøysund Register Centre,[1] could make it more efficient. A common corporate
database for collecting relevant data required by The Act data would better align with
today's digital capabilities.

6.2 Case 2: Company Beta

Beta is a well-recognized machine contracting firm serving both the public and private
sectors with a specialization in infrastructure projects. Established nearly a quarter-
century ago, Beta is predominantly family owned and operated, boasting a workforce
of over 100 employees and an annual turnover exceeding 250 million NOK. The com-
pany is deeply committed to environmental stewardship and ensuring fair working
conditions, integral to its operations and brand identity. Beta actively contributes to
three Sustainable Development Goals (SDGs), highlighting its positive societal impact.
Furthermore, it holds ISO 9001 certification for its management system and ISO 14001
for its environmental management system, underscoring its dedication to quality and
sustainability.

6.2.1 Interview

The interviewee, referred to as "the translator", was invited to participate in the study
through a course on sustainability within an executive programme. Beta had not es-
tablished a programme in line with the Transparency Act (The Act) when it came into
force. The company had furthermore not encountered any customer inquiries regard-
ing human rights or working conditions within its supply chain. With no request to

---

1 The Brønnøysund Register Centre is a public centre collecting and providing information about en-
tities in both the public and private sector.

do due diligence and not being aware of the law, it was the annual ISO 9001 revision requiring third-party certification in September 2022 which brought The Act to their attention.

6.2.2 The translator's incentives/motivation

It was presumed that Beta, as a local firm compliant with Norwegian regulations, fulfilled all legal requirements. Moreover, the voluntary attainment of ISO 9001 and ISO 14001 certifications underscored the company's commitment to environmental and social matters. Winning bids in this sector often necessitates stringent management of social aspects, such as working conditions and wages, ensuring they align with regulatory standards. The company's esteemed reputation in addressing environmental and social issues fostered a sense of confidence among both owners and employees that everything was well managed. Therefore, it came as an unexpected revelation when the ISO certifier pointed out the company's non-compliance with The Act regulations.

6.2.3 The characteristics of the translator

The company's administration consists of a small team operating within a relatively flat organizational structure, where strategy and formal procedures were not clearly defined. As the head of accounting and finance, and with a personal interest in environmental and social issues, the translator took the initiative to address the challenges related to The Act highlighted by the ISO certifier.

Interest in pursuing The Act compliance was primarily confined to the administrative team. In contrast, employees engaged in operational tasks and machinery management on the factory floor showed little interest or concern regarding environmental, social, and governance (ESG) matters.

6.2.4 Translation rules

The translator's initial step was to assess the company's suppliers to pinpoint those with the highest risk of violating human rights within the supply chain. Significant concerns were noted particularly among suppliers of work clothing and certain minerals imported from developing countries. Around 20 suppliers flagged for potential issues were subsequently monitored.

The method involved scrutinizing these companies' websites for their disclosures on social issues within their supply chains, complemented by conducting several interviews. The translator estimated that this due diligence process took about a week to complete. Although this review did not uncover any actual or potential adverse im-

pacts within the supply chain, it proved instrumental in establishing a formal framework for systematic evaluation and reporting on the current state of the supply chain.

A customer's criticism of the company's offer, attributing it to inadequate Act-mandated reporting, intensified the commitment to compliance. Consequently, a more detailed report concerning The Act was prepared, and the company is currently awaiting the contractor's feedback. Once approved, this report will be published on the company's website and made available to the public.

6.2.5 The effect of the translation

The primary outcome of the translator's efforts was an enhanced understanding and a structured approach to addressing supply chain issues. The translator gained a deeper insight into the company's supply chain dynamics and identified potential areas of concern. It was an unexpected discovery for the translator that, despite initial beliefs, the company's operations in Norway were not exempt from social challenges, indicating that the working conditions in Norway also warranted further scrutiny.

The frequency of conducting these reviews remains a point of uncertainty for the translator. To facilitate smoother integration with The Act requirements, the development of more standardized contracts is planned. Additionally, a code of conduct aligned with The Act has been formulated and is now accessible on the company's website. However, there is a recognized need for a more formalized framework to effectively implement The Act guidelines.

7 Case comparative analysis

Comparing Posten's and Beta's approaches to the Transparency Act (The Act) reveals both similarities and differences in how they manage and perceive their responsibilities under the law. The comparison is structured around the four interview points: the translator's incentives/motivation, the characteristics of the translator, translation rules, and the effect of the translation.

7.1 The translator's incentives/motivation

Both companies recognize the importance of integrating The Act issues into their operations.

In Posten one person was assigned the task of managing the implementation of The Act, whereas in Beta it was one person picking up on how to address The Act.

In Posten, the translator's role emerged from an existing commitment to decent working conditions and human rights in the supply chain. The implementation of the Act was seen as an extension of ongoing practices rather than a new initiative. Posten has had employees working on ensuring acceptable operations in the supply chain for many years. This is due to, among other things, the visibility of the company, which is a leading provider of services and partly owned by the government. A key element, however, is that the company size allows for resources to follow up on new regulations.

In Beta, the translator was motivated by the discovery of the company's non-compliance through an external ISO certification process, indicating a reactive approach to the Act. With a small administration Beta did not have resources for carefully following up the continuous flow of new regulations, some more and some less relevant for the company. Whereas Posten was following the development of the Act, Beta was working on ensuring operations in line with existing laws and regulations. Being a family-owned and -operated company, Beta was taking pride in following existing laws and regulations while missing The Act being introduced and becoming into force. The main incentive and motivation for Beta was thus information from the ISO external certifier and later the customer's demand for due diligence of operation.

7.2 Characteristics of the translator

Both translators have a significant degree of responsibility for integrating The Act issues within their respective organizations.

In Posten, the translator operates within a larger, more complex organizational structure with a dedicated team and support from various departments. Posten having a resource pool to keeping track of development and expectations related to sustainability issues and human rights, the company had the opportunity to appoint one person dedicated to follow up the Act. Given this more formal structure and division of responsibilities, the translator had, however, less room for taking personal choices on how to follow-up the Act.

In Beta, the translator's role is more solitary, with the challenge primarily confined to the administrative level of a smaller organizational structure. The person picking up on The Act was free to choose how to address the issues. No detailed steering of the translator job gives him/her the opportunity to do this in his/hers own way. The latter had a unique interest in sustainability issues, and the approach was marked by his/her understanding of what was necessary.

7.3 Translation rules

Both companies have taken steps to assess and monitor their supply chains in response to the Act, indicating an awareness of the law's requirements for due diligence and transparency.

Even though Posten already had a system in place to follow up the supply chain, this was fragmented with different approaches in different part of the operations. The introduction of The Act provided a "push" to get a better structure on the task of following up the supply chain. Also, because The Act required the top management to endorse and sign the Transparency Act Report, following up the due diligence process in the supply chain became higher on the priority list of the top management team. Questions and information requests from the outside regarding the company's work on human rights in the supply chain operations further required better knowledge among employees. These circumstances enforced the demand for a more explicit structure of response. So, for Posten the response to the Act was driven both by internal pull and by external push.

In Beta, the approach was initially more *ad hoc*, focusing on identifying high-risk suppliers and conducting due diligence to establish a formal framework for compliance. Beta's efforts were catalysed by customer feedback and the need to meet specific reporting requirements. The main driver or push was from the outside. Still, also in Beta a structured system was established for reviewing suppliers and identifying red flags. This was managed by the translator by him/herself.

7.4 Effect of the translation

Both companies have experienced an enhanced focus on social issues within their supply chains as a result of implementing The Act. It was to be expected that for Beta, since it was not initially aware of The Act, the implementation resulted in both increased knowledge and action. However, Posten too, even with an already existing system for following up the supply chain found issues requiring actions as a result of complying with The Act.

In Posten, the translation process has led to a more formal and aligned approach to supply chain management, with an emphasis on risk assessment, reporting, and transparency. The company has faced challenges in balancing the law's reporting requirements with actual improvements in supply chain conditions. External demand for information and media requests have made Posten more aware of issues to follow up. The primary outcome has been a structured approach to addressing supply chain issues and the realization that even domestic operations can face social challenges.

Beta is still in the process of formalizing its framework for compliance with The Act. Individual and media requests for further information in line with The Act have not been an issue for Beta. For Beta, the focus on the Act was initiated by an external

consultant, then followed up by a dedicated employee. Still, eventually Beta would have had to adhere to The Act given that customers would require it. So, for most companies, in a longer perspective, The Act is a driver for awareness and action on human rights and transparency issues.

In conclusion, the comparison highlights that while both Posten and Beta are committed to complying with The Act, their approaches reflect their size, existing commitments to social responsibility, and the maturity of their sustainability programs. Posten's response is characterized by an extension and formalization of existing practices, supported by a robust organizational structure. In contrast, Beta's approach is more emergent, driven by external certification and customer feedback, reflecting the challenges smaller companies may face in adapting to new regulatory requirements. Still, it is evident that individuals' attitude, role and engagement impact how the Act is acted upon.

8 Conclusion

This study sheds light on the nuanced ways in which corporations navigate the interpretation and implementation of the Norwegian Transparency Act, revealing the multifaceted nature of legal compliance within diverse organizational contexts. By employing the Knowledge Transfer as Translation (KTT) theory, the research delves into the interpretive processes that underpin corporate actions in response to the Act, highlighting the pivotal role of individual actors within the organizational framework. The KTT has been frequently applied to study translation of corporate cultural issues across countries, whereas this study applies the KTT to study the translation of laws and regulations into corporate engagement and output. The comparative analysis of two distinct companies illustrates that despite the Act's uniform requirements, its implementation is anything but homogenous, shaped significantly by the companies' sizes, resources, individuals and internal cultures.

The findings underscore the complexity of translating legal mandates into practical corporate strategies, where the broad and open-ended nature of The Act serves as both a challenge and an opportunity for innovation in compliance practices. The study reveals that the Act's interpretation is not merely a legal exercise, but a dynamic process influenced by corporate values, stakeholder expectations, and the strategic priorities of individual translators.

In the introduction, the concept of *voluntary* Corporate Social Responsibility (CSR) was contrasted to statutory *regulations*, now intriguing thoughts emerge from this study. Both voluntary initiatives and legal requirements lead to changes within companies. However, regardless of whether these actions are voluntary or mandated by law, companies often struggle with how to proceed. One might expect that legal requirements would provide clear guidance, prompting companies to undertake straightfor-

ward and comparable activities. Nevertheless, our case indicates that this is not always the case.

In essence, the research contributes to a deeper understanding of the legal compliance landscape, suggesting that the path to effective implementation of The Act lies not in the pursuit of uniformity but in embracing the diversity of corporate interpretations and actions.

This approach not only enriches the discourse on corporate responsibility and transparency but also opens avenues for future research on the interplay between law, corporate culture, and individual agency in the realm of ethical business practices.

Conducting the same KTT method and data collection in the study of additional corporations will contribute to an improved understanding of the effect of The Act. Also comparing companies of similar sizes in the same sector would also be an interesting approach in further research.

This study analyses two company cases. The first possesses extensive resources to comply with the law. The second, Beta, is a smaller company, yet relatively large compared to typical Norwegian firms. The contrast between our two case samples and the broad target of The Act raises questions about the other 9,000 companies required to fulfil it. This illustrates both a limitation of this study and an opportunity for further research.

This study explores approaches to human rights and due diligence in the supply chain from a *purchaser* point of view. However, the analysis of the *suppliers'* comprehension and interpretation of legal standards is limited. The extent to which suppliers embrace the requirements of the Transparency Act in their operations remains unclear, as does the degree of alignment between local or national authorities' views and the buyers' perceptions of human rights. Additionally, it is uncertain how much end-use customers support the measures implemented to comply with the Transparency Act. Further research in this field is necessary.

Additional questions are suggested by chapter 6, where Midttun examines whether initiatives like the Transparency Act represent yet another effort by Western countries to impose their values and norms on emerging and developing economies. It appears that Western standards encounter significant resistance from more influential Eastern nations, such as China. This may raise challenges for the effective implementation of the supply chain monitoring and management required by The Act. By the time this book is published in December 2024, a year after the interviews, the scenarios described in this chapter will likely have evolved. A follow-up study will be particularly valuable for gaining deeper insights into the implementation of the law.

References

Andersen, S. S. (2003). *Case-studier og generalisering* (2. opplag ed.). Fagbokforlaget Vigmostad & Bjørke AS.

Business and Human Rights Information Center. (2024). *France's Duty of Surveillance Law* https://www.business-humanrights.org/pt/temas-relevantes/corporate-legal-accountability/frances-duty-of-vigilance-law/

Chao, H.-K. (2022). Two Types of Travelers in the History of the Transmission of Western Economic Knowledge in Late Qing and Early Republican China [Article]. *History of Political Economy*, *54*(3), 437–458. https://doi.org/10.1215/00182702-9779600

Czarniawska-Joerges, B. (1997). *Narrating the organization: dramas of institutional identity* (Vol. Chapter 1). Chicago University Press.

Demir, R., & Fjellström, D. (2012). Translation of relational practices in an MNC subsidiary: Symmetrical, asymmetrical and substitutive strategies [Article]. *Asian Business & Management*, *11*(4), 369–393. https://doi.org/10.1057/abm.2012.13

Dilling, P. F. A. (2011). Stakeholder Perception of Corporate Social Responsibility. *International Journal of Managment and Marketing Research*, *4*(2), 23–33.

Ditlev-Simonsen, C. (2016). Rapportering om samfunnsansvar i små foretak. In H. o. B. T. Anderssen (Ed.), *Moderne forretningsjus III*. Fagbokforlaget.

Ditlev-Simonsen, C. D. (2010). From corporate social responsibility awareness to action? *Social Responsibility Journal*, *6*(3), 452–468. https://doi.org/10.1108/17471111011064807<http://dx.doi.org/10.1108/17471111011064807>

Eisenhardt, K. M. (1989a). Building theories from case study research [Article]. *Academy of Management Review*, *14*(4), 532–550. http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=4308385&site=bsi-live

Eisenhardt, K. M. (1989b). Building Theories from Case Study Research. *Academy of Management Review*, *14*(4), 532. http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=4308385&loginpage=login.asp&site=ehost-live

Eisenhardt, K. M., & Graebner, M. E. (2007a). Theory building from cases. Opportunities and challenges. *Academy of Management Journal*, *50*(1), 25–32. http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=24160888&site=bsi-live

Eisenhardt, K. M., & Graebner, M. E. (2007b). Theory building from cases. Opportunities and challenges *Academy of Management Journal*, *50*(1), 25–32. http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=24160888&site=bsi-live

Endringslov til regnskapsloven. (2013). Lov om endringer i regnskapsloven og enkelte andre lover [Act on amendments to the Accounting Act and certain other acts] (LOV-2013-04-19-15). Lovdata. http://lovdata.no/dokument/LTI/lov/2013-04-19-15

European Coalition for Corporate, J. (2017). French corporate duty of vigilance law: FAQs. https://corporatejustice.org/publications/faqs-french-duty-of-vigilance-law/

European Commission. *Corporate sustainability and responsibility*

Gold, J., & Pedler, M. (2022). Bridging research to practice via action learning [Article]. *Action Learning: Research & Practice*, *19*(1), 3–18. https://doi.org/10.1080/14767333.2021.2018288

Harris, K. D. (2015). *The California Transparency in Supply Chain Act – A Resource Guide*.

Modern Slavery Act 2015, https://www.legislation.gov.uk/ukpga/2015/30/contents/enacted (2015).

Posten. (2023). *Our work with the Transparency Act – Posten Bring*.

Posten, B. (2024). *About us*. Retrieved January from

Posten Bring. (2022). *Trust in tough times Integrated annual report 2022 Posten Norge*.

Røvik, K. A. (2007). *Trender og Translasjoner – Ideer som former det 21. århundrets organisasjon* (Vol. ISBN 978-15-01116-5). Universitetsforlaget.

Røvik, K. A. (2016). Knowledge Transfer as Translation: Review and Elements of an Instrumental Theory [Article]. *International Journal of Management Reviews*, *18*(3), 290–310. https://doi.org/10.1111/ijmr.12097

Weick, K. E. (1995). *Sensemaking in Organizations*. Sage Publications, Inc.

Wig, K. a. S. T. (2023, June 8th). Espresso House slettet timer ansatte jobbet ekstra. *Dagens Næringsliv*.

Yin, R. K. (2000). *Case study research* (Vol. Volume 5). Sage Publications, Inc.

Aarseth, T. (2023, Augutst 21st 2023). *Her jobbet polakker til 69 kroner timen og uten rettigheter*. NTB. https://frifagbevegelse.no/nyheter/her-jobbet-polakker-til-69-kroner-timen-og-uten-rettigheter-6.158.648631.861fdc683a

Åpenhetsloven. (2021). Lov om virksomheters åpenhet og arbeid med grunnleggende menneskerettigheter og anstendige arbeidsforhold (åpenhetsloven) [Act relating to enterprises' transparency and work on fundamental human rights and decent working conditions (Transparency Act)] (LOV-2021-06-18-99). Lovdata. https://lovdata.no/dokument/NL/lov/2021-06-18-99

Siri Granum Carson and Matthias Kaiser

Conclusion
A sea change in science and technology ecologies? Prospects of socially responsible research and innovation towards 2030

1 A crisis of science and technology?

Even the most ardent critics of science and technology would not want to turn back the clock and forego the knowledge-based benefits with which we now endow our lives. Our industries are based on growth by introducing ever new inventions on the market. Accordingly, people's expectations of wellbeing and a good life have increased steadily. At the same time, the Janus face of science and technology, the realization of its goods together with its bads, has become a common perception. The free production of scientific and technological knowledge needs to be balanced by recognition that freedom must have its counterpart in responsibility. We have seen that often new breakthroughs and products with some obvious immediate benefits bring about other and negative impacts, which create new problems to solve for science and technology.

And by now, we find ourselves caught in a world in which the complexity of impending problems is the norm of the day. The systems portrayed in scientific disciplines do not develop in isolation but interact with each other, and the machine-human interaction crosses all fields with ever-new risks and impacts. Climate change is probably the best example of this. Recognized as a threat by science early on, it did not make the policy agendas before it actually and recognizably materialized at a breathtaking speed. But it is only one of several instances of the so-called global grand societal challenges. The recent COVID-19 pandemic finally brought home the realization of our vulnerability and the interconnectedness of most fields of our life. The environmental, political, economic, health, social, and cultural spheres came together in a cascade of risks which have not gone away with the speedy resolve of the pandemic as such, greatly facilitated by the scientific development of vaccinations. Instead, the cascade of risks is becoming more complex and threatening.

Meanwhile, we have social cohesion falling apart in many countries. We have increased economic inequity with the spread between the rich and the poor becoming bigger, while the global economy is more volatile than ever. And the public trust in science, the belief that science carries the keys to a better life, is under siege. What even large parts of science regard as facts often become just mere opinions or even elements of conspiracies in the eyes of significantly many people.

So, are we facing a crisis of science in our days? People judge this differently, and we may (as our colleague Silvio Funtowicz has done) quote Antonio Gramsci in this connection, with reference to the social system:

> *The crisis consists precisely in the fact that the old is dying and the new cannot be born; in this interregnum a great variety of morbid symptoms appear* (Gramsci, 1971).

The word "crisis" is certainly value laden, also given its role in socialist or communist parlance. But it seems one can assert that science and technology development is faring in troubled waters. The ship of science that Otto Neurath wanted to repair while afloat in the open seas[1] needs a thorough overhaul. In the literature and certainly in the present book, we detect signs of an incumbent *sea change* in our current research and innovation ecologies. Old boundaries are broken down, new challenges in the scientific community are realized, bridges to policy making are sought, new models of cooperation with industry are experimented with, and many more scientists are driven not only by their wishes for their own career but more prominently by their social conscience.

Scientific and technological knowledge has become more and more specialized. The late 19th and then the 20th centuries saw the emergence of specialized disciplines and sub-disciplines. Derek De Solla Price (1963) noted already in the early 1960s the exponential growth of virtually all indicators of scientific research and knowledge production. Millions of annually published scientific articles seldom meet the expectations of actionable knowledge. Outputs from disciplinary "siloes" suffer typically from tunnel view. Even techno-science and technology seldom produce innovation which is robust in socio-cultural environments.

What has this implied for higher education and the funding of research? Reaction to this varies in different regions of the world. The European Commission reacted with a series of efforts to counterbalance the situation described above. From ELSA (ethical, legal, and social aspects) to science-in-society, to science-with-and-for-society, one moved to Responsible Research and Innovation (RRI), and further on to Open Science and mission-oriented research and innovation. Today, we are in a situation where dedicated funding for transdisciplinary efforts to steer research and innovation in a sustainable direction remains low, all while the verbal endorsement for such ideas remains high. What is needed in order to move from talking the talk to walking the walk? Let us take a closer look at the situation in Norway to better assess this.

---

1 Neurath, 1921, p. 199: "We are like sailors who must rebuild their ship on the open sea, never able to dismantle it in dry-dock and to reconstruct it there out of the best materials. Where a beam is taken away a new one must at once be put there, and for this the rest of the ship is used as support. In this way, by using the old beams and driftwood the ship can be shaped entirely anew, but only by gradual reconstruction".

2 The Norwegian case revisited

Given the international state of affairs presented in the last section, let us now turn to how we are faring in the Norwegian research and innovation system. As shown in previous chapters (see Introduction, chapters 1 and 3), Norway has been an early mover and, in many ways, "best in class" when it comes to implementing respectively the CSR and RRI agendas. So, to paraphrase Gro Bruntland, is it typically Norwegian to be socially responsible?[2] Furthermore, in what sense and to what extent has AFINO delivered on the task it was given by the Research Council of Norway, to help transform and make the Norwegian research and innovation system (even more) socially responsible?

The RRI agenda, understood as an attempt to improve the adaptivity of scientific information to socio-economic realities, found particular resonance in Norway (Owen & Pansera, 2019). A remarkable response was noted in the published literature on RRI originating from Norway (cf. chapter 3). In the CSR literature, it has been argued that Nordic countries seem to share a culture of trust in institutions and flat-decision structures which is conducive to CSR (Strand & Freeman, 2015). Cooperation and dialogue between different sectors of society, including industry, were promoted and often realized. This book bears witness to some of it.

Ethics of and in science has also been promoted by Norwegian authorities for a considerable time, starting with the establishment of three national committees of research ethics, one for each sector of science (NENT = natural science and technology, including agriculture and fisheries; NESH = social sciences and humanities, psychology; NEM = biomedical research and health sciences). These three committees were established by the Norwegian parliament (Stortinget) already in 1990.[3] Guidelines for ethical research have been developed by these committees and are referred to in the Norwegian Act on research ethics. It is also remarkable that ethical issues of the sciences receive quite a lot of attention in the public media and in official statements from the authorities. At the time of the writing of this article, the directorate for higher education and competence (HK-dir) together with the national committees for research ethics invited for a seminar on how ethical issues of research can be handled in international collaborations. It will be argued that the question of the responsibility of science is central to research ethics, and how to deal with ethical commitments according to the ethical guidelines in research with "difficult" countries.

---

2 This formulation is an attempt at an inside joke, referring to then prime minister Gro Harlem Brundtland's (in Norway) famous new year speech from 1992, where she stated that it is "typically Norwegian to be good".

3 Since then they have been supplemented by a National Commission for the Investigation of Research Misconduct (GRU, since 2007) and a National Committee on the Ethics of Human Remains (Sjelettutvalget, since 2008).

The guidelines for the NENT committee include a relatively comprehensive section on the obligations of scientific research towards society. Here it is stressed that research has an independent responsibility towards societal development. Furthermore, it is stressed that the sciences should contribute to sustainable development and to increased global justice. One may suspect that this is just a lot of words without significance for how the actual research is conducted. But not quite so. For instance, in 2014 when a complaint was made that the University of Bergen uses funds from the oil industry to promote what is essentially an unsustainable form of industry, the Rector decided to send this question to the NENT with reference to the ethical goal of sustainability in science. NENT concluded, among others:

> *NENT believes that it is irresponsible in terms of research ethics if the framework conditions and research activities of petroleum research hinder adjustment processes so that the UN's climate goals, which Norway has committed to, cannot be achieved.*

As a consequence, the funding practices at UiB were changed,[4] and the case sparked lunchroom discussions, op-eds, panel debates, and meetings across Norwegian research environments.

Based on this background, one may take a slightly optimistic outlook on how socially responsible science is practiced in Norway. Perhaps we can point to some achievements up here in the far North of Europe that are not mirrored in the rest of Europe? Opinions may differ on this, specifically in terms of distinguishing between real achievements and mere window dressing.

Taking a closer look behind the scenes at Norwegian research, and in particular the funding practices of it, coordinated by the Research Council of Norway (RCN), may also bring in some second thoughts. The SAMANSVAR program which funded the AFINO centre was discontinued in 2020 as a result of an internal reorganization process in the Research Council. This program was founded on an ambitious idea of challenging the *modus operandi* of the Norwegian research and innovation system through dedicated funding of transdisciplinary projects and centres that were given room and resources to experiment and train a new generation of researchers in the art of cocreation and foresight.

We may view our own centre, AFINO, as a 5-year "spa treatment," allowing for testing out of new learning arenas that point beyond the traditional scientific training. However, while this "spa treatment" has no doubt been refreshing and beneficial for those of us who have been directly or indirectly involved in the activities of the centre, the intended systemic transformation of the research and innovation system has failed to materialize (as stated in chapter 4). Of course, one may argue that such a transformation cannot be expected to emerge from a single "stunt" such as this. On

---

4 Cf. also: https://www.theguardian.com/business/political-science/2014/feb/10/norwegian-universities-consider-the-ethics-of-oil-and-gas-research

the other hand, the AFINO hub with its nodes (four connected research projects) marked the endpoint of a longer stretch of SAMANSVAR-financed projects and activities, including the "sister centre" Digital Life Norway. In the call text, it was emphasized that AFINO should engage the institutions and change the role of researchers and that this means developing new competence, skills, and capacity in business and research organizations as well as in the Research Council itself. It was required that the project description should detail "how the Research Council can have an active role in the coproduction that will take place in the undertaking."[5]

However, this call for systemic transformation and true transdisciplinarity was in reality issued from just one corner of the Research Council and failed to diffuse throughout the organization (cf. chapters 3 and 4). The AFINO experiment has coincided in time with a period of considerable economic and organizational unrest in the Research Council (and well, of course, a pandemic), which in part accounts for the lack of resonance, and even dialogue. Meanwhile, the Research Council, and in general, the Norwegian research and innovation system, continues to operate on the basis of a quite linear view of the relationship between science and the world. Ideologically speaking, the research policy is still led from a relatively unilateral perspective of what (relevant) knowledge is. One clear example is the lack of integration of local and indigenous knowledge into the core of the Norwegian research and innovation policy, an area where countries such as New Zealand, Canada, and South Africa seem to be miles ahead.

The Norwegian system has proven ambitious when it comes to establishing a paradigm of socially responsible research and innovation. What may be lacking is the reflexivity needed to become aware of our own blind zones, as well as a recognition of the diversity of Norwegian value systems. The Norwegian self-image of a homogenous society gives little room for diversity, and this may be a hindrance when it comes to fulfilling our ambitions. We still have a distance to walk from the RRI talk in the sense of reflexivity, inclusiveness, and responsiveness. There are also challenges to solve in terms of making genuine transdisciplinary research possible, not least when it comes to the unevenly distributed political and economic power of the actors involved. Here, the relatively flat structure of Norwegian society may make it a fertile ground to experiment with new forms of dialogue between research, business, and civil society.

---

5 Quoted from the Research Council's response letter to our project sketch received in December 2018, our translation.

3 Where to from here? Policy recommendations towards 2030

With SAMANSVAR and AFINO, seeds have been sowed in the Norwegian research and innovation community. How do we fertilize the ground and make sure these seeds bloom in the years to come? In 2024, the Norwegian Research Council launched its new RRI web pages,[6] sending a strong signal that, at least in the Norwegian research and innovation system, the RRI agenda endures. Whether the concept as such makes a comeback in the next framework program or not, a recent OECD report (OECD, 2024) indicates with all clarity that the European dream of a research and innovation system directed towards social justice and sustainability is alive and kicking. The tag we put on our attempts to realize this dream may not be so important. However, the tendency to "re-brand" the agenda every few years comes with the risk of losing some of the capacities and lessons learned within one program in the eagerness to reinvent the wheel within the next one. On the other hand, it is important to continue to develop rather than stand still. In our case, a natural next step would be to continue the learning processes of the SAMANSVAR centres and projects together with the funding agencies, as the plan was from the beginning. Transforming the research and innovation system cannot happen unilaterally on the side of the researchers and entrepreneurs – without a corresponding transformation on the side of the funding and research policy instruments, all good intentions will remain intentions.

We see several areas and topics where institutional reform will be necessary. As a matter of fact, we would claim that a change in the basic science culture is emerging and that institutional reform is necessary to bring this about. One overarching priority should be to counteract the increasing fragmentation of, and reliance on, scientific disciplines. When calls are out for funding, scientific disciplinary excellence is still one of the main priorities. But the excellence covers smaller and smaller areas of expertise, defined by disciplinary and sub-disciplinary standards. If we compare this to the complexity of our realities, it implies that ever smaller segments of realities are studied, and thereby typically excluding other segments and their interactions. Many scholars have long since realized that we need to break out of the disciplinary siloes and cross academic boundaries. Once we realized the complexity of the world, we tried multi-disciplinarity and then interdisciplinarity schemes, hoping for radical interdisciplinarity, crossing all faculties and taking on social science and humanities. But the results were meagre, and we realized that people had a hard time working on developing even such basic things as a common language. Scientific journals are also not accustomed to publishing contributions which transgress disciplinary boundaries.

Related to this is another factor deeply rooted in dominant scientific culture, namely the view that scientific knowledge is basically a linear route to what can be

---

6 http://www.forskningsradet.no/en/research-policy-strategy/rri

called the final truth, even when that may be unattainable in practice. Many people stick to the idea that "objectivity" as a scientific gold standard implies the exclusion of non-reductionist alternative approaches, and that all indeterminacies (except in quantum physics) are due to our epistemic uncertainty which in the long run can disappear and lead to convergence. It is only in more recent years that some philosophers of science have promoted the view that the object of our scientific study is not the systemically closed and united world but a much more fragmented perspective on the world. The Norwegian philosopher Arne Naess promoted a similar view already in 1972 (Naess 1972; cf. also Bogdan 1974). The renowned philosopher and anti-fundamentalist Nancy Cartwright talked about a patchwork of mutually unrelated laws which help us highlight aspects of reality (Cartwright 1999). More recently, Giere (2006) and Massimi (2022) have defended perspectivism in science, a view also promoted by Saltelli and colleagues (2020). The upshot is no less than this sea change we mentioned, a sea change in the culturally entrenched view about the potential of science to help us manage our realities. While the old dogma of objectivist fundamentalism provided the illusion of the linear road to the "good" outcome (e.g. sustainability), these new perspectives now tie the knowledge closely to us as subjects with diverse interests which open up multiple perspectives on reality. When knowledge is no longer de-coupled from the knower, then we may readily realize how social responsibility is constitutive of the quality of the knowledge we produce. We see this as the core of the RRI and CSR frameworks.

We mentioned already that AFINO failed with respect to implementing these new ideas and approaches within the main pillars upon which the funding of Norwegian research is built, as in the organizational and institutional setting of the RCN. While recognizing that cultural changes like those discussed above cannot be achieved overnight or with a few twists of the functional nuts and bolts of research funding, we maintain that more adequate and targeted measures to promote genuine transdisciplinary research would be significant first steps to the cultural change we advocate.

We venture, however, some more concrete advice to the RCN which seems to follow from the above reflections and the current literature (cf. also Kaiser & Gluckman, work in progress). From AFINO's side, towards the end of our funding period, here are our top three recommendations for how to best promote a transformative research and innovation system in the years to come:

– Secure dedicated (and stepwise) funding mechanisms for transdisciplinary research. These steps need to differentiate between (i) shaping, (ii) supporting, and (iii) evaluating and communicating (cf. Kaiser & Gluckman 2023; Carew & Wickson 2010).
– Allow for adequate, long-term (and flexible) time dimensions for developing new capacities in the research community – the traditional 2- to 5-year project fundings are not the way to build up inter- and transdisciplinary expertise that goes beyond disciplinary boundaries.

–   Promote open, non-hierarchical, and collaborative leadership of projects, allow-
    ing non-academic partners ownership and data sovereignty, and include people/
    units for brokerage with policy.

References

Bogdan, R. J. (1974). Arne Naess, "The Pluralist and the Possibilist Aspect of the Scientific Enterprise" (Book
    Review). Theory and Decision, 5(3), 353.
Carew, A. L., & Wickson, F. (2010). The TD wheel: a heuristic to shape, support and evaluate
    transdisciplinary research. Futures, 42(10), 1146–1155.
Cartwright, N. (1999). The dappled world: A study of the boundaries of science. Cambridge University
    Press.
Funtowicz, S. O. & Ravetz, J. R. (1993), "Science for the Post-Normal Age", Futures 26/6: 568–582.
Giere, R. N. 2006. Scientific Perspectivism. The University of Chicago Press, Chicago.
Gramsci, A. (1971). Selections from the Prison Notebooks. International Publishers, New York.
Kaiser, M. & Gluckman, P. (in progress) "Are scientific assessments and academic culture impeding
    transformative science?"
Kaiser, M. & Gluckman, P. (2023). Looking at the future of transdisciplinary research. International Science
    Council, April 2023. DOI: 10.24948/2023.05. Available at:https://council.science/publications/looking-
    at-the-future-of-transdisciplinary-research/
Massimi, M. (2022). Perspectival ontology: Between situated knowledge and multiculturalism. The Monist,
    105(2), 214–228.
Naess, A. (1972). The pluralist and possibilist aspect of the scientific enterprise.
Neurath, O. (1921). Anti-Spengler: Dedicated to the young and the future they shape. In Empiricism and
    sociology (pp. 158-213). Dordrecht: Springer Netherlands.
Owen, R., & Pansera, M. (2019). Responsible innovation and responsible research and innovation
    (pp. 26–48). Edward Elgar Publishing.
Saltelli, A., Benini, L., Funtowicz, S., Giampietro, M., Kaiser, M., Reinert, E., & van der Sluijs, J. P. (2020). The
    technique is never neutral. How methodological choices condition the generation of narratives for
    sustainability. Environmental Science & Policy, 106, 87–98.
Strand, R., & Freeman, R. E. (2015). Scandinavian cooperative advantage: The theory and practice of
    stakeholder engagement in Scandinavia. Journal of business ethics, 127, 65–85.

List of figures

List of tables

Index